

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

Document Title: Evaluation of Bullyproofing Your School
Author(s): Scott Menard ; Jennifer Grotmeter ; Daniella Gianola ; Maura O'Neal
Document No.: 221078
Date Received: January 2008
Award Number: 2004-IJ-CX-0082

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

Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

**EVALUATION OF BULLY-PROOFING YOUR SCHOOL:
FINAL REPORT**

Scott Menard
Jennifer Grotspeter
Danielle Gianola
Maura O'Neal

Research on this project was supported by grants from the National Institute of Justice (NIJ 2004-IJ-CX-0082), the Office of Juvenile Justice and Delinquency Prevention (OJJDP 1998-MU-MU-K005) and a second grant from OJJDP in collaboration with the Centers for Disease Control and Prevention (OJJDP 1999-JN-FX-K006).

Executive Summary

Bullying in school is a major social problem with severe consequences to physical and mental health, and it has been implicated in the most severe forms of school violence. Schools are in need of effective programs to reduce bullying and improve school safety. Bully-Proofing Your School (BPYS) is a school-based intervention program designed to reduce bullying and school violence; it differs from other anti-bullying programs by providing teachers with a specific curriculum that can be implemented in the classroom. The present study is an evaluation of BPYS at the elementary school and middle school level.

Program Targets

BPYS targets primarily elementary and middle school students in the school context. As a whole-school intervention, adult faculty and staff in the school are also involved as both secondary targets of the intervention (to change their behavior to produce a school climate more unfavorable to bullying and more favorable to school safety) and as agents for delivering the intervention directly to the students. As part of the program, teachers are given information and strategies which help them to recognize bullying and intervene appropriately in bullying situations.

Program Content

Bully-Proofing Your School (BPYS) is designed as a comprehensive, school-based intervention with three major components: (1) heightened awareness of the problem of bullying, involving a questionnaire to assess the extent of bullying in the school, and the creation of classroom expectations and rules regarding no tolerance for bullying; (2) teaching protective skills for dealing with bullying, resistance to victimization, and providing assistance to potential victims of bullying; and (3) creation of a positive school climate through promotion of a “caring majority” in the school which works to alter the behavior of bystanders.

As part of the first component, all members of the school community, adults and students, commit themselves to a nontolerance policy about bullying, and to creating a caring community. School rules and expectations are established that are understood and enforced throughout the community. All systems in the school are addressed, from administration to transportation, and specific steps for implementing the school-wide program are included. The second component of the program teaches skills and strategies to help individuals avoid being victimized, including knowing how and when to get help from others, when to (and when not to) stand up to a bully, how and when to walk away from a threatening situation, how to use humor to defuse a threatening situation, and thinking positively about oneself. The third component of the program involves broad efforts to change the overall school culture, rather than only focusing on specific individual skills, with the goal of creating a positive, prosocial school climate that feels safe and secure for all members of the school community.

The intervention includes a classroom curriculum, consisting of seven sessions, with two optional sessions on conflict resolution and diversity. This curriculum is taught by the classroom teacher or mental health staff at the school, once a week, from thirty to forty-five minutes per

session, depending on the age of the children. There is an abbreviated three-session curriculum for first grade and kindergarten students. After the classroom curriculum is completed, work continues to reinforce the caring behavior of the majority of students who will not tolerate bullying. Teachers are encouraged to reward caring behaviors and to hold weekly classroom meetings to discuss and acknowledge the behaviors exhibited during the previous week. An in-service component provides parents with the same information as the students and staff. Other information is provided through newsletters and follow-up workshops. Individual parents of students involved in bullying perpetration and victimization are also consulted. Complete implementation of BPYS spans three years, the first year being devoted to implementing the full curriculum, and the second and third years involving booster sessions to reinforce the material presented in the first year.

Research Design

The present evaluation used a multiple nonequivalent control group pretest-posttest design to test the effectiveness of BPYS in elementary and middle schools. An attempt (more successful at the elementary school than the middle school level) was made to match treatment and comparison schools at baseline, in order to be able to infer that post-baseline differences between treatment and comparison schools were attributable to the program. Data were analyzed to compare treatment and comparison schools on individual items and composite scales related to the three major components of the program, including bivariate analysis with the treatment-comparison contrast as the predictor and either the individual items or the composite scales as outcomes. In addition to bivariate analyses of the relationship between treatment and outcome, analysis was also done to assess the impact of fidelity of implementation on the outcome, and multivariate analysis was done to assess the impact of treatment in a broader context including intervening variables (peer environment and attitudes toward aggression) which might be expected to mediate the effects of the intervention.

For the item-level analysis, Somers' d (a directional measure of association for ordinal outcomes) and its associated test for statistical significance was used to assess the strength of the relationship between treatment and outcomes. For the analysis of the composite scales, which could be treated as being measured at the interval level of measurement, and also for the analysis of the impact of quality of implementation on the results of the intervention, Pearson's r and its associated test for statistical significance were used. The multivariate analysis was performed using ordinary least squares multiple regression analysis.

Evaluation Results

There was considerable variation in the degree to which the program was faithfully implemented in the elementary schools, and it was not implemented especially well in the middle schools.

The results of the evaluation at the middle school level were inconclusive; but they suggest that the program does no harm and may do some good.

The results of the evaluation at the elementary school level are more persuasive, and they indicate that the program has the intended beneficial effect in reducing bullying behaviors and

school violence more generally, and in changing the attitudes of students toward bullying and school violence.

Where the program was implemented faithfully at the elementary school level, favorable results were quicker to materialize, more pervasive, and more long-lasting than in schools where implementation was weaker; but even where implementation was weaker, there were some positive effects of the program.

The program appears promising as an intervention to reduce bullying and school violence at the elementary school level.

Further research would be needed (and, given the results here, would be appropriate) before it can be concluded that the program demonstrates effectiveness at the middle school level.

TABLE OF CONTENTS

Part 1	Background and Significance	1
	Prevalence and Consequences of Bullying	3
	Risk Factors for Bullying	4
	Identification of Effective Intervention Programs	5
	Implementation Fidelity and Program Effectiveness	6
	The Intervention: Bully-Proofing Your School	7
	Preliminary Studies	10
	Research Question and Hypotheses	11
Part 2	Methods	13
	Schools	13
	Program Stages	15
	Individual Subjects	17
	Data Collection Procedures	21
	Measurement	24
	Analytical Methods	26
	Scaling	28
Part 3	Process Evaluation Results	30
	Aldine Middle School and Beacon Elementary School	32
	Chapman Middle School	37
	Doubleday Elementary	43
	Elsevier Elementary	47
Part 4	Elementary School Outcome Evaluation	51
	Item-Level Results	51
	Program Components and Hypotheses: Composite Scale Outcomes	69
	Quality of Implementation and Overall Impact	72
	Multivariate Analysis of the Impact of BPYS	73
	BPYS and Faculty and Staff Perceptions of School Climate	77
	Conclusion: The Impact of BPYS at the Elementary School Level	78
Part 5	Middle School Outcome Evaluation	79
	Item-Level Results	79
	Program Components and Hypotheses: Composite Scale Outcomes	86
	Quality of Implementation and Overall Impact	88
	Multivariate Analysis of the Impact of BPYS	89
	BPYS and Faculty and Staff Perceptions of School Climate	93
	Conclusion: The Impact of BPYS at the Middle School Level	93
Part 6	Conclusions	94
	Recommendations for Future Research	96
	References	97

Part 1: BACKGROUND AND SIGNIFICANCE

A student is bullied or victimized when he or she is exposed, repeatedly and over time, to negative actions on the part of one or more other students (or others). Bullying is characterized by three criteria: It is (1) aggressive behavior or intentionally inflicting physical or emotional harm, (2) carried out repeatedly over time, and (3) done in the context of an interpersonal relationship characterized by an imbalance of power (Olweus 1993; Olweus et al. 1999). A playground fight between two children of approximately equal strength, or a single incident of violence or harassment, does not constitute bullying, although it is aggressive or violent behavior. Direct bullying is a relatively open attack on a victim, which can be physical or verbal in nature. Indirect bullying, which is more subtle and may be more difficult to detect, may include social isolation, intentional exclusion, making faces, obscene gestures, or manipulating friendship relationships. Table 1, from Garrity et al. (2000b) details the different types of behavior which may constitute bullying when perpetrated repeatedly on a victim markedly less powerful than the perpetrator.

TABLE 1 Bullying Behaviors Chart					
MILD	MODERATE			SEVERE	
	PHYSICAL AGGRESSION				
<ul style="list-style-type: none"> ◆ Pushing ◆ Shoving ◆ Socking 	<ul style="list-style-type: none"> ◆ Kicking ◆ Hitting 	<ul style="list-style-type: none"> ◆ Defacing property ◆ Stealing 	<ul style="list-style-type: none"> ◆ Physical acts that are demeaning and humiliating, but not bodily harmful (e.g., degrading) ◆ Locking in a closed or confined space 	<ul style="list-style-type: none"> ◆ Physical violence against family or friends 	<ul style="list-style-type: none"> ◆ Threatening with a weapon ◆ Inflicting bodily harm
	SOCIAL ALIENATION				
<ul style="list-style-type: none"> ◆ Gossiping ◆ Embarrassing 	<ul style="list-style-type: none"> ◆ Setting up to look foolish ◆ Spreading rumors about 	<ul style="list-style-type: none"> ◆ Ethnic slurs ◆ Setting up to take the blame 	<ul style="list-style-type: none"> ◆ Publicly humiliating (e.g., revealing personal information) ◆ Excluding from group ◆ Social rejection 	<ul style="list-style-type: none"> ◆ Maliciously excluding ◆ Manipulating social order to achieve rejection ◆ Malicious rumormongering 	<ul style="list-style-type: none"> ◆ Threatening with total isolation by peer groups
	VERBAL AGGRESSION				
<ul style="list-style-type: none"> ◆ Mocking ◆ Name calling ◆ Dirty looks ◆ Teasing 	<ul style="list-style-type: none"> ◆ Teasing about clothing or possessions 	<ul style="list-style-type: none"> ◆ Teasing about appearance 	<ul style="list-style-type: none"> ◆ Intimidating telephone calls 	<ul style="list-style-type: none"> ◆ Verbal threats of aggression against property or possessions 	<ul style="list-style-type: none"> ◆ Verbal threats of violence or of inflicting bodily harm
	INTIMIDATION				
<ul style="list-style-type: none"> ◆ Threatening to reveal personal information ◆ Gossip ◆ Publicly challenging to do something 	<ul style="list-style-type: none"> ◆ Defacing property or clothing ◆ Playing a dirty trick 	<ul style="list-style-type: none"> ◆ Taking possessions (e.g., lunch, clothing, toys) 	<ul style="list-style-type: none"> ◆ Extortion ◆ Sexual/racial taunting 	<ul style="list-style-type: none"> ◆ Threats of using coercion against family or friends 	<ul style="list-style-type: none"> ◆ Coercion ◆ Threatening with a weapon

The behaviors represented in Table 1 range from relatively mild to relatively severe. Both are of concern in the context of prevention of bullying. Research on illegal behavior generally and on violence in particular indicates that relatively minor illegal or violent behaviors tend to be initiated prior to more serious forms of violence and other illegal behavior, and the less serious behaviors may even be prerequisite to the more serious behaviors for the vast majority of individuals (e.g., Elliott 1994; Elliott et al. 1989), so prevention of even relatively mild forms of aggression and violence may forestall escalation to more severe forms. Also

evident from Table 1 is that it is not the behavior by itself that distinguishes bullying from other forms of behavior. Indeed, the content of the behavior is identical, and it is the repetitious nature of the behavior and the power differential between the actors that sets bullying off from other forms of behavior. The differences between bullying and normal peer conflict are summarized in Table 2, also taken from Garrity et al. (2000b). Bullying characteristically involves greater social and emotional distance (including imbalance of power), greater repetitiveness and seriousness of threat or harm, and greater purposefulness on the part of the offender.

TABLE 2 Recognizing the Difference

Normal Peer Conflict	Bullying
Equal power between friends	Imbalance of power between friends
Individuals often play together	Individuals rarely play together
Happens occasionally	Repeated negative actions
Accidental	Purposeful
Not serious	Serious with threat of physical or emotional harm
Equal emotional reactions	Strong emotional reaction from victim and little or no emotional reaction from bully
Not seeking power or attention	Seeking power, control, or material things
Not trying to get something	Attempt to gain material things or power
Remorse—will take responsibility	No remorse—blames victim
Effort to solve the problem	No effort to solve the problem

There are two implications that follow from the observation that bullying does not represent a unique behavior or set of behaviors, but rather a set of behaviors which can also occur outside the context of bullying. First, from the perspective of injury prevention, minor forms of bullying may be regarded as an indirect risk factor, and more serious forms of bullying may be regarded as a direct risk factor, for minor and serious injury and, in extreme cases, death, including homicide and suicide (see the review pp. 3-4 below). Second, successful prevention of bullying will necessarily have an impact on the more general forms of the behaviors in question. This follows not only logically but also empirically, based on findings from previous implementations of another similar anti-bullying program. Olweus et al. (1999:19-20) found that in addition to reducing bullying behavior, the program also resulted in reductions in other problem behaviors, including vandalism, fighting, theft, truancy, and disciplinary problems. In the following discussion, therefore, there is a dual concern, both specifically with bullying, and also more broadly with victimization and perpetration involving aggression and violence, given the expectation that any intervention successful in reducing bullying should also have an impact on the *contextually less specific* but *behaviorally more specific* problems of violent victimization and perpetration. In other words, there is substantial overlap between, on the one hand, bullying, and on the other hand, aggression and violence, particularly minor forms of violence. One cannot expect that a program that has an impact on bullying will have no impact on other forms

of violence and aggression, or that a program that has an impact on violence and aggression will have no impact on bullying; nor would it be particularly desirable to have an impact on one but somehow deliberately leave the other undiminished.

Prevalence and Consequences of Bullying

Garofalo et al. (1987:321) found that about half of all adolescent victimizations were school-related, and suggested that most school related victimizations among adolescents “appear to consist primarily of bullying, injured pride, and misguided mischief.” Kaufman et al. (2000) reported that 10% of students in 6th and 7th grades reported being bullied at school during the previous six months. The School Crime Supplement to the National Crime Victimization Survey (NCVS) reported similar prevalence of bullying for 12-18 year olds in 2001, with 8% reporting having been bullied in the past 6 months, with higher prevalence of bullying for younger students than for older students: 14% among sixth graders, 9% among ninth graders, but only 2% among twelfth graders (Devoe et al. 2003; see also DeVoe et al. 2004 and subsequent years of the School Crime Supplement to the NCVS). Cross-national data reported by Olweus (1993) suggests that as many as one in seven students may be victims of bullying during the school year, and other studies report even higher rates of victimization by bullying over the course of the elementary or middle school career (Batsche and Knoff 1994; Hoover et al. 1992). Garrity et al. (2000a) suggest that bullying occurs once every seven minutes on elementary school playgrounds. Given the wide range of behaviors defined as bullying, these numbers may not appear surprising. As noted by Lawrence (2007), bullying has come to be recognized as one of the most serious problems in schools, with consequences affecting victims for months and years after victimization.

Like other forms of criminal victimization, bullying may have one or more of four important impacts on its victims: (a) physical or medical, (b) financial, (c) cognitive or emotional, and (d) behavioral. Physical harm refers specifically to bodily injury or death. Financial costs include costs associated with physical harm (e.g., costs of medical treatment), and also direct financial losses resulting from property theft or damage. Cognitive and emotional costs include subjective emotional pain and suffering, and are likely to be manifested in the form of mental health problems such as depression, anxiety, and post-traumatic stress disorder (PTSD). Behavioral impacts refer to voluntary and involuntary behavioral consequences of victimization, some of which may also be related to mental health diagnoses. Behavioral consequences can include subsequent victimization, perpetration of criminal acts of one's own, and problem use of alcohol or illicit drugs. In addition to the direct impacts of violent victimization on victims and offenders, there may also be indirect consequences to the family, friends, acquaintances, and even medical and mental health professionals who know or know about the individual or the victimization incident (Ruback and Weinberg 2001). In the present context of evaluating BPYS, the focus is on the direct consequences to the victims and the perpetrators.

Limited evidence, much of it anecdotal or from case studies, suggests that bullying has an impact on both the perpetrators and the victims. For the victims, consequences include painful and humiliating experiences that can cause young victims to be unhappy, distressed, and

confused; loss of self-esteem; anxiety and insecurity; negative effects on concentration and learning in school; refusal to attend school or avoidance of school or specific places at school; feelings of stupidity, shame, unattractiveness, and failure; psychosomatic symptoms such as stomach aches and headaches; depression; physical injury; perpetration of violent behavior; and, at the extreme, suicide or perpetration of homicide (Fried and Fried 1996; Kaufman et al. 2000; Olweus 1992; Olweus et al. 1999; O'Moore and Kirkham 2001; Rigby 1998). Perpetrators of bullying are more likely to engage in other antisocial/delinquent behavior (e.g., vandalism, shoplifting, truancy, and drug use) into adulthood, to be convicted of crimes by age 24, and to engage in serious violence during adolescence and adulthood (Farrington 1993; Farrington 1995; Olweus et al. 1999). In addition to its impacts on the individual level, bullying also affects the school climate more generally. Students tend to feel less safe and are less satisfied with school life in schools where bully/victim problems occur, although there is some question about whether bullying is the cause, the effect, or both, with respect to school-related stress and alienation (Natvig et al. 2001). In schools where bully/victim problems are ignored, students may start to regard bullying behavior as acceptable. This may result in more bullying behavior as well as other, possibly more severe, problems (Olweus et al. 1999).

When bullying involves actual or attempted violence, consequences can be severe and long-lasting, as indicated by information on the consequences of victimization, particularly violent victimization, in adolescence and adulthood. The adverse impacts of victimization, particularly violent victimization in adolescence, are pervasive, severe, and sometimes enduring, consisting not only of physical injury, financial loss, and emotional distress, but also including elevated risks of subsequent victimization (which may result in further injury and also exacerbate the emotional distress from earlier victimizations), problem substance use, and criminal behavior, a cost which goes beyond the initial victim of crime to new victims, who in turn may perpetuate the cycle of harm and personal suffering (Berton and Stabb 1996; Blumberg 1979; Boney-McCoy and Finkelhor 1995; Bureau of Justice Statistics 1994; Kilpatrick et al. 1987; Klaus 1994; Laub 1997; Lurigio 1987; Menard 2000; Menard 2002; Miller et al. 1996; Norris et al. 1997; Resick and Nishith 1997; Simon et al. 2001). To the extent that any program has an impact on bullying, it should, based on both logic and past empirical research, also have an impact on violent victimization, which, as noted earlier, is behaviorally more specific but contextually broader than bullying. As a consequence, a successful intervention to reduce bullying should directly result in decreased injury, correspondingly result in decreased costs associated with treating injury, and also at least indirectly reduce the risks of future violent victimization, future violent offending, and future substance use and mental health problems, thus further reducing injury and the costs associated with injury from violence.

Risk Factors for Bullying

Bullying takes place in the classroom, on the playground, in hallways, in gyms, in locker rooms, and in bathrooms. Bullying is two to three times more likely to occur at school as on the way to and from school (Olweus 1993; Olweus et al. 1999). There are individual, familial, peer, and school factors that can place a youth at risk for participating in bullying behavior. Generally, boys are much more likely to engage in bullying behavior than girls. Girls who bully are less likely to be physically abusive than boys are. Although most bullying occurs between students in

the same grade, older students sometimes bully younger students. Individual risk factors include impulsivity, short temper, dominant personality lacking empathy, difficulty conforming to rules and low frustration tolerance, positive attitudes toward violence, physical aggressiveness, and gradually decreasing interest in school. Family risk factors include lack of parental warmth and involvement, overly permissive or excessively harsh discipline/physical punishment by parents, and lack of parental supervision. In the peer group, friends/peers with positive attitudes toward violence and exposure to models of bullying constitute risk factors. Risk factors in the school context include lack of supervision during breaks (e.g., lunchrooms, playgrounds, hallways, locker rooms, and bathrooms), unsupervised interactions between different grade levels during breaks, indifferent or accepting teacher and/or student attitudes toward bullying, and inconsistent enforcement of the rules

There are also individual, familial, peer, and school factors that can place a youth at risk for being bullied. Both boys and girls are most likely to be victimized by boys. Younger and weaker students are most likely to be bullied. Individual risk factors include a cautious, sensitive, insecure personality, difficulty in asserting oneself among peers, and physical weakness (particularly in boys). Other risk factors may include over-protection by parents in the family context, lack of close friends, and the same constellation of school-based risk factors described in the previous paragraph. To counter the risk of bullying, several steps have been proposed, including (1) Awareness and warm, positive involvement of adults (e.g., teachers, principals, school counselors, parents); (2) Setting and maintaining firm limits regarding what behavior is unacceptable (i.e., Bullying is not accepted in our school); (3) Consistent application of non-hostile, nonphysical negative consequences for rule violation and unacceptable behavior; and (4) Encouraging adults to act as authorities and positive role models in students' academic learning and social relationships in school. These principles have been implemented in one model program for bullying prevention (Olweus 1993; Olweus et al. 1999), and are also included in the BPYS program which is currently being evaluated.

Identification of Effective Intervention Programs

In the early 1970s, evaluations of existing rehabilitation and prevention programs produced the pessimistic conclusion that few if any programs could be demonstrated to be effective according to scientific criteria, leading to the (at least slightly overstated) conclusion that *nothing* works in intervention to reduce or prevent crime and delinquency (e.g., Lipton et al. 1975; Martinson 1974). In the past decade, however, this pessimistic view has given way to a focus on what *does* work in the prevention of violence, illicit substance use, and other criminal behavior. Lists of “successful” and “promising” interventions include Montgomery et al. (1994), National Institute on Drug Abuse (1997), Sherman et al. (1997), and Waller et al. (1979). One of the most demanding and rigorous attempts to identify effective prevention programs has been the Blueprints for Violence Prevention project (Mihalic et al. 2001), begun in 1996 at the Center for the Study and Prevention of Violence (CSPV) at the University of Colorado at Boulder, working with the Colorado Division of Criminal Justice (CDCJ). The objective was to identify truly outstanding programs and to describe these interventions in a series of “Blueprints.” Each Blueprint describes the theoretical rationale for the intervention, the core components of the program as implemented, the evaluation design findings, and the practical experiences the

program staff encountered while implementing the program at multiple sites. The Blueprints are designed to be very practical descriptions of effective programs which allow states, communities, and individual agencies to (1) determine the appropriateness of each intervention for their state, community, or agency; (2) provide a realistic cost estimate for each intervention; (3) provide an assessment of the organizational capacity required to ensure its successful start-up and operation over time; and (4) give some indication of the potential barriers and obstacles that might be encountered when attempting to implement each type of intervention.

The evaluation standards established for the selection of the programs were: (1) an experimental design or a strong quasi-experimental design, (2) evidence of a statistically significant or marginal prevention or deterrent effect, (3) replication at multiple sites with demonstrated effects, and (4) evidence that the prevention effect was sustained for at least one year after treatment. This set of selection criteria establishes a very high standard, one that proved difficult to meet, but it reflects the level of confidence necessary if it is to be recommended that communities adopt these programs with reasonable assurance that they will prevent violence. Given the high standards set for program selection, the burden for communities mounting an expensive evaluation to demonstrate their effectiveness is removed; this claim can be made as long as the program is implemented well. Documenting that a program is implemented well is relatively inexpensive but critical to the claim that the program is effective.

Programs reviewed for the Blueprints project were classified into three categories. Eleven programs were classified as exemplary or Blueprint programs because they met or came close enough to meeting all the criteria (strong research design, demonstrable effect, multiple site replication, and sustained effects) for inclusion as exemplary programs. Other programs were classified as promising because they met some but not all of the criteria. Most commonly, programs in the promising category demonstrated evidence of some prevention effect in a strong or fairly strong research design, but multiple site replication or evidence of sustained effects were absent. The third category, other programs, consists of programs for which none of the four criteria has been adequately addressed. Both promising and other programs may, in fact, be effective in preventing or reducing violence, but their effectiveness has not yet been demonstrated through sound evaluation research. At present, the BPYS program is classified as an “other” program. A goal of the proposed research is to evaluate BPYS with respect to these criteria. The evaluation, if BPYS demonstrates the required prevention effect, would allow us to promote the program as a promising program or, with replication elsewhere, an exemplary or model program.

Implementation Fidelity and Program Effectiveness

Even the implementation of effective anti-bullying programs is unlikely to affect the incidence bullying in schools unless careful attention is given to the degree to which a program is delivered as it was designed. Programs must be implemented with fidelity to the original model that was evaluated in order to preserve the behavior change mechanisms that made the original model effective (Mihalic, 2001). As suggested in Mihalic (2001), in order for a program to be implemented with fidelity, it is crucial that *all core components* of the program be provided

at the intended dosage. This is important because when research studies move from the original trials where they are well controlled by the program designer to implementation in less controlled naturalistic settings, the tendency for key program components to be “watered down” increases. Evaluations of such implementations that yield negative results may incorrectly conclude that the program does not work, when instead what was evaluated was not the program as it was designed. Alternatively, changes in a evaluated program that are specific to a particular implementation may yield the false conclusion that the program *as designed* is effective, when it was the program with modifications that was effective.

To underscore the importance of implementation fidelity, Mihalic (2001) describes several studies that conclude that analysis of fidelity of implementation (which can be part of a process evaluation) has shown consistently stronger outcomes when programs are implemented with fidelity. First, a meta-analysis study of 200 intervention programs (Lipsey, 1999) indicated that implementation effects were larger when attention was given to program implementation. Second, an evaluation of the Life Skills Training program (Botvin et al. 1995) compared results from a high fidelity sample to the full sample, and while both showed significant improvement, the high fidelity sample had significantly better results.

Mihalic (2001) further argues that of greatest concern is that some programs may only show significant effects in the high fidelity samples. Citing an evaluation of the Child Development Program (Battistich et al. 2000), it is noted that the program was implemented in 12 schools, and that the overall results would have resulted in the conclusion that the program did not work. However, the program was conducted with high fidelity at only five schools, and results were significant and positive for the students at those schools.

The Intervention: Bully-Proofing Your School

Bully-Proofing Your School (BPYS) is designed as a comprehensive, school-based intervention with three major components: (1) heightened awareness of the problem of bullying, involving a questionnaire to assess the extent of bullying in the school, and the creation of classroom expectations and rules regarding no tolerance for bullying; (2) teaching protective skills for dealing with bullying, resistance to victimization, and providing assistance to potential victims of bullying; and (3) creation of a positive school climate through promotion of a “caring majority” in the school which works to alter the behavior of bystanders. As part of the first component, all members of the school community, adults and students, commit themselves to a nontolerance policy about bullying, and to creating a caring community. School rules and expectations are established that are understood and enforced throughout the community. All systems in the school are addressed, from administration to transportation, and specific steps for implementing the school-wide program are included.

The second component of the program teaches skills and strategies to help individuals avoid being victimized. Figure 2, from Garrity et al. (2000b), is a presentation of the individual strategies that can be used to avoid victimization by bullying, and which are applicable to avoidance of interpersonal violent victimization more generally. The mnemonic HA HA SO is used to stand for the six strategies: (1) Help; (2) Assert yourself; (3) Humor; (4) Avoid; (5) Self

talk; and (6) Own. Different coping strategies may be appropriate to different situations, and students are taught different skills they can use, depending on the situation, and in which situations the skills are appropriate. Help means knowing when and how to get help from others. Assert yourself means knowing when to stand up to a bully, and also when not to, e.g., in instances of severe bullying or high risk of injury. Humor means trying to turn a difficult situation into a funny one, a surprise strategy, which may be difficult for a frightened child, but which may be learned with practice. Avoid means knowing how and when to walk away, the “how” referring to combining disengagement with self-assertion. Self talk is the practice of thinking positively about oneself even when one is being “put down” by someone else. Owning the insult combines agreement with the bully and making light of the insult, a strategy which may be appropriate for insults to appearance (e.g., clothing or hair style), but may be inappropriate for insults based on gender, ethnicity, disability, religion, or heritage. The HA HA SO strategies are appropriate as individual responses to bullying or the threat of violence.

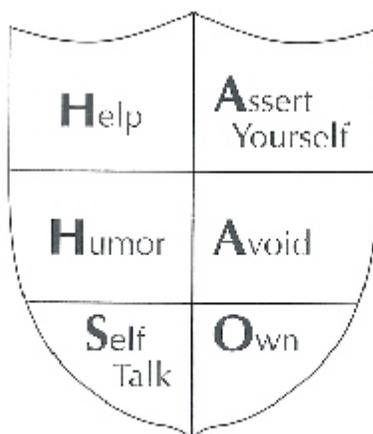


Figure 2: HA HA SO: Individual Strategies to Prevent Bullying

In the third component of the program, the focus is on climate change, creating a positive, prosocial school climate that feels safe and secure for all members of the school community. This requires broad efforts to change the overall school culture, rather than only focusing on specific individual skills. Thus, in addition to teaching specific individual and group level skills, BPYS focuses on the 85% of students in school who are neither bullies nor victims, but who are in the role of bystanders. These are the students who generally have well-developed prosocial skills, but do not know how to or are afraid to intervene to prevent bullying. Strategies available to assist victims or potential victims in the broader group context are illustrated in Table 3, also from Garrity et al. (2000b). These strategies include not joining in, getting adult help, mobilizing the group, taking an individual stand, and befriending the victim. By emphasizing not only personal but also group responses to bullying, the program works to develop a caring majority to help prevent bullying and to respond in a way that empowers the would-be victims and disenfranchise the would-be bullies.

As part of the BPYS program, teachers are given information and strategies which help them to recognize bullying and intervene appropriately in bullying situations. The intervention includes a classroom curriculum, consisting of seven sessions, with two optional sessions on conflict resolution and diversity. This curriculum is taught by the classroom teacher or mental health staff at the school, once a week, from thirty to forty-five minutes per session, depending on the age of the children. There is an abbreviated three-session curriculum for first grade and kindergarten students. After the classroom curriculum is completed, work continues to reinforce the caring behavior of the majority of students who will not tolerate bullying. Teachers are encouraged to reward caring behaviors and to hold weekly classroom meetings to discuss and acknowledge the behaviors exhibited during the previous week. An in-service component provides parents with the same information as the students and staff. Other information is provided through newsletters and follow-up workshops. Individual parents of students involved in bullying perpetration and victimization are also consulted. The curriculum component of BPYS with its associated instructional materials, including materials for parents (Garrity, Jens, et al. 2000; Garrity et al. 2000a; Garrity et al. 2000b; Bonds and Stoker 2000), is one major feature that distinguishes it from Olweus' bullying prevention intervention. The thoroughness of the coverage of the BPYS curriculum was noted by Fried and Fried (1996:162), who characterized it as "the most complete curriculum contained in one book that we have been able to identify." The existence of this specific and detailed curriculum has been one reason for the greater interest of some Colorado school districts in BPYS as opposed to alternative interventions, including the Olweus program. Complete implementation of BPYS spans three years, the first year being devoted to implementing the full curriculum, and the second and third years involving booster sessions to reinforce the material presented in the first year.

Preliminary Studies

There has been extensive research on programs other than BPYS, and that research suggests that antibullying programs can reduce bullying and other problem behavior, although results may be different for different programs (Olweus 1993; Olweus et al. 1999; Smith and Sharp 1994; Stevens et al. 2001). Stevens et al. (2001), in particular, found that a Flemish antibullying program was more effective at the primary school than at the secondary school

level, and suggested that antibullying programs are more developmentally appropriate to primary school (i.e., to elementary and middle school as opposed to high school). Olweus' (1999) anti-bullying program is one of the few programs to date to meet the Blueprints criteria for a model program, and as noted earlier, BPYS incorporates elements of the Olweus program and adds additional elements of its own, suggesting that it should be at least as successful as the Olweus program, if properly implemented.

BPYS itself was implemented at one suburban elementary school, with a predominantly (95%) majority (white) population, in Englewood, Colorado, in 1995, with students in kindergarten through fifth grade. The full curriculum was implemented in the first year, and during the second year, continued classroom sessions were provided in the form of booster sessions (three sessions for kindergartners, five sessions for first graders, and a three session review for grades 2-5). During the third year, kindergartners again received the three sessions, first graders five sessions, and second through fifth graders the three session review. There was no comparison school in the study. The bullying survey was administered in the fall of 1995 (n=351), spring 1996 (n=339), and to grades 2-5 in spring 1997 (n=328), 1998 (n=345), and 1999 (n=367). The results indicated that there was a statistically significant improvement in the sense of safety at school and on the way to and from school, and a reduction in bullying behaviors over time. In addition, analyses were conducted comparing assessments of school safety in the school which had implemented BPYS to data from 3,223 third through fifth graders from 17 elementary schools in suburban Colorado. Results were a mixture of no differences or else differences favoring the program school (i.e., greater feelings of safety in the program school), depending on the year of the comparison and the specific context (classroom, hallways/lunchroom, playground, going to and from school) being compared. Broadly, there were no pretest differences in the program school and the subsequently added comparison schools, but there appear to have been improvements in perceived safety experienced in the program school which were statistically significantly above and beyond those experienced in the other schools, and no significant decreases in perceived safety in the program school.

Research Question and Hypotheses

The overarching research question for this evaluation is, obviously, whether BPYS can be effectively implemented and, if it is effectively implemented, whether it is effective in reducing bullying and related aggressive and violent behaviors in the school context. Based on previous research on this and related programs, we can offer three hypotheses, based on the three program goals.

Hypothesis 1: Compared to students in schools in which BPYS is not implemented, students in schools in which BPYS is implemented will perceive greater intolerance of bullying. The first major component of the program involves (via heightened awareness of bullying) the creation of classroom expectations and rules regarding no tolerance for bullying. Indicators relevant to measuring success on this objective will focus on perceptions by students (the target population) that bullying is discouraged in the school context.

Hypothesis 2: Compared to students in schools in which BPYS is not implemented, students in schools in which BPYS is implemented will report lower rates of (a) victimization by, (b) perpetration of, and (c) witnessing physical nonphysical aggression. This should follow from the second major component of the program in which protective skills are taught to help individuals avoid being victimized. Indicators relevant to measuring success on this objective will focus on self-reported victimization by, perpetration of, and witnessing violence, threats, relational aggression, and students picking on and being picked on by other students. Both items specifically addressing the repetitive nature of bullying and items representing physical and nonphysical aggression more generally are included as central to this hypothesis, consistent with previous research on the effectiveness of anti-bullying programs (e.g., Olweus et al. 1999).

Hypothesis 3: Compared to students in schools in which BPYS is not implemented, students in schools in which BPYS is implemented will report higher rates of feeling safe and lower rates of feeling unsafe at school. This should follow from the third major component of the program, creation of a positive school climate that feels safe and secure. Indicators relevant to measuring success on this objective will consist primarily of items on students' perceptions of school safety.

The three hypotheses above focus on the core objectives of BPYS with respect specifically to bullying and related physical and nonphysical aggression. It may be expected, however, that (a) in order to reduce bullying in particular or physical and nonphysical aggression more generally, more generally favorable conditions may need to be created with respect to school climate and individual behavior, and that (b) successful reduction of bullying in particular or of physical and nonphysical aggression more generally may have spillover effects beyond those directly involving bullying and physical or nonphysical aggression. For example, students in schools where BPYS is implemented should be more likely than students in schools where BPYS is not implemented to perceive that rules are clear and discipline is fair in general. Also, students in schools where BPYS is implemented should be more likely than students in schools where BPYS is not implemented to perceive the school climate as generally more favorable. These issues, unlike the three hypotheses described above, are not directly related to the assessment of the effectiveness of the intervention; they will not be examined in detail, but will be considered in the item-level analysis in the results section.

Another consideration is the possibility that BPYS affects bullying and related aggressive behaviors not only directly, but also by influencing risk factors for aggression and violence. Two of the most consistent predictors of violence, aggression, and other forms of delinquency are association with delinquent friends and one's own attitudes toward delinquency, and it is expected that the impact of the program may be mediated at least in part through these two variables. Moreover, improvement in either or both of these variables, given their well-established relationships with delinquency in general and aggression and violence more specifically, would constitute at least indirect evidence of program effectiveness (insofar as the program reduces one or more risk factors for violence). We expect that students in schools where BPYS is implemented should be more likely than students in schools where BPYS is not implemented to report (a) that their own attitudes are less favorable to physical and nonphysical aggression, and (b) that the attitudes of their friends are less favorable to physical and

nonphysical aggression; and additionally, based on past research, that the peer environment and students' own attitudes will have a direct impact on bullying and related aggressive behaviors in the school context. These concerns will be addressed by examining peer environment and attitudes toward aggression and violence as outcomes of the intervention, and also by examining the impacts of BPYS on peer environment and attitudes, and the impact of peer environment and attitudes on bullying and related aggressive behaviors, in a multivariate analysis that includes these variables along with sociodemographic background variables and family bonding as controls.

Part 2: METHODS

The evaluation uses a multiple nonequivalent control group pretest-posttest design (Riecken and Boruch et al. 1974:110) with *ex ante* selection of treatment and comparison groups (Rossi et al. 1999). As discussed in Campbell and Stanley (1963), Cook and Campbell (1979), Reichardt and Mark (1999), and elsewhere, the nonequivalent control group design with both pretest and posttest is one of the most common designs in evaluation research. In the nonequivalent control group design with only one treatment and one comparison group, the two principal questions that arise regarding internal validity are the possibility of regression effects, particularly if one group has been selected for its extreme scores on the variable being tested at the pretest and posttest, and the possibility that the selection process interacts with maturation or testing. Potential threats to external validity include interaction of testing and treatment, interaction of selection and treatment, and reactive arrangements (experimenter effects). The credibility of the observed results, or of alternative explanations involving threats to internal or external validity, depends on the pattern of the results themselves when there are only two groups (Cook and Campbell 1979).

A distinction needs to be made, however, between the nonequivalent control group design involving only two natural groups (e.g., two schools) and the stronger design in which experimental and comparison groups are made up of multiple natural groups (e.g., several schools). As Riecken and Boruch et al. (1974:110) succinctly observe, “In the two-group case, the cause of an apparent effect is ambiguous because there were undoubtedly many differences between the two schools over and above the presence of the treatment. Any of these differences could have produced the differential gains. In the multiple-group version at its best, there are likely to be few differences except the experimental treatment that would operate systematically in the same direction to differentiate the experimental from the comparison schools.” This is because random effects such as regression to the mean or other selection-related effects are likely to cancel one another out when there are multiple comparisons instead of only one. Systematic effects associated with selection, however, remain a threat to the validity of the results (e.g., if all of the treatment schools were selected to have lower test scores than the comparison schools). With multiple independent treatment and comparison groups, pre-existing differences between treatment and comparison groups are no longer inextricably confounded with the treatment/comparison distinction itself, insofar as differences between treatment and comparison schools are not the same but vary across treatment/comparison pairs or groups of schools.

Schools

As part of a broader initiative, the Safe Communities ~ Safe Schools (SCSS) study (Delbert Elliott, principal investigator), conducted through the Center for the Study and Prevention of Violence (CSPV) at the University of Colorado in Boulder, schools in the state of Colorado were offered the opportunity to participate in implementation and testing of a range of school programs designed to meet specific needs of the respective schools. Bully-Proofing Your School (BPYS) was one of several such programs. Initially, elementary and middle schools interested in implementing an anti-bullying program were identified. Interest in implementation

was considered crucial in order to secure staff cooperation and increase the likelihood of implementation fidelity. Once the schools interested in implementing BPYS were identified, potential comparison schools from the SCSS study were identified and compared with the prospective treatment schools. From among those schools indicating a willingness to participate, comparison schools were selected to match treatment schools as closely as possible on grade levels, the sociodemographic characteristics of the schools (percent majority and minority ethnic groups, percent eligible for free and reduced school lunch), and average student standardized test scores (Colorado Student Assessment Program, or CSAP, tests). Preference was also given to matching schools for similarity of location (i.e., urban with urban and rural with rural schools), but since a given school district may have only a single elementary or middle school, matches within the school district or the geographic area were not always possible. Schools were not matched on the outcome variables (violence and other problem behaviors) because the available measures were unreliable (such as official disciplinary referrals, which may reflect official policy as much as or more than actual behavior; see, for example, Menard 1987; Menard and Covey 1988; O'Brien 1985). Overall, the schools selected as treatment and comparison schools represent a range from one-third to two-thirds eligible for free or reduced lunch; from less than one-third to over half minority/Hispanic; and all have low (in one treatment school) to average test scores.

A total of two treatment middle schools plus one comparison middle school, and three treatment elementary schools plus three comparison elementary schools, were selected for evaluation. (Originally another treatment middle school, another treatment elementary school, and another comparison middle school had been planned for inclusion, but dropped out of the study early in the evaluation.) Five comparison schools, consisting of four elementary schools and one middle school, were also selected. For all ten schools, approval was received from the school districts, principals, and teachers involved, and signed memoranda of understanding were received from each of the schools in the study. Treatment schools agreed to a process of data collection and planning, intervention, and evaluation. Comparison schools agreed only to data collection, parallel to the data collection in the treatment schools. Initially, only nine schools were matched together, five treatment to four comparison, because one of the comparison elementary schools was initially selected to match a potential treatment school (mentioned above) that ultimately decided not to participate. The decision was made to continue to collect data from the “extra” comparison school because it had similar characteristics to the other comparison schools, and the data collected from that school could be used if one of the comparison schools dropped out of the study. This decision proved fortunate, as one of the comparison schools did drop out after the third year of data collection, and so the original “extra” comparison school’s five years of data were substituted. Though some of the initial sociodemographic characteristics varied between the original comparison school and its replacement, as will be noted later in this report, baseline survey data indicated that the new comparison resulted in a satisfactory match.

The following table (Table 4) presents the characteristics of the treatment and comparison schools with respect to percent on free or reduced lunch, percent minority (of which the predominant minority is Hispanic), and test scores. Pseudonyms are used for all of the schools. It is evident from the table that some matches are closer than others. For the middle

schools, Treatment 1 and Treatment 2 have lower percentages minority and Hispanic than their control, are matched on test scores, and although not matched well individually on free and reduced lunch eligibility, the differences for the two pairs are in the opposite direction (one treatment school with equal and one with higher eligibility for free/reduced lunch). Additionally the three middle schools are all located in the same region in Colorado.

For the elementary schools, the treatment schools match fairly well with the comparison schools on percentages eligible for free or reduced lunch. Both comparison schools have much higher percentages of minority (predominantly Hispanic) students. Only Grades Three and older were surveyed in the elementary schools due to the need for the students to be capable of completing a pencil and paper survey with all items read aloud by a trained researcher.¹

Table 4: Treatment and Comparison Schools

School	Percent Free and Reduced Lunch	Percent Minority/ Percent Hispanic	Test Scores (CSAP)*	Grades
Middle Schools:				
Treatment MS 1 (Aldine)	68	40/35	Average	6-8
Treatment MS 2 (Chapman)	57	44/43	Low	6-8
Comparison MS 1 (Fawcett)	57	62/57	Average	6-8
Elementary Schools:				
Treatment ES 1 (Beacon)	65	43/40	Average	1-5
Treatment ES 2 (Doubleday)	59	45/45	Average	2-5
Comparison ES 1 (Guilford)**	65	81/80	Low	1-5
Comparison ES 2 (Harcourt)	65	68/62	Average	2-5
Treatment ES 3 (Elsevier)	32	28/18	Average	1-6
Comparison ES 3 (Ingram)**	33	41/38	Average	1-5
* Test scores are summarized by CSAP as Unsatisfactory, Low, Average, High, or Excellent.				
**Comparison ES 1 is the school that was added back in after the original comparison ES 3 dropped. The original comparison ES 1 became comparison ES 3.				

Program Stages

The evaluation of BPYS took place over five years. In Year One, the 2001-2002 school year, the treatment and comparison schools were identified in the fall, and baseline student and staff outcome evaluation surveys were conducted with the schools during the spring semester. Concurrently, the teachers in the treatment schools were to begin receiving training, but did not

¹Surveys completed by the sixth graders in the only elementary school in the study that had Grade Six in the school were not used in the analyses contained in this report. All other sixth graders in this study were students in a different context, middle school, and received the more advanced middle school survey.

yet begin any program implementation. In Years Two, Three, and Four, 2002-2005, the treatment schools implemented BPYS. During those three school years, process evaluations designed to measure fidelity of program implementation were conducted with the treatment schools during each semester. Further, outcome evaluation surveys were completed during the spring semesters in the treatment and comparison schools. Year Five, the 2005-2006 school year, was the post-implementation year. The treatment schools could still deliver the program if they desired, but without technical assistance or feedback from BPYS or CSPV staff. The purpose of the one-year followup survey is to see whether any improvements resulting from the implementation could be sustained by the school staff once the BPYS staff have completed their training, or whether it appears that favorable effects only occur during the period in which BPYS staff are actively involved in the schools. An overview of the schedule of program implementation and evaluation follows in Tables 5 and 6 for Treatment and Comparison schools, respectively.

Table 5: Treatment Schools

Year One		Year Two		Year Three		Year Four		Year Five	
Fall 2001	Spring 2002	Fall 2002	Spring 2003	Fall 2003	Spring 2004	Fall 2004	Spring 2005	Fall 2005	Spring 2006
Staff Training Only: No Program Implementation		Implementation with Technical Assistance Year 1		Implementation with Technical Assistance Year 2		Implementation with Technical Assistance Year 3		Follow up Year: No Technical Assistance	
		Implementation Fidelity Ratings:		Implementation Fidelity Ratings:		Implementation Fidelity Ratings:			
		CSPV & BPYS	CSPV & BPYS	CSPV & BPYS	CSPV & BPYS	CSPV & BPYS	CSPV & BPYS		
		Process Evaluation Surveys		Process Evaluation Surveys		Process Evaluation Surveys			
		Teachers and Cadres	Teachers and Cadres	Teachers and Cadres	Teachers and Cadres	Teachers and Cadres	Teachers and Cadres		
	School Climate Surveys: Students and Teachers		School Climate Surveys: Students and Teachers		School Climate Surveys: Students and Teachers		School Climate Surveys: Students and Teachers		School Climate Surveys: Students and Teachers

Table 6: Comparison Schools

Year One		Year Two		Year Three		Year Four		Year Five	
Fall 2001	Spring 2002	Fall 2002	Spring 2003	Fall 2003	Spring 2004	Fall 2004	Spring 2005	Fall 2005	Spring 2006
No BPYS Implementation									
	School Climate Surveys: Students and Teachers		School Climate Surveys: Students and Teachers		School Climate Surveys: Students and Teachers		School Climate Surveys: Students and Teachers		School Climate Surveys: Students and Teachers

Individual Subjects

All students in each of the third- through fifth-grade classrooms in the seven elementary schools and all students in the sixth- through eighth-grade classrooms in the three middle schools were invited to participate in the study each of the five years of survey data collection. The research design required active parental consent for participation in the survey. Active parental consent means that parents were requested to fill out a form indicating their willingness to have their child participate in the survey. If a parent said no, or if a parent did not return the form, the child was not included in the study. The alternative to active consent is passive consent, in which a parent is informed that the child will be included in the study unless the parent specifically withdraws the child from participation, in which case individuals would be included in the study if the parent gave permission or did not return the form, but not if the parent returned the form withdrawing the child from participation. Active parental consent provides a heightened level of protection for human subjects, while passive consent typically results in higher response rates. As is often the case in research requiring active parental consent as opposed to passive consent, response rates can be expected to be lower than would be obtained using passive consent procedures. There is also a fairly extensive literature indicating that in cross-sectional research, active consent results biased samples and tends to underestimate the extent of bullying and other problem behavior (Anderman et al. 1995; Bifulco 2002; Ellickson and Hawes 1989; Esbensen et al. 1996; Henry et al. 2002; Kearney et al. 1983; Severson and Biglan 1989; Unger et al. 2004). Past research on obtaining consent indicates that where active consent is required, the fact that a form is not returned more often signifies nonparticipation (neglecting or forgetting to return the form) than a denial of consent, and that when efforts are successful to secure a high rate of return for the consent forms, not only does the response rate for the survey increase, but in addition minority children, children with lower grades in school, and lower income children are better represented in the sample (Anderman et al. 1995; Esbensen et al. 1997).

In the context of the proposed research, there are three mitigating factors that reduce this bias. First, as noted in Henry et al. (2002), the greatest difference is not between parents who consent and parents who actively refuse, but rather between parents who do not respond (giving neither written consent nor written refusal) compared to parents who turn in the consent form, regardless of whether consent is provided or refused. In the present study, rates of return for the consent forms is typically around 90%. Secondly, as long as such high response rates are obtained (around three-fourths of the sample), type of parental permission does not affect self-reported prevalence of risk behaviors (Eaton et al. 2004), and participation rates among students in the present study are typically around 72%. Thirdly, the use of repeated measures (for three years of implementation plus one year post-implementation in all schools, plus an additional baseline measurement for the treatment schools) mitigates the problem of sample selection bias. Insofar as the bias is systematic, occurring in the sample each year, the change measures within schools should provide valid indicators of change. Alternatively, to the extent that the bias changes from year to year, it should behave like random error, and merely attenuate the apparent relationship between treatment and outcome. The principal results of the biased sampling that results from active parental consent, then are mitigated by high response rates and by the use of repeated measures to assess change over time.

Table 7: Completed Student Surveys by Year

School	Sample (pop) Size 2002	Sample (pop) Size 2003	Sample (pop) Size 2004	Sample (pop) Size 2005	Sample (pop) Size 2006	Total % Participation
Middle Schools	280 (488)	306 (467)	339 (469)	354 (471)	348 (431)	70%
Treatment 1 (Aldine)	26 (36)	20 (36)	30 (38)	26 (37)	32 (40)	72%
Treatment 2 (Chapman)	82 (176)	127 (178)	123 (167)	114 (155)	129 (150)	70%
Total Treatment	108 (212)	147 (214)	153 (205)	140 (192)	161 (190)	70%
Comparison (Fawcett)	172 (276)	159 (253)	186 (264)	214 (279)	187 (241)	70%
Elementary Schools	708 (1095)	636 (1023)	708 (935)	735 (897)	710 (887)	72%
Total Treatment	190 (308)	176 (261)	222 (288)	232 (270)	210 (238)	75%
Total Comparison	518 (787)	460 (762)	486 (647)	503 (627)	500 (649)	71%
Block 1: Treatment 1 (Beacon)	24 (36)	31 (38)	22 (29)	17 (20)	15 (20)	76%
Treatment 2 (Doubleday)	82 (162)	84 (134)	120 (156)	124 (144)	118 (128)	73%
Comparison 1 (Guilford)	119 (182)	119 (176)	133 (171)	123 (165)	136 (178)	72%
Comparison 2 (Harcourt)	211 (297)	168 (288)	191 (247)	188 (222)	186 (216)	74%
Block 2: Treatment 3 (Elsevier)	84 (110)	61 (89)	80 (103)	91 (106)	77 (90)	79%
Comparison 3 (Ingram)	188 (308)	173 (298)	162 (229)	192 (240)	178 (255)	67%
Total	988 (1583)	942 (1490)	1047 (1404)	1089 (1368)	1058 (1318)	72%

Overall,² this resulted in 5,124 completed student surveys and 1,108 completed staff surveys.³ These represented 72% and 66% participation rates, respectively. Of the student surveys, 3,497 were completed by elementary school students and 1,627 were completed by middle school students. Table 7 indicates the total number of participants (sample) and the total possible in the school (population) for each year, split out by treatment and comparison groups and the specific blocks of treatment and comparison school. Of the elementary school students, 1,812 (52.1%) identified themselves as girls, 1,665 (47.9) identified themselves as boys, and 20 surveys (0.6%) did not indicate sex of respondent. Student reports of grade level revealed 1,105 third graders, 1,155 fourth graders, and 1,237 fifth graders. Elementary students were asked about race/ethnicity, but the question appeared to be confusing to many of them, and responses on this question for the elementary school students were deemed unreliable. Of the middle school students, 784 (48.2%) identified themselves as girls, 741 (45.5%) identified themselves as boys, and 102 (6.3%) did not indicate their sex on their survey. Student reports of grade levels revealed 526 sixth graders, 541 seventh graders, and 556 eighth graders, total, across the five years of the evaluation. Middle school respondents who indicated their race or ethnicity identified themselves as Hispanic/Latino (43%), Mixed Race (19%), White (27%), or Native American (2%), reasonably close to the official estimates for ethnic composition of the middle schools. These older respondents less often responded “Don’t Know” (6%), and “Other,” which included Asian and African American (2.6% combined).

Additionally, all teachers and staff members who have any contact with any students throughout a typical school day were invited to complete a school climate survey. Sex of respondent was not reported on staff surveys because such a small percentage of staff were male and reporting this would have jeopardized the anonymity of the surveys. A total of 807 staff members (75.1%) completed surveys about their elementary schools, 213 staff members (17.6%) completed surveys about their middle schools, and an additional 88 surveys (7.3%) were completed by teachers affiliated with both an elementary and middle school in this evaluation project and would cause the totals to be 895 and 301, respectively. The percentages of returned completed surveys were much lower in the comparison schools than they were for the treatment schools. This was likely due to the teachers in the treatment schools knowing about the project and having a vested interest in returning the surveys. In the comparison schools, the front office staff and the teachers in the third, fourth, and fifth grade classrooms were very aware of the survey and its importance, but the only way to explain the study to the rest of the school staff was a note attached to the survey. Despite annual assurances that the school principals would highlight its importance to the school at staff meetings, this did not appear to increase staff participation in their portion of the study.

²We dropped from analysis the data from the elementary school that only completed three years of the study (n= 476) and the data from the sixth graders in one elementary school (n=146) because they were the only sixth graders in an elementary school in the study, and all other sixth graders were in the study were in a middle school environment and completed the more advanced middle school survey.

³Because the study took place over five years but individual participants were not tracked over time, students and staff members could have participated up to five times.

Table 8: Completed Staff Surveys by Year

School	Sample Size 2002	Sample Size 2003	Sample Size 2004	Sample Size 2005	Sample Size 2006	Total % Participation
Middle Schools	70 (98)	64 (101)	51 (67)	59 (68)	57 (87)	71%
Treatment 1 (Aldine)*	23 (39)**	22 (37)	8 (11)	14 (15)	21 (30)	67%
Treatment 2 (Chapman)	19 (22)	18 (22)	19 (19)	16 (19)	13 (16)	88%
Total Treatment	42 (61)	40 (59)	27 (30)	30 (31)	34 (48)	76%
Comparison (Fawcett)	28 (37)	24 (42)	24 (37)	29 (37)	23 (39)	67%
Elementary Schools	173 (270)	167 (277)	186 (267)	183 (275)	186 (294)	65%
Total Treatment	83 (129)	75 (124)	91 (102)	88 (97)	106 (120)	77%
Total Comparison	90 (141)	92 (153)	95 (165)	95 (178)	80 (174)	56%
Block 1: Treatment 1 (Beacon)*	23 (39)**	22 (37)	8 (11)	14 (15)	21 (30)	67%
Treatment 2 (Doubleday)	28 (42)	28 (38)	33 (35)	30 (34)	32 (34)	79%
Comparison 1 (Guilford)	42 (63)	42 (62)	37 (67)	36 (66)	30 (77)	56%
Comparison 2 (Harcourt)	37 (52)	27 (46)	36 (49)	30 (49)	26 (40)	66%
Block 2: Treatment 3 (Elsevier)	32 (48)	32 (49)	50 (56)	44 (48)	53 (56)	82%
Comparison 3 (Ingram)	11 (26)	23 (45)	22 (49)	29 (49)	24 (57)	45%
Total	220 (329)	209 (341)	229 (323)	228 (328)	222 (351)	66%
<p>*Middle School Treatment 1 and Elementary School Treatment 1 are housed in a single building and the school does not observe distinctions between elementary and middle school staff .</p> <p>** The total number of staff in those combined schools was not recorded in 2002, and so the total for 2003 was used to calculate the total participation rate for the school.</p>						

Data Collection Procedures

Process evaluation in treatment schools. In the early fall of each year, the treatment schools received training on the BPYS curriculum. The training was led by BPYS technical consultants. Ideally, the staff received eight hours of training in both the first and second years of implementation, and four hours of staff training in the fall of the last year of implementation. Each school selected a coordinator that would help to form and lead a cadre and be the contact person for the CSPV and BPYS personnel. The cadre was charged with facilitating the implementation of the BPYS program, including having regular meetings, addressing concerns from staff, parents, and students, and directly helping teachers with BPYS lessons when needed. BPYS technical assistants were available to provide support both on the phone and in person when necessary.

Process evaluation was done separately by a CSPV evaluator and a BPYS technical consultant. The BPYS technical consultants were trained in-house, and CSPV did not participate in that training. BPYS developed the 10-item rating scale used in the process evaluation of this study. In consultation with CSPV, BPYS personnel developed specific descriptions for rating a construct as low, medium, or high in the process evaluation. These descriptions were used by both BPYS and CSPV in the process evaluation to rate implementation fidelity to the program components upon completing observations in the schools each semester in which the program was actively being implemented. The CSPV evaluator and the BPYS technical consultant contacted the school coordinator to set up days to observe the implementation of the BPYS curriculum in the classrooms at each school. Both gave ratings on the fidelity of implementation throughout the evaluation. The BPYS and CSPV ratings were independent, and while they are generally in reasonable agreement (see Part 3), to the extent that there were discrepancies between the BPYS and CSPV ratings, this may be taken as evidence of some (generally minor) unreliability. These ratings were condensed into two ratings per year, one per semester. In addition, the CSPV personnel surveyed members of the cadre, the teachers, and conducted in person interviews with both the coordinator and the principal concerning the implementation process. These surveys and interviews were summarized into single year-end reports. These year-end reports are attached as Appendix 1 to this report.

Outcome evaluation in treatment and comparison schools: student surveys. In the spring semesters of 2002, 2003, 2004, 2005, and 2006, all participating schools completed a school climate survey. The parental consent forms were written in both English and Spanish. Some schools preferred that all forms were sent home in both languages, and in some schools, we asked students to indicate whether they wanted to take home a form in English or in Spanish. Copies of the parental consent forms are available in Appendix 2. In order to gain consent from the parents, members of the research team brought consent forms to the school, and gained permission from the principal and affected teachers to address the students. The format varied by school: sometimes the researcher spoke briefly with individual classrooms, and sometimes whole grades were gathered together in a “pod” outside of their classrooms. The researcher explained the study to the children in developmentally appropriate language (e.g., relating the study to a science fair project and the students were the “experts” who could tell them what their school was like), and asked the children to take home the consent forms to their parents to be

signed and then return the signed forms to school. Members of the research team would return to the school several times per week for approximately two weeks to pick up returned forms and to serve as a reminder to the students to bring back their signed consent forms. An incentive was offered to the children for returning the signed forms, regardless of whether their parents checked “yes” or “no” on the consent form. The incentive varied by school and was at the discretion of the principal (e.g., individual rewards such as an ice cream bar for each student who returned a signed form, or classroom wide rewards such as a pizza party for the classroom that brought back the highest percentage of forms, or for any classroom that brought back all of their forms). This resulted in a return rates of around 90%.

Once the consent process had begun, the research team scheduled times with the teachers to return to the school and conduct the survey with the students. The surveys could be completed in third grade classrooms within one hour, and took less time with older students. The surveys were also available in Spanish, which was the most common non-English first language of students in the treatment and comparison schools. The research team checked in advance to see if any students would want to take the survey in Spanish instead of English, and also brought extra Spanish surveys in case any students changed their mind on the day of the survey. The research team always included a fluent Spanish speaker in order to increase the degree to which primarily Spanish speaking respondents would feel comfortable answering questions on the survey. It should be noted that most schools intentionally housed all students whose primary language was English within the same classroom at each grade level, which eased the task of providing bilingual assistance to the primarily Spanish speaking respondents in the study.

On the day of the survey, when the researchers first entered the classroom, they consulted with the teacher to be sure only those students with consent remained for the survey. When possible, more than one researcher went to a classroom to assist with answering question the students had and to maintain positive behavior in the room. Those students whose parents declined consent on a signed form or who didn’t bring back a signed form left with the teacher to read or do homework. The researcher then handed out surveys, pens, and non-white pieces of paper to use as cover sheets.

The researcher guided the students in completing their assent forms, which asked the students to write and sign their names and then to indicate whether or not they wanted to participate in the survey. Copies of the student assent forms are available in Appendix 3. If the student declined to participate, they were sent out of the room to join their teacher and their fellow non-participating classmates. The researcher then discussed with the students the three ways they would keep their answers private: (1) the researchers would not divulge their responses to anyone they knew; (2) the students needed to use their colored cover sheets to keep their classmates from seeing their responses; and (3) the students were asked not to talk during the survey unless they had a question for the researcher, in which case they needed to raise their hand. Students were also told that they could leave blank any questions they did not understand or that they did not want to answer.

Next, the researcher guided the students through the sociodemographic portion of the surveys and showed the student how to use their colored piece of paper to cover up their

answers. At this point, in the middle schools, the students were allowed to read and reply to the questions on their own, but research staff were available to read a question for them if were not able to understand it on their own. In the elementary school classrooms, the researcher read each question aloud, twice, and students marked their own answers. If a student had a question about an item on the survey, the researcher would give the student the best possible response without leading the student in any way. In some instances, it simply required re-reading the question, as it was clear the student had mis-heard or mis-read the question. The research team was specifically instructed not to re-define any words or to work through a question with a child in any way. Often the response given was that it was OK to leave any questions blank if they didn't understand them. Specific items that consistently led to questions are identified in the measurement section of this document.

At the end, the research team collected the completed surveys and thanked the students for their participation. The research team allowed the students to keep their pens and left pens for the nonparticipating students with the teachers. At this time, the team made arrangements with the school and the individual teachers to conduct make-up surveys with children who were absent on the day their classroom was surveyed. When the research team returned to CSPV, the surveys were logged in and put into locked file cabinets. A final accounting to match parental consent forms with student assent forms was conducted, and these consent and assent forms were locked in separate file cabinets. The surveys were only removed to enter the data into the computer and were then returned to the locked cabinets as soon as possible thereafter.

Outcome evaluation in treatment and comparison schools: staff surveys. On the first day of student surveys in a school, the research team placed a packet in the mailbox of each teacher and each school staff member who had any contact with the students, including cafeteria workers and maintenance staff. The packet consisted of a letter explaining the staff survey, a staff survey, and a postage-paid envelope that allowed the staff member to mail the survey back to the research staff anonymously. Once the data arrived, the surveys were logged in and put into locked file cabinets. They were only removed to enter the data into the computer and were then returned to the locked cabinets as soon as possible thereafter.

Data preparation and interim reports to schools. Data were entered into the computer, and then a system of checking the entered data was implemented. First, a random twenty percent of the entered data was hand checked with the raw data, and if errors were found, they were corrected and then another random twenty percent was hand checked, and if errors were found they were corrected and then another random twenty percent was hand checked. This system continued until a clean twenty percent was found. At this point, descriptive analyses were conducted on the data to check for any out of range responses. If any were found, they were corrected and the items around that item were checked. Once this process was complete, the data were deemed "clean" and ready for analysis.

Annually, these data were prepared into a simple report of descriptive data for the schools. Both treatment and comparison schools were given this feedback in the late summer or very early fall and were given permission to use the results as they wished in planning for the upcoming school year.

Measurement

Overview. The elementary and middle school surveys were composed of 83 and 131 questions, respectively. The staff surveys were composed of 38 questions. Not all of these questions were directly related to the evaluation of BPYS; instead, some are part of the broader Safe Communities ~ Safe Schools study (Delbert Elliott, principal investigator) within which, as noted earlier, the present evaluation of BPYS was undertaken. The responses for questions about attitudes and relationships were yes/no or extent of agreement (strongly agree/agree/disagree/strongly disagree) scales. The responses measuring frequency of behavior differed between the two student versions: In the elementary school, responses were “Never,” “Once,” “A Few Times,” and “A Lot,” whereas the middle school students were asked to answer “yes” or “no,” and if they answered “yes,” they were asked to provide the number of times in the past month that these behaviors occurred. The staff survey used a five point scale for frequency of behavior: “never,” “almost never,” “occasionally,” “often,” and “very often.” Copies of the elementary school, middle school, and staff surveys are available in Appendix 4.⁴

The surveys were translated into Spanish (and then validated by being translated back into English) for use by students whose primary language was Spanish. When a child professed proficiency in both languages, the researchers asked them to select whichever survey they would feel most comfortable completing. Copies of the elementary and middle school surveys in Spanish are also included in Appendix 5.

Measurement domains. Measures used on the outcome evaluation are divided into the following domains:

(1) Sociodemographic information: gender, ethnicity, grade in school, self-reported school performance (what grades the student is getting in school).

(2) Relationships with parents and other adults. The impact of a school-based intervention may be limited by the family environment, either because students from favorable family environments are less likely to be involved in bullying or other problem behavior regardless of the school-based intervention, or because the intervention cannot completely overcome the negative influence of a dysfunctional family environment.

(3) Friends attitudes toward aggression and violence. As noted earlier, association with friends whose attitudes and behaviors are favorable toward violence and aggression is one of the most robust predictors of violence and aggression. To the extent that BPYS has an impact, either on choice of friends or on attitudes and behaviors of friends, it can be expected to reduce violence. Exposure to prosocial or antisocial friends is an important mediating variable in

⁴Note that the middle school surveys are labeled Middle/High School. This is because they while the surveys were being developed, it was still a possibility that the program would be implemented and evaluated in high schools in addition to elementary and middle schools.

predicting school violence. Also, as with family environment, the impact of the school-based intervention may be limited by exposure to prosocial or antisocial friends; students with prosocial friends may be less likely to be involved in bullying and other problem behavior regardless of the intervention, and the program may be insufficient to overcome the effects of antisocial peer groups, except insofar as it changes the peer group or the attitudes of peers.

(4) The student's own attitudes toward aggression and violence. Like friends' attitudes toward aggression and violence, this is an important mediating variable in predicting school violence, and also like the family and peer group environment, it may limit the effectiveness of the intervention, except insofar as the intervention is able to change the attitudes of the students.

(5) General attitudes toward school and school climate, including questions about liking school, whether teachers try to prevent bullying, whether there is gang activity at the school, whether teachers and students respect one another. As noted earlier, school-related stress and alienation may be a cause, an effect, or both with respect to bullying. For middle school but not elementary school students, perception of school climate includes perceptions of substance use at school, and the availability of and student's own participation in school or other activities. Whether teachers and other adults try to prevent bullying is one of the targeted outcomes for BPYS, specifically for the first of the three major components of the program as described earlier.

(6) Questions that ask specifically about perceptions of school safety and whether students avoid school because of fears for their safety. This is one of the targeted outcomes for BPYS, specifically for the third of the three major components of the program. Included here are questions about perceptions of other students being bullied at school, also one of the targeted outcomes for BPYS, specifically for the second of the three major components of the program.

(7) Questions about victimization and perpetration of relational aggression (Crick and Grotpeter 1995; Crick and Grotpeter 1996; Grotpeter and Crick 1996), physical aggression, and physical violence, including encouragement of aggression and violence. The literature on bullying indicates that while males are much more likely to be victims and perpetrators of physical aggression, females are more likely to engage in relational aggression, the types of behaviors listed under "social alienation" in Table 1 above. Failure to include relational aggression would risk missing an important aspect of bullying in the school context which, although less likely than physical aggression to produce physical injury, may nonetheless have serious emotional consequences to the victim and perpetrator. Reduction of relational aggression, physical aggression, and physical violence are all explicit goals of BPYS, specifically for the second major component of the program, and they are also goals of the state of Colorado's mandate for schools to enhance school safety. The use of self-report measures is, after decades of research on the method, well known to produce more accurate estimates of behavior than official reports for both illegal and deviant behavior generally (Elliott et al. 1989) and for bullying in particular (Ireland 2002).

For middle school but not elementary school students, questions were also asked about substance use (tobacco, alcohol, marijuana, and other illicit drugs, both at school and more

generally). The inclusion of substance use for middle school but not elementary school students simply reflects the fact that substance use is strongly age-related, and the use of illicit substances tends to be relatively rare at earlier ages (e.g., Elliott et al. 1989). These questions were included as part of the broader Safe Communities ~ Safe Schools project, and are not central to the evaluation of BPYS, but we will briefly address whether BPYS had any collateral impact in this area.

To summarize, then, our outcome measures include questions about personal victimization by violence and relational aggression, questions about school safety, and questions about risk factors for violence, aggression, and other problem behaviors including exposure to delinquent or deviant peers and beliefs about how wrong it is to violate the law, two of the more robust predictors of problem behavior (Elliott and Menard 1996; Menard and Elliott 1994; Menard and Huizinga 1994) derived from the integrated theory of Elliott et al. (1979; 1985; 1989; see also Roitberg and Menard 1994). Even if BPYS had no effect on bullying *per se*, but was effective in reducing violent victimization and perpetration of violence in the school (as well it might be, given the elements of the program designed to build skills to avoid victimization of oneself and to help prevent victimization of others), the program would still be considered a success. We are also interested in whether the impact of the program (if any) on violent victimization appears to operate indirectly, via the program's impact on bullying, or whether the program has effects on violence above and beyond its impact on bullying.

Analytical Methods

The design of the data collection for the study is a repeated cross-sectional design (Menard 2002c). Within each cross-section, the basic design is a hierarchically nested multilevel design (Boyd and Iversen 1979; Brown and Melamed 1990; Jackson and Brashers 1994). Students are nested within the schools, and schools are split between treatment and comparison conditions, with unequal numbers of cases and potentially different distributions, ranges, and variances on outcome variables. High turnover of students as a result of both geographic mobility and the normal processes of entry into and graduation from elementary and middle school over a three-year period effectively precludes the use of intraindividual change models such as multilevel growth curve modeling (Raudenbush and Bryk 2002) or latent growth curve structural equation models (Little et al. 2000; Stollmiller 1995). Instead, individual-level data are used to assess inter-school differences and intra-school change. Inter-school differences are assessed by examining the impact of the treatment, BPYS, on the targeted outcomes: perceptions of school safety, perceptions that bullying is discouraged, perceptions that others are being physically or otherwise attacked or picked on, and self-reports by students of their own victimization by and perpetration of physical and relational aggression.

In research involving extensive lists of illegal behaviors, the best approach would be to analyze prevalence and frequency separately. Because of the age of the elementary students, only prevalence of a limited number of behaviors was asked. For the middle school students, both prevalence and frequency were obtained, but again for a limited number of behaviors. These measures of illegal behavior are best regarded as manifest indicators of a latent variable, extent of aggression, violence, and bullying, and in this context, standard psychometric scaling

techniques are appropriate for the outcome measures involving behaviors (self-reported relational and physical aggression victimization and perpetration) as well as the perceptual and attitudinal measures (extent to which others are bullied, extent to which bullying is discouraged, feelings of school safety).

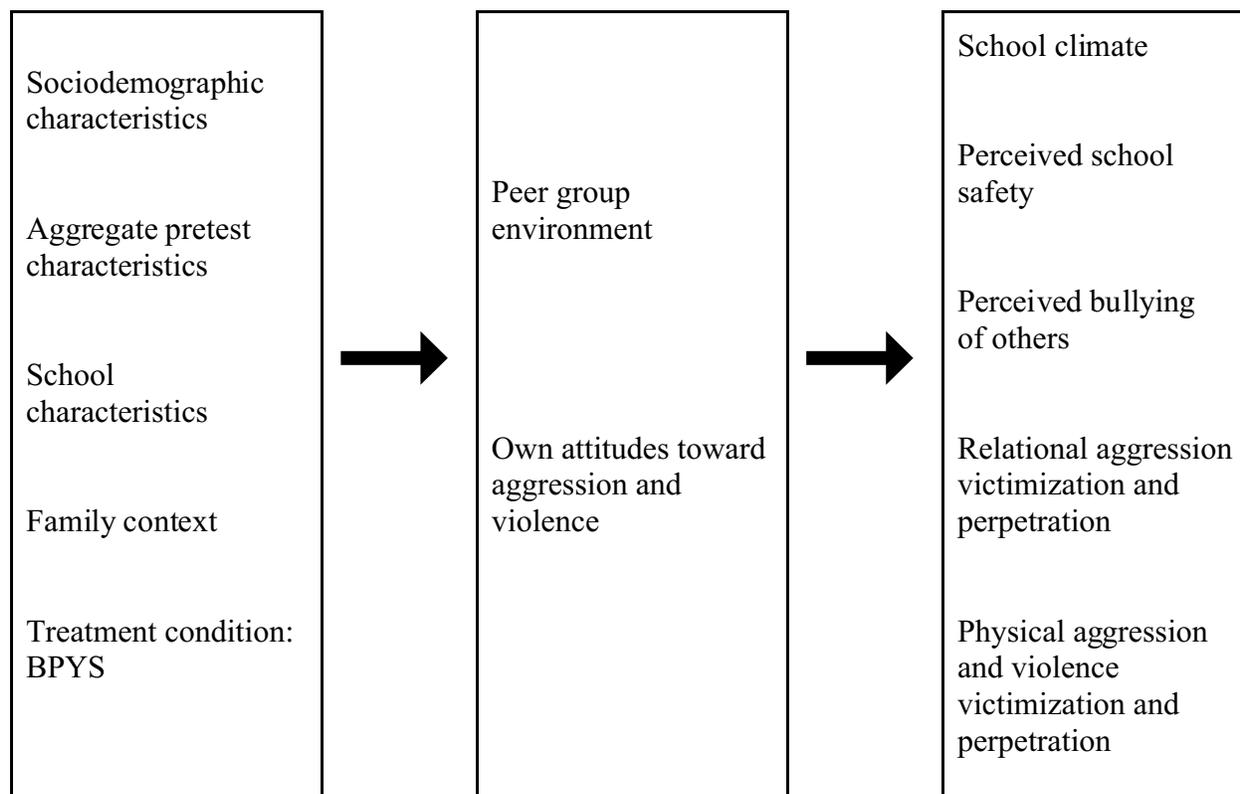


Figure 3: Evaluation Analysis Model

The basic model to be tested is represented in Figure 3, above. Sociodemographic characteristics of the students (gender, ethnicity, class in school - i.e., third, fourth, or fifth grade in elementary school) are measured directly. We should note that particularly at the elementary school level, students seemed confused about the question of race or ethnicity, and the classification by race and ethnicity may have limited reliability and validity. Aggregate pretest characteristics are taken into consideration by comparing results for year 1, the pretest year, with subsequent years. Similarly, differences between treatment and comparison schools at year 1 serve as a point of comparison for results for subsequent years. The critical family background variable for present purposes is family bonding, which, based on past research (e.g., Elliott et al. 1989) may have a substantial impact on students' attitudes toward aggression, their peer group climate, and (most likely indirectly through attitudes and peer group climate) their own

involvement in aggressive or violent behavior. These variables, along with the treatment/comparison distinction are treated as purely independent, exogenous variables, and are hypothesized to affect the attitudes of the students themselves, and the peer group climate, the intervening variables. The exogenous and intervening variables, in turn, are hypothesized to affect the outcomes of interest: school climate, bullying, relational and physical aggression, and physical violence.

Scaling

A number of scales were incorporated into the design of the Safe Schools ~ Safe Communities research from which the present study is derived. Here, however, we use only those scales directly pertinent to our evaluation of the BPYS intervention. Scales were constructed in two stages. First, factor analysis was used to test for dimensionality. Second, for each scale identified as representing a single dimension, the items were added, and the resulting additive scale was examined for reliability using Cronbach’s alpha (α), a standard measure of additive scale reliability (see, e.g., Zeller and Carmines 1980 for a discussion of reliability testing for additive and factor scales). Scales were not constructed for sociodemographic characteristics, measured separately, or for aggregate pretest characteristics and school characteristics, which are assessed using the baseline (pretest) year results; and the treatment condition is a simple dichotomous variable, coded 1 for the treatment schools and 2 for the comparison schools. The reliabilities for the remaining scales used in the model depicted in Figure 3 are summarized in Table 9 below; the full listing of the items included in each of the scales is included in Appendix 6.

Table 9: Scale Reliabilities (Cronbach’s α)

Scale	Elementary School	Middle School
Family context: family bonding	.76	.82
Peer group environment: attitudes toward aggression	.85	.87
Respondent’s own attitudes toward aggression	.75	.84
School climate: discouragement of bullying	.60	.77
Perceived school safety	.62	.73
Perceived bullying of others	.75	.63
Relational aggression perpetration	.74	.62
Relational aggression victimization	.84	.69
Physical aggression and violence perpetration	.82	.64
Physical aggression and violence victimization	.76	.64

It can be seen from the results in Table 9 that the reliabilities fall within generally acceptable ranges, with reliabilities on the perceptual and attitudinal measures being higher for middle school students, but reliabilities for the behavioral measures being higher for elementary school students. This may be a function of different (age-appropriate) wording in the elementary and middle school items, or it may reflect a change in the degree to which behaviors are clustered as individuals get older. In either case, the reliabilities appear to be adequate to proceed with the planned analysis. For some outcomes of interest, (perceptions of general school climate, perceptions that rules were well known and fair), the items failed to have satisfactory scale properties; analysis of these items is limited to the item-level results.

The *school climate: discouragement of bullying* scale is central to evaluation of the first component of the program. According to Hypothesis 1, as stated earlier, *compared to students in schools in which BPYS is not implemented, students in schools in which BPYS is implemented will perceive greater intolerance of bullying*. The school climate: discouragement of bullying scale is used here as our indicator of perception of greater intolerance of bullying.

The *perceived bullying of others* scale, along with the *relational aggression perpetration, relational aggression victimization, physical aggression and violence perpetration, and physical aggression and violence victimization* scales are used to evaluate the second component of the program. According to Hypothesis 2, as stated earlier, *compared to students in schools in which BPYS is not implemented, students in schools in which BPYS is implemented will report lower rates of (a) victimization by, (b) perpetration of, and (c) witnessing physical nonphysical aggression*. The perceived bullying of others scale is the scale most directly related to bullying, but the other scales, as noted earlier, are also critical to this hypothesis. One could hardly call the program a failure if it had no significant impact on bullying, but was successful in reducing physical and nonphysical aggression more generally.

The *perceived school safety* scale is central to the evaluation of the third component of the program. According to Hypothesis 3, *compared to students in schools in which BPYS is not implemented, students in schools in which BPYS is implemented will report higher rates of feeling safe and lower rates of feeling unsafe at school*. The perceived school safety scale is used here as our indicator of feeling safe (or unsafe) at school.

In addition to the analysis of the model depicted in Figure 3, separate analysis is performed examining the differences between treatment and comparison schools at the item level, taking each question separately rather than as part of a scale. The first concern is with how well the treatment and comparison schools were matched at baseline (year 1, pre-treatment). To the extent that the treatment schools at baseline resemble the comparison schools, we would have evidence that differences between the treatment and comparison schools subsequent to program implementation could plausibly be attributed to the treatment. If, however, the treatment schools at baseline are dissimilar to the comparison schools, such inference would be inappropriate. In this case, the focus needs to be on comparing changes over time in the treatment and comparison schools, rather than on cross-sectional differences between the treatment and comparison schools.

Part 3: PROCESS EVALUATION RESULTS

Overview. The following is a narrative describes the process evaluation measuring the fidelity of program implementation for each treatment school. To preserve confidentiality of the schools, per their agreement to participate, pseudonyms are used to identify the schools. (The pseudonyms are based on past and currently existing publishing companies.) The schedule for the present implementation of BPYS called for the treatment schools to receive eight hours of training from BPYS in the early fall of 2002. The training specifically addressed ten elements of implementation at training sessions at the schools: (1) Staff acknowledgment of the problem of bullying and their commitment to the creation of a safe school; (2) Administrative support for the program; (3) School-wide discipline plan in place; (4) Bully-Proofing cadre formed to design and guide implementation of the program; (5) Assessment of current school climate and safety issues; (6) Training of staff; (7) Training of students; (8) Support from the parent community; (9) Strategies for ongoing development of the caring community; and (10) Evaluation of the program. Each semester after this, both CSPV and BPYS conducted observations of teachers implementing the program in the schools, surveyed members of the cadre and the teachers, and conducted in person interviews with both the coordinator and the principal concerning the implementation process. After doing this, CSPV and BPYS staff each completed forms rating the fidelity of implementation on those same ten implementation elements (i.e., low, medium, or high). The specific parameters used for giving a school a low, medium, or high rating are included in Appendix 7.

Interrater reliability was estimated by calculating the correlation coefficient (Pearson's r) between BPYS and CSPV average implementation scores across all of the treatment schools (three elementary and two middle schools) at each assessment (twice annually during the three years in which the program was actively being implemented). Based on this approach, the interrater reliability $r = .75$, which is acceptable, but which does reflect some disagreement in some schools in some years between BPYS and CSPV raters. To the extent that such discrepancies occurred, they are illustrated in the process evaluation reports for the individual schools; see, in particular, the graphs assessing average implementation for each school.

The process evaluation narratives are a combination of qualitative and quantitative data. For each school, we constructed a chart illustrating fidelity of implementation. Specifically, we computed scores for each school that represented the average fidelity ratings given by the (1) CSPV staff and (2) BPYS staff who observed the implementation. For each semester, we assigned numeric values to the nominal categories (e.g., low=1, medium=2, high=3) and averaged across the ten categories. This summary view is provided for each school in this narrative. This summary view is by its nature very broad, but greater detail is available in Appendix 1 of this report.

In addition to these ratings, teachers were asked to debrief at the end of each school year in a survey that assessed their experience with BPYS implementation. The questions asked are available in Appendix 8 of this report. Teachers rated each item on a four point scale (i.e.,

strongly agree, agree, disagree, strongly disagree). For the sake of parsimony, the results are presented here as dichotomous scales (i.e., strongly agree/agree vs disagree/strongly disagree).

The reports that follow also provide qualitative narratives describing the degree to which the schools implemented the program with fidelity. These narratives are drawn from the narratives written each semester by the CSPV technical assistance providers and thus provide commentary that is more editorial than purely objective in nature. This commentary is based on the degree to which the following plan was followed:

The eight hour introductory training sessions were to be conducted at each of the treatment sites individually. For each subsequent year following the initial BPYS training, there were to be shorter training sessions in the early fall. These training sessions served as a refresher course for teachers who had already been through the program and helped to bring new teachers up to date with the program so that they could implement the curriculum in their classroom.

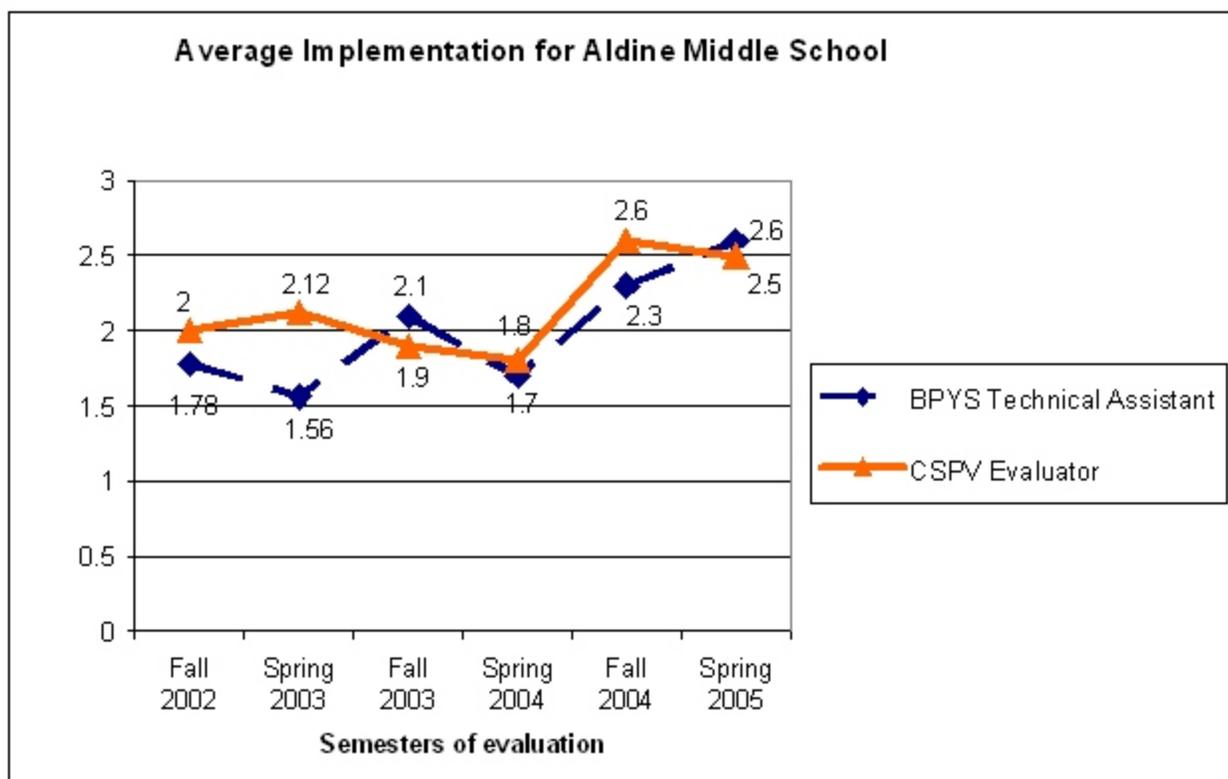
Each site was also charged with creating a cadre to facilitate the implementation of the program. Ideally, cadres were to be composed of members that represented a cross-section of the school: teachers, administrators, support staff, mental health, classified staff, parents, students, and community members. Each cadre was charged with selecting a chairperson or coordinator who would be in charge of cadre meetings, serve as a contact person for CSPV and BPYS technical assistants, and be active in addressing any problems and concerns that arise from the BPYS implementation. BPYS recommended that the cadre be given the same status as other school committees. In order to help the cadres improve their techniques and strategies for implementing the curriculum, regional cadre meetings were held once each year of implementation.

Throughout the year, BPYS technical assistants were charged with providing telephone and on-site support when needed. This proved challenging for some schools located further from the BPYS offices (e.g., a days drive away). These sites often received less than the program's designed amount of on-site assistance and thus more work was conducted over the phone.

The following are summaries of the process evaluation for each school. Each school writeup contains the above described scores for fidelity of implementation and a narrative that discusses both the successes and the barriers to implementation faced by each school. Graphs constructed using responses from the teacher surveys are also included to provide relevant information that varies by school. Complete semester by semester reports that contain process evaluation data for each of the five treatment schools is available in Appendix 1 of this report.

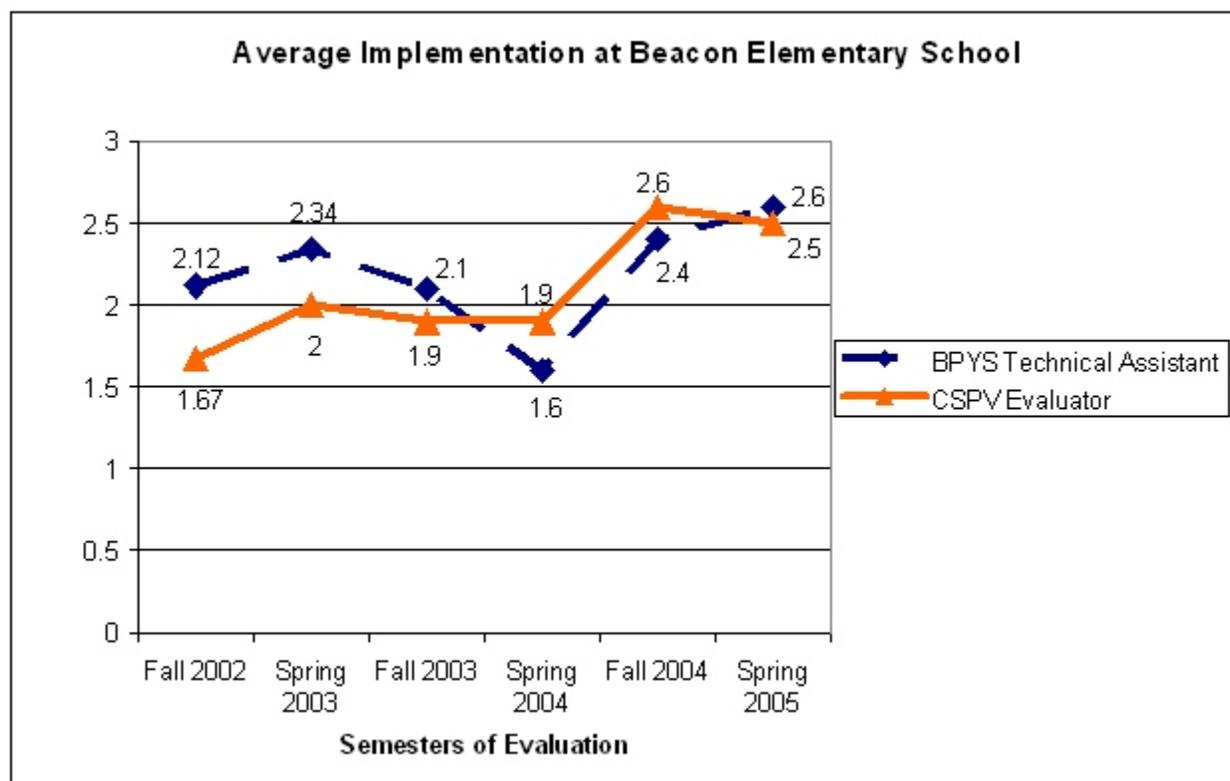
Treatment middle school 1 and Treatment elementary school 1 (Block 1)
Aldine Middle School and Beacon Elementary School

This report describes the process evaluation of the BPYS implementation in Aldine Middle School and Beacon Elementary School, which are located in the same school district. When possible, the report will distinguish between the elementary and middle school implementations; however, both schools are served by one principal and the same cadre, and the schools do not distinguish between elementary and middle school staff, thus most aspects of this summary will not distinguish between elementary and middle schools. Overall, the BPYS implementation in the Aldine Middle School and Beacon Elementary School was slow to gain momentum due primarily to suboptimal administrative buy-in to the program. The staff initially reflected this lack of interest on the part of the administration, and the program implementation did not reflect fidelity to the program as designed. During the final year of implementation, however, the principal was highly motivated to implement the program with fidelity. This support encouraged the teachers and other school staff, and program implementation fidelity dramatically improved in the final year of the evaluation.



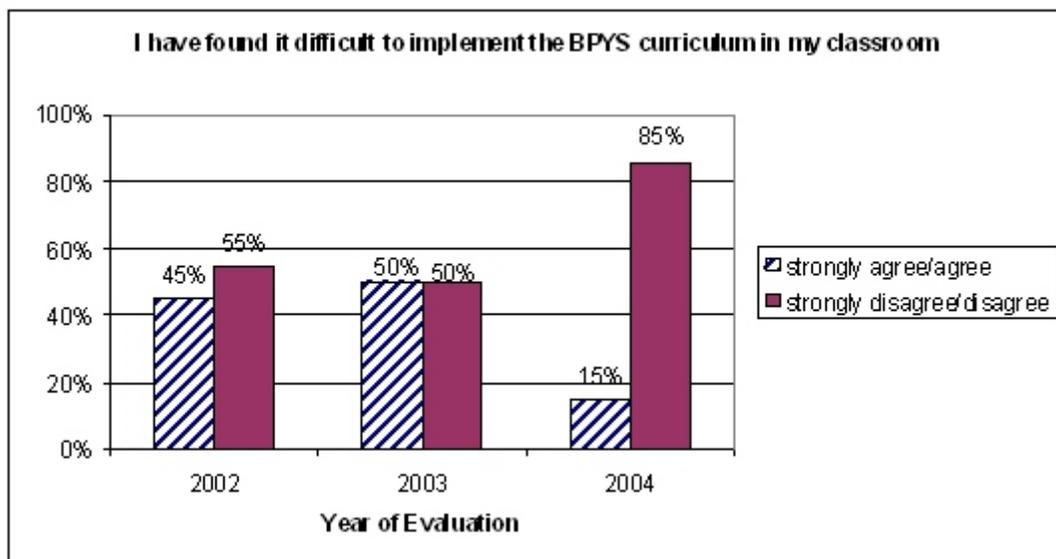
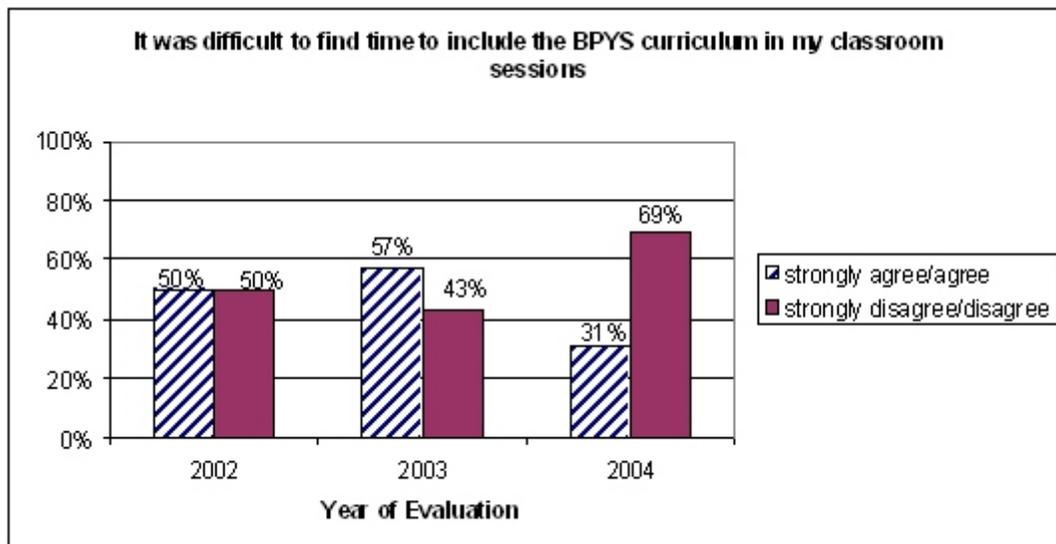
Year One. During the first semester of the program, the implementation of BPYS was hampered by a lack of leadership. No staff members, including the principal, were willing to take on the role of coordinator. Without strong leadership from above, staff at both the elementary and middle schools were left without direction for implementing the program. When BPYS technical assistance providers and the CSPV evaluator met with staff in October 2002,

some teachers indicated that they were not aware that they were supposed to be implementing the curriculum. In December 2002, the principal reluctantly agreed to step in as coordinator. This helped to facilitate communication with BPYS and CSPV, who could then schedule observation times with the schools in order to collect information on the fidelity of implementation. During the first year of implementation, Aldine Middle School and Beacon Elementary School’s overall fidelity scores were brought down specifically by this low administrative support and also low support from the parent community. The CSPV evaluator also gave the schools low scores for their assessment of the current school climate and safety issues. Additionally, communication between the principal and the evaluator (CSPV) and technical assistant (BPYS) was poor. As a result, the CSPV evaluator was unable to observe and document many teachers implementing the BPYS program. Moreover, members of the cadre noted that one elementary teacher had not even looked at the BPYS curriculum. At the middle school, the teachers were only slightly more invested in implementing the program. The two sixth grade teachers had conducted all of the lessons, but the seventh grade teachers were not as involved. Of note is that all of the teachers who completed the teacher questionnaire found the techniques provided in training to be helpful. However, the positive feelings about the training did not translate into successful implementation for many teachers.



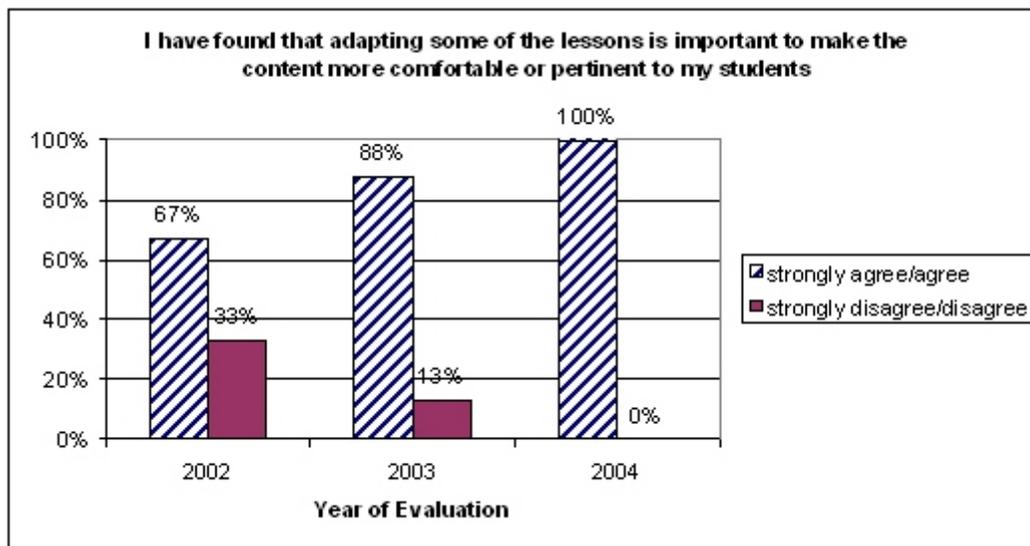
Year Two. The second year of evaluation arrived with several changes at Aldine Middle School and Beacon Elementary School. First, one of the teachers on the cadre left the district. The district decided not to hire a new teacher to replace her, but instead combined the 2nd and 3rd grade into one classroom, thus reducing the already small teaching staff from five to four. Second, the elementary and middle schools got a new principal. He provided support for the

BPYS program in theory though he did not demonstrate much commitment to the program at the school level during his first school year. Despite an August 2003 meeting with the CSPV evaluator where she explained the program and evaluation, the principal did not view the program as an integral part of the curriculum and did not consider it a priority for the district. Instead of scheduling time for the full eight hours of BPYS staff training in the fall, the principal only scheduled three hours. Further, in the principal interview, he indicated that the teachers are “squeezing other academic programs to get time for BPYS.” Rather than viewing the BPYS program as an integral part of the curriculum which can help support the academic achievements of students, he considered it to be an “add on.”



The graphs above give an indication of the struggle teachers reported with finding time to implement the program in their classrooms. During the 2003-04 school year, the cadre was engaged in trying to facilitate implementation and struggled with little support from the

administration. Despite these challenges, the cadre continued to push forward with its effort to help teachers implement the BPYS program and to create a caring community, and ease of implementation improved in the final year. Unlike problems reported by the other treatment middle school teachers, the teachers at the schools did not report much, if any, resistance to the BPYS curriculum. Teachers did, however, feel that adapting some of the BPYS lessons was important in order to make the program more pertinent to their students.



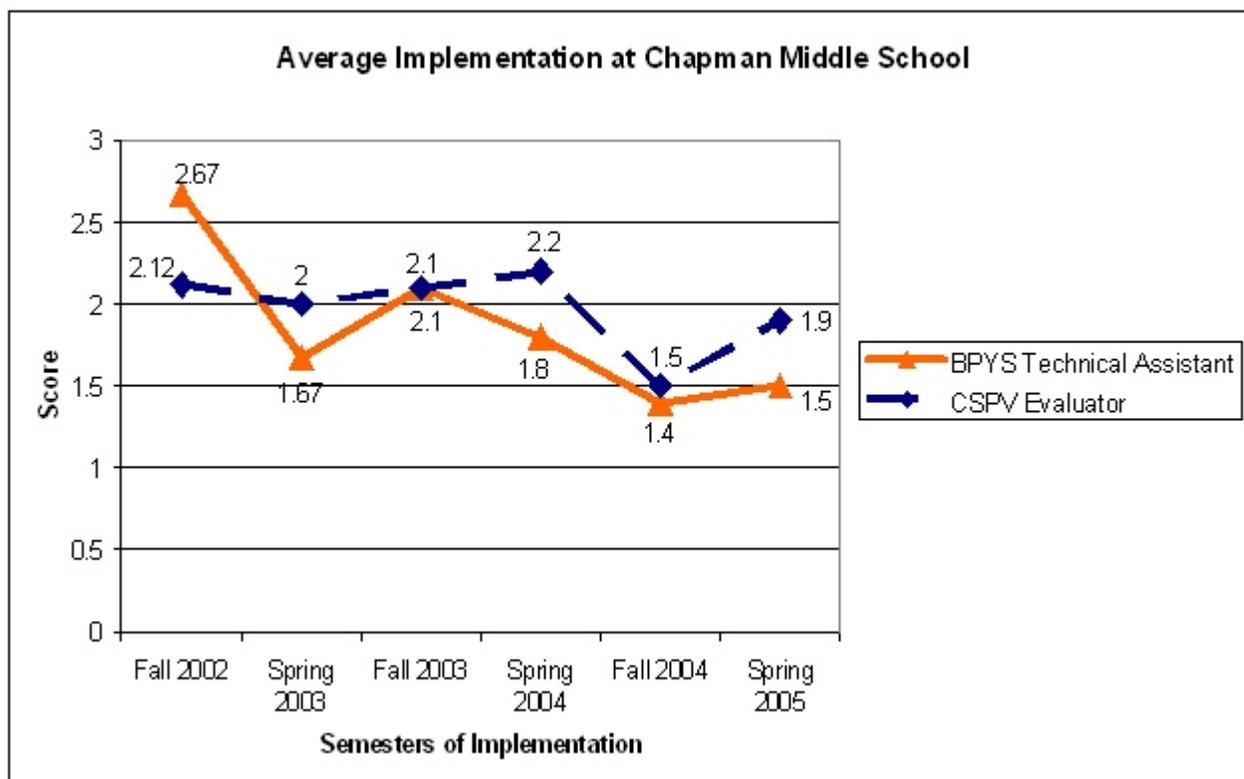
Year Three. One of the biggest reasons for the positive change in the Aldine Middle School and Beacon Elementary School was a complete shift in attitude by both the principal and the superintendent. Low CSAP (Colorado Student Assessment Program) scores from the spring 2004 semester put the school district “on watch” by the Colorado Department of Education. In recognition of these low scores, the principal and superintendent decided to make an effort to implement BPYS with fidelity, hoping it would help to improve their schools overall. The change was profound. First, the principal redeveloped the district’s discipline code, which had a marked effect on discipline referrals and staff morale. The teachers responses to the statement, “School rules regarding bullying have been enforced,” for example, reflected the change that occurred at the schools. Whereas in 2002, only 50% of teachers agreed with the statement, in 2004, 100% agreed that the rules were being enforced. In addition, the principal seemed more committed to the BPYS technical assistant consultations and he also committed to the full four hours of staff training during the last year of implementation. It appeared that by the principal placing more emphasis on the program and implementing it with fidelity, the staff saw the value of the curriculum. The cadre noted these changes by giving the teachers high ratings for implementation during the last year. The CSPV evaluator also noted a change at the school: Classroom observations conducted by the CSPV evaluator revealed a marked improvement in several areas, particularly teacher understanding, student understanding, and student receptivity. Moreover, for the first time since the evaluation began, the CSPV evaluator was able to schedule classroom observations with little or no teacher resistance. Teachers were more open and cooperative in both scheduling as well as following through with implementing on the agreed

upon time and date. The BPYS technical assistant also requested that the teachers make a schedule, which may explain some of the change as well. Many teachers admitted that once they set a standing time and day for BPYS curriculum, it became much easier for them to follow through.

Another notable improvement was the increased effectiveness of the cadre. In earlier years, though they attempted to facilitate the implementation of the BPYS program, they were suffering a lack of support and cooperation from the principal and several teachers. Despite this lack of support, the cadre had been successful at meeting regularly to create strategies to motivate the students. For example, as a group, the cadre worked to create a “caring community” by rewarding students “caught” using the BPYS skills and language. The principal, who had attended the meetings during the previous school year, was much more involved in discussion and planning with the cadre in 2004-05, and he made BPYS a standing agenda item at monthly staff meetings which was taken seriously. Once the cadre had the support of the staff and administration, their effectiveness dramatically increased. Over time, the teachers reported a better understanding of the timeline for implementation, that the school rules regarding bullying were being better enforced, that a “caring community” was developing at their school, and that they had more parent support for the program. In addition, a greater percentage of teachers reported each year that the faculty supported each other in the implementation of the curriculum and that they were consistently using classroom meetings to facilitate a “caring community.”

Treatment Middle School 2
Chapman Middle School

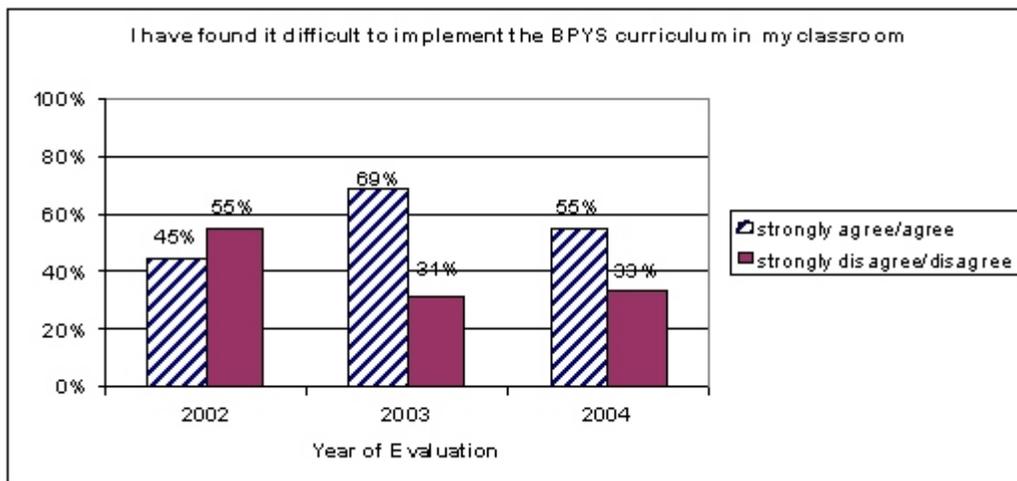
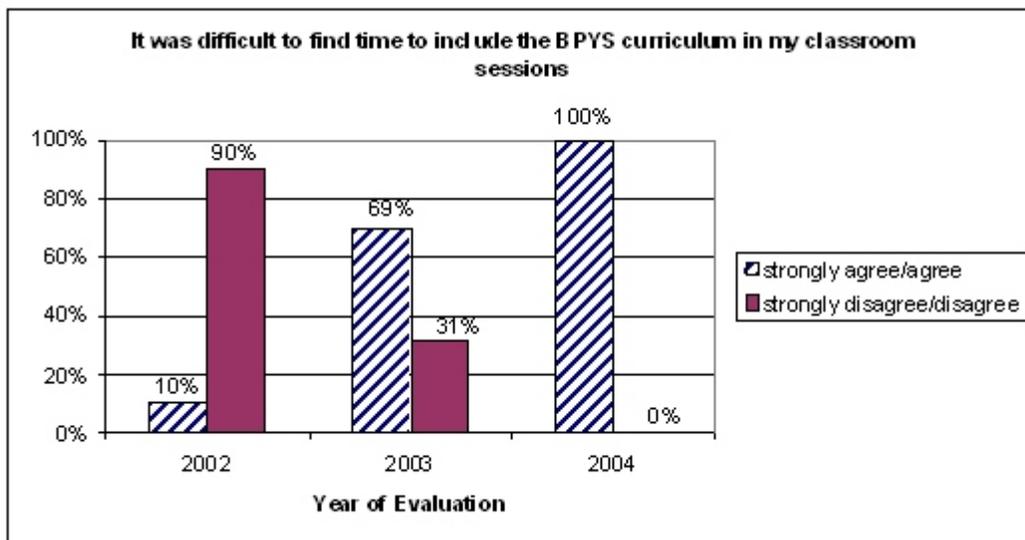
Overall, Chapman Middle School had great difficulty implementing the BPYS curriculum and with fidelity to the program as it was designed. Specifically, the absence of administrative support was an overwhelming barrier to successful implementation. Without full support from the principal, the teachers and cadre struggled to keep the program going throughout the evaluation. Additionally, the teachers and students reported that the curriculum presented to the middle school students was inappropriate for the students' ability level and was better suited for younger children. The school elected to depart from the program as designed and made some significant adaptations to the program in an effort to decrease resistance from students.



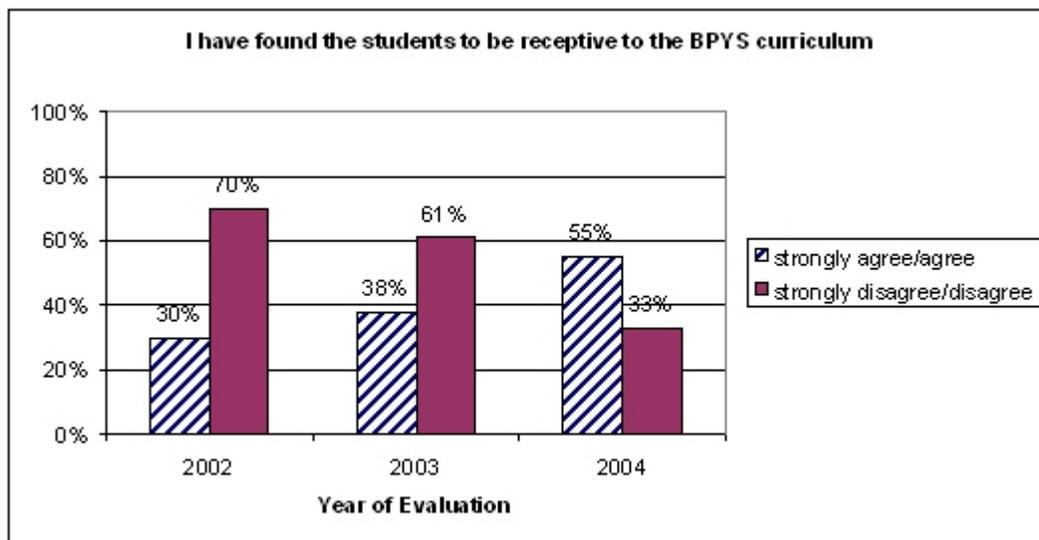
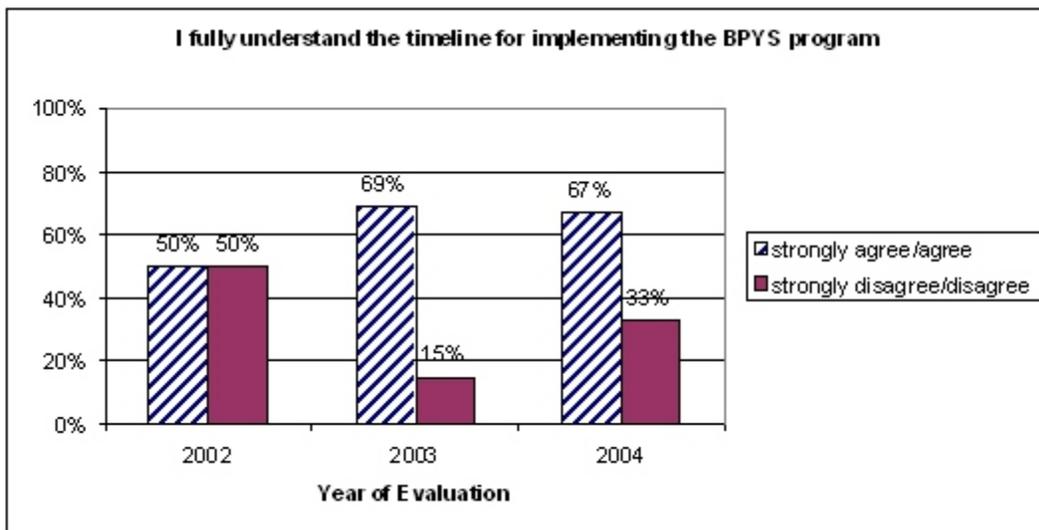
The administration at Chapman Middle School were overloaded with many responsibilities and as a result, finding time for the BPYS implementation proved challenging. During the first year of implementation, Chapman Middle School's principal was unable to find time to meet with the CSPV evaluator for the planned in person interview. Thus, the principal's insights regarding the first year of the implementation were lost, and this principal did not return the following year. Her successor's leadership style caused some uneasiness among the staff. The new principal's style of leadership was more directive and it was his preference that he handle discipline problems alone rather than collaborating with the school counselor. He elected not to join the cadre, though he did make some attempts to be involved with the BPYS curriculum and program implementation during his first year as principal. The administrative

instability contributed to low implementation fidelity, specifically in cadre functioning, staff training, and strategies for ongoing development of a caring community. Parent support was also consistently low throughout the evaluation.

It is important to note that the school counselor, who served as the coordinator for the BPYS implementation, was out of school with a serious illness during the final year of implementation. The principal did not assign someone to take her place as the coordinator of the cadre until late in the fall semester. By that point, the teachers were already behind the timeline in their implementation. Chapman Middle school only completed four of the eight hours of staff training in early October 2003 (the second year of implementation). Training for the new staff in 2004 took place on November 15th, which is several months later than is ideal for keeping to the implementation timeline. The difficulties in implementation for teachers at Chapman Middle School were illustrated in their responses to the teacher questionnaire. In general, the teachers thought that the training was helpful but there was a marked decrease in the number who felt helped by the training in the last year (i.e., the year that the principal reduced the amount of time available for training).

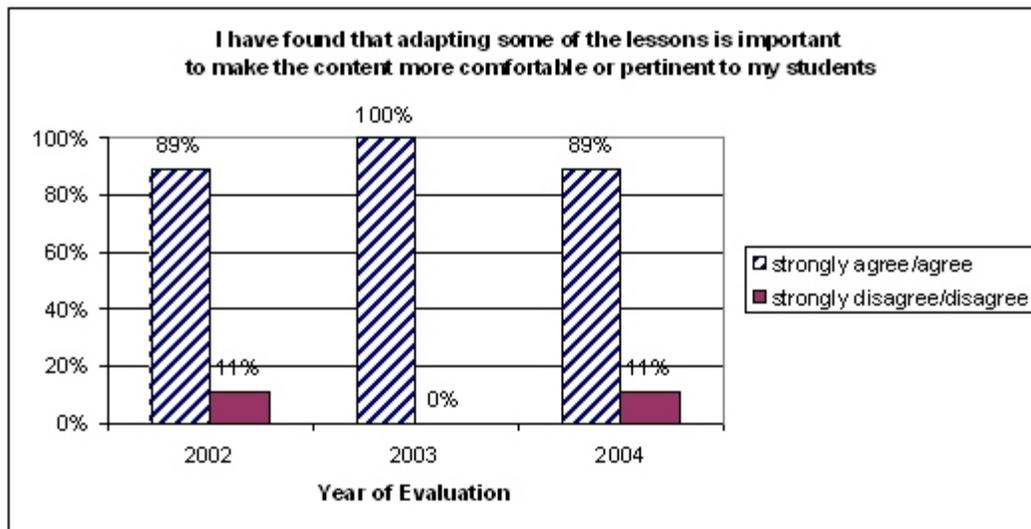


During the second and third years of implementation, larger percentages of teachers reported that they had a difficult time implementing the curriculum and that it was difficult to find the time to include the curriculum in classroom sessions. The new principal was only weakly supportive of the program and greatly limited the training and time available to teachers for implementation. Incongruously, teachers overall seemed to have a better understanding of the timeline for implementation in the last years of evaluation than they did in the first year of the program. The percentage of staff reporting their understanding of the timeline dipped slightly from year two to year three of evaluation. The limited and delayed training sessions and the absence of the school coordinator are two very strong possibilities for this decline.



The teachers reported that the students at Chapman were not very receptive to the program. Some of the reasons given by students were that the BPYS material was “juvenile” and their claim that “it doesn’t work”. They also questioned whether the program would have real world application. Teachers worked with the BPYS technical assistant to learn some

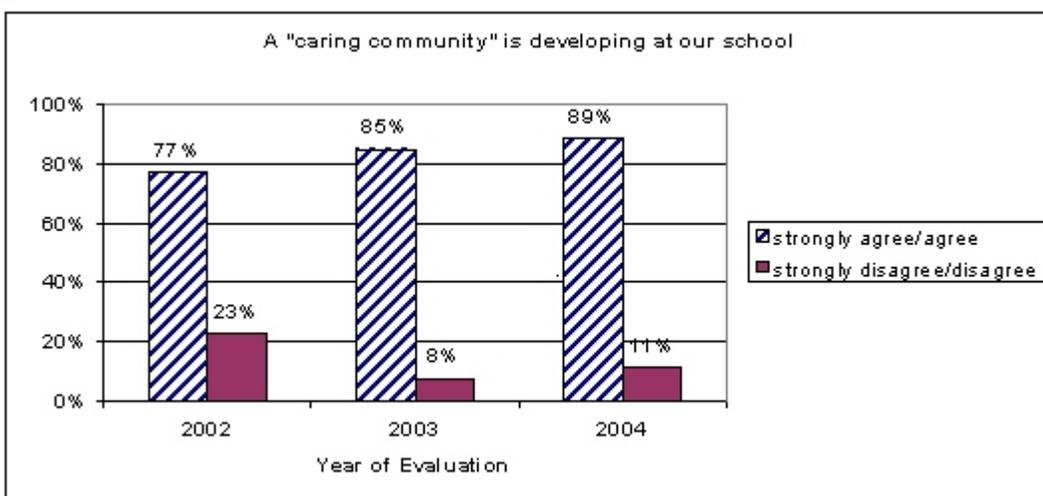
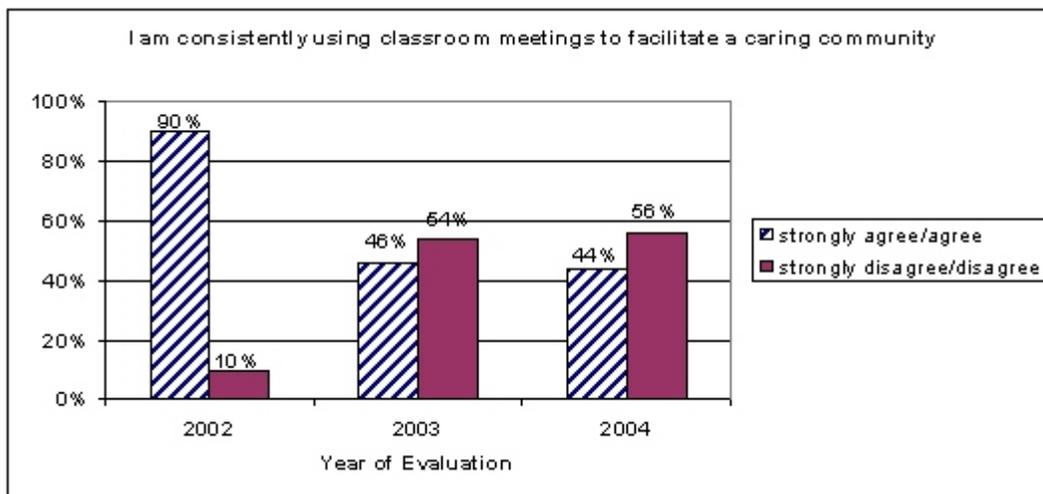
strategies for dealing with this persistent issue. The BPYS technical assistant hypothesized in her report that receptiveness of teachers was part of the problem, though that was not noted by CSPV.



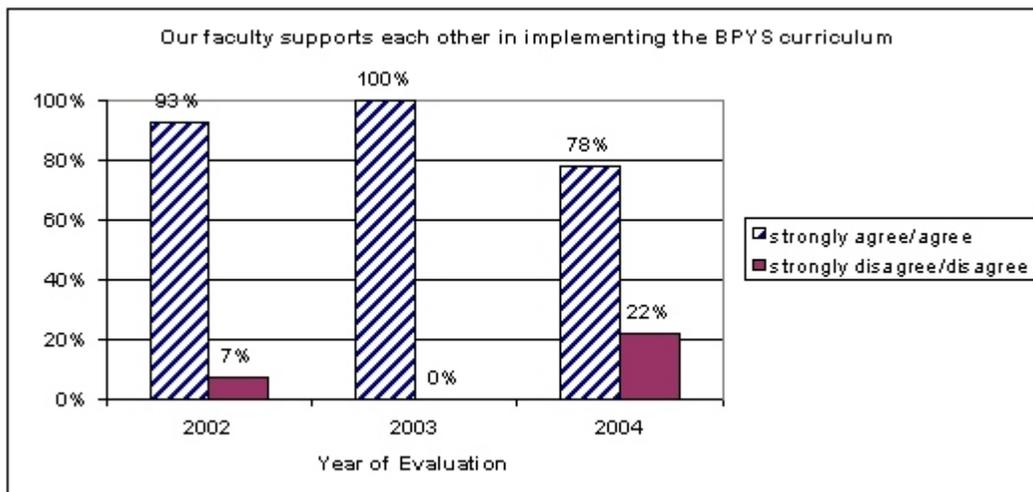
High percentages of teachers at Chapman Middle School reported that adapting the BPYS lessons was important in order to make the content more pertinent to their students. These adaptations were made in an attempt to get the students to connect with the curriculum and to address some of their complaints about the lessons. In particular, as a group, Chapman Middle School designed and implemented a substantial change in program delivery during the final year of evaluation. The 8th grade teachers finished presenting the lesson material to their students in the first two weeks of the school year. Upon completion of the lessons, the staff asked their students to brainstorm ideas for developing the “caring community” at their school. One of the primary ideas developed was for the 8th grade students to present the program lessons for the 6th grade classrooms. The idea behind this was that it would give the 8th grade students an opportunity to “own” their leadership role in the school and in addition utilize the influence they have over the younger students. The 8th grade students were broken into teams and assigned a lesson for which to prepare. Each group selected their own material (overheads, charts, handouts, etc) and practiced the way in which they would present the lesson. The 8th grade staff gave guidance and suggestions to the student groups when needed. The 8th grade students began presenting the lessons in the 6th grade classrooms in November. They were presented in a 35 minute time slot once a week. The CSPV evaluator observed once in each of the 6th grade classrooms during the fall semester and gave all of those sessions either medium or high ratings.

During the first year of evaluation, 93% of the teachers agreed that “School rules regarding bullying have been enforced.”. The next two years saw fewer teachers agreeing with that statement, with only 69% of teachers agreeing in 2003 and 76% agreeing in 2004. One factor that may influence these reports are the change in principal and adjusting to his discipline style.

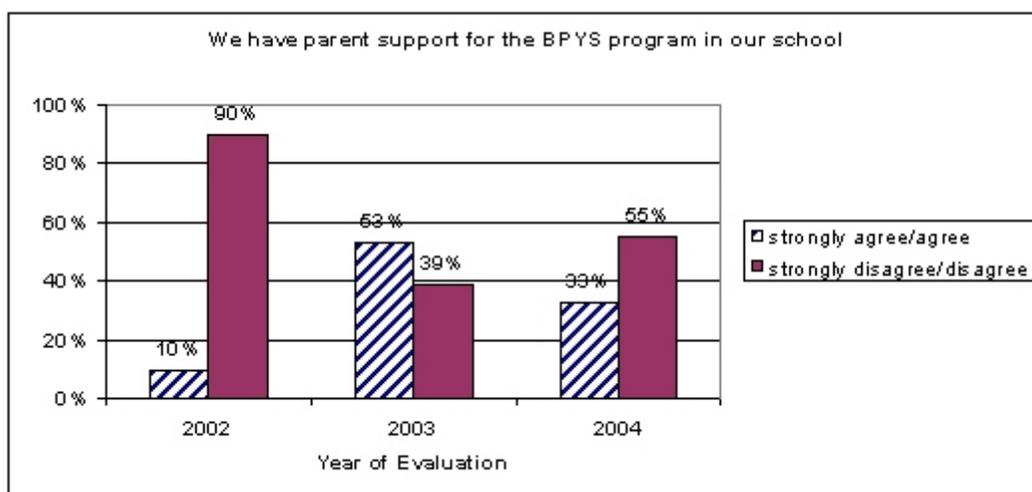
On a positive note, there was a steady increase in the percentage of teachers that agreed that “A caring community is developing at our school”. Given the struggles that Chapman Middle School faced when implementing this program, these numbers were surprising and encouraging. Interestingly, teachers felt that a “caring community” was developing at their school even though the percentage of teachers reporting consistently using the classroom meetings to facilitate a caring community declined each year.



In general, the faculty reported feeling supported by one another in implementing the BPYS program. During the second year of evaluation, 100% of teachers reported that the faculty supported each other. Though this was the first year for Chapman’s new principal, this is also the year in which the cadre and school coordinator were most active and involved. After meeting with other BPYS cadres at the regional cadre meeting in 2003, the Chapman Middle School cadre members realized that they were not doing enough to implement the program. When they attended the regional meeting in 2004, they were able to talk about all of the progress they had made and left the meeting with a feeling of accomplishment. These accomplishments were in no small part due to the cadre leader, the school counselor.



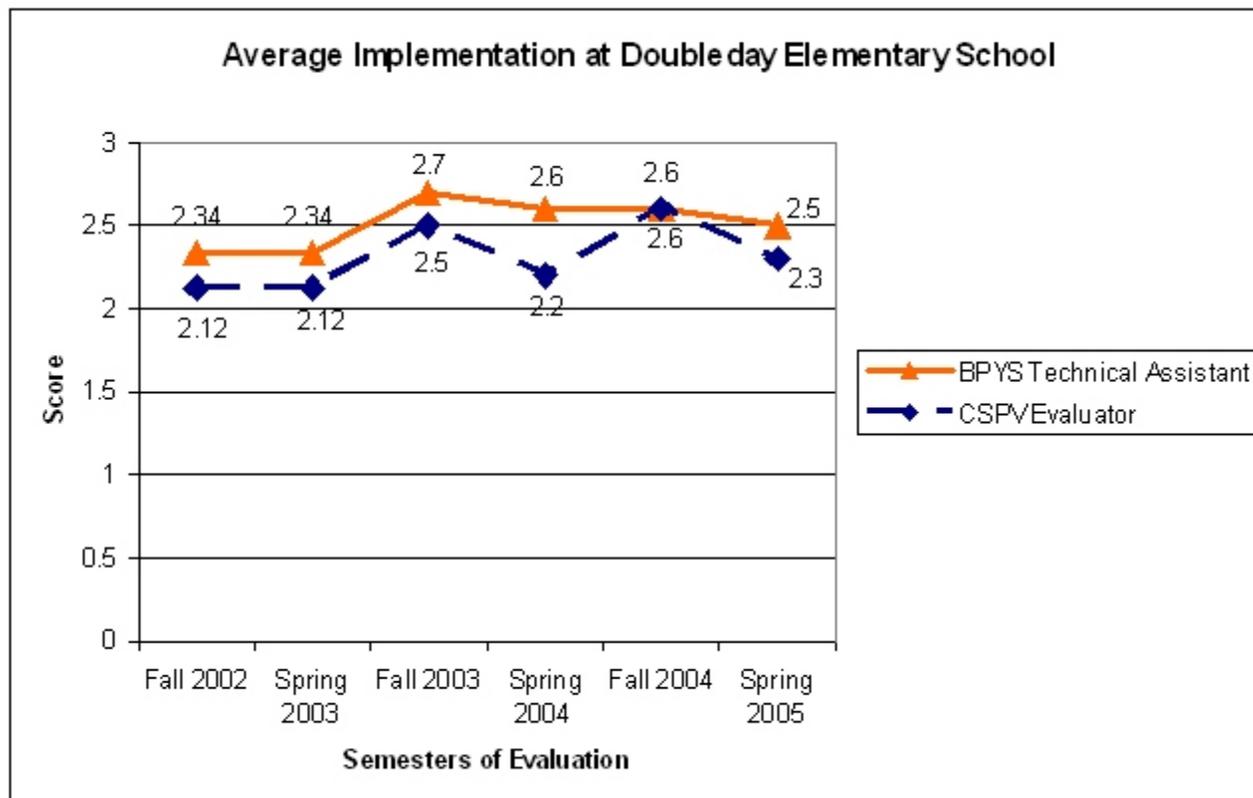
As the graph illustrates, in the final year the percentage of teachers reporting that they felt supported by faculty decreased from 100% to 78%. In the final year, both BPYS technical assistants and CSPV evaluators noticed that administrative support for the program was very weak. The principal assured the BPYS technical assistant that all cadre members would be able to attend a joint cadre meeting on the scheduled date, however, only two were able to attend. In addition, when it was clear that the school counselor was going to be unable to fulfill her duties as cadre coordinator due to illness, the principal did not take charge of the project nor did he inform her that he had cancelled the staff training at the beginning of the school year. Staff members at the school appeared to be aware of his weak support for the program. They reported that they were reluctant to communicate their concerns with the principal for fear of retaliation. The BPYS technical assistant noted that communication between the staff and the principal seemed strained, particularly surrounding discipline issues. The lack of administrative support undermined the staff's ability to implement the program and feel secure in their school environment. Another barrier to successful implementation was low parent support for the BPYS program. Over time, teachers reported more parent support than in the first year of evaluation. A likely explanation is that the school did not have a parent meeting about the program until the second year of evaluation.



Treatment School 2 (Block 1) Doubleday Elementary

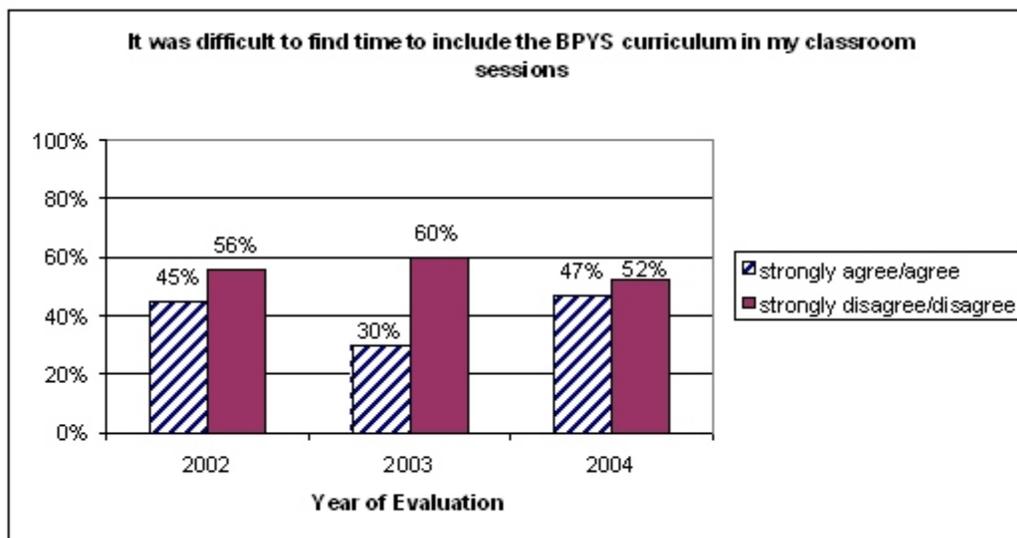
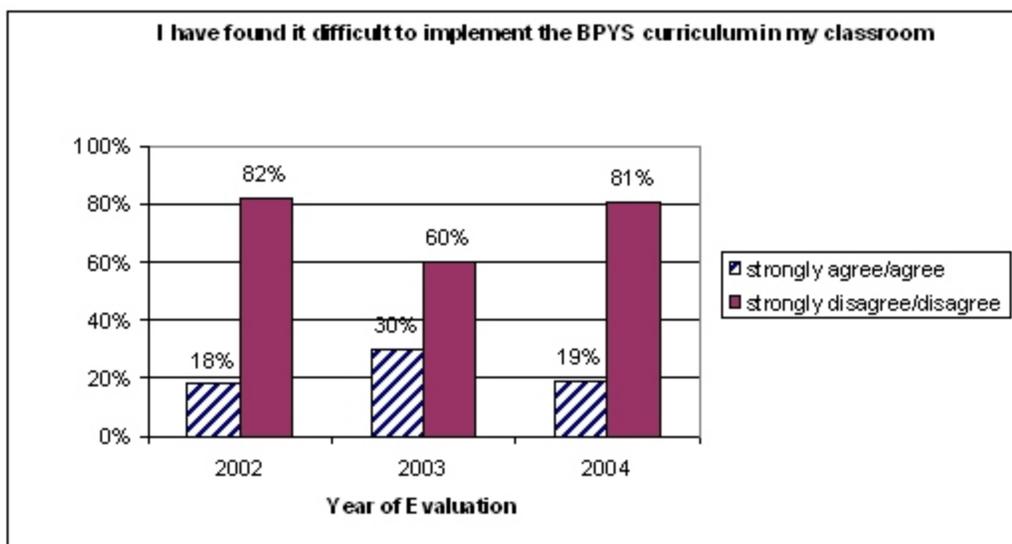
Overall, despite good intentions from the principal, the school counselor, and the teachers, the BPYS program at Doubleday Elementary never reached full implementation. The enthusiastic new principal struggled to implement a new discipline policy, the counselor struggled to increase the effectiveness of the cadre, and the teachers reported some difficulty implementing the program as it was designed. As a result, BPYS never became integrated into the school culture.

Doubleday Elementary began the 2002-2003 school year with a new principal. This new principal had previously been a teacher at Doubleday and his enthusiasm for the BPYS project was evident. In this first year as principal, he worked with the cadre to develop a new discipline policy for the school. Despite this effort, however, the process, detail, and follow-through on the new policy struggled during the first year, and the CSPV evaluator gave low fidelity ratings to the school on both “School wide discipline plan in place” and “Assessment of current school climate and safety issues.” Over time and with refinement, the policy became more effective. Low parent support also brought the fidelity ratings down for the first two years. During the second year, the cadre added a parent representative. In both the second and third year, the school hosted a parent night and workshops devoted to the BPYS program. This helped greatly to garner parent support for the program and raise Doubleday Elementary’s fidelity scores.



The staff at Doubleday were enthusiastic about the BPYS program after attending the training given by the BPYS technical assistants. Close to 100% of teachers agreed each year that the techniques provided during the training were helpful. Even though Doubleday only had four of the eight hours of training during the second year of evaluation, all of the teachers reported that the training was helpful. Despite this, the teachers continued to struggle with implementing the program in the classroom.

Although teachers found the training helpful, some reported difficulty implementing the lessons in their classrooms. One reason the teachers cited for this difficulty was finding the time to do the lessons. The principal agreed that everyone had a hard time keeping momentum because they were already so overloaded.

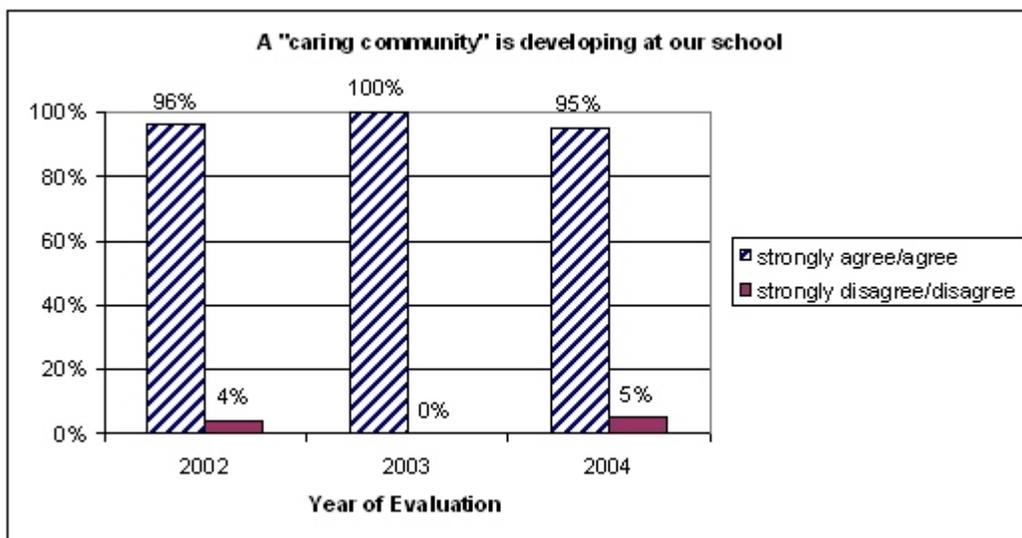


Interestingly, in 2002, 2003, and 2004, the percentage of teachers reporting consistently using the classroom meetings to facilitate a caring community was only 63%, 50%, and 85%, respectively. Even though not all of the teachers felt that they were implementing consistently,

they felt that in general a caring community was developing. In 2003, the school only completed half of the staff training and did not receive the 7 hours of on-site consultation from BPYS. The school is located a day's drive from the BPYS offices, which made it difficult for the BPYS technical assistant to make regular visits to the school. Moreover, the phone consultation also was lacking. Only one and a half hours of phone consultation out of the five hours allotted were delivered to the school.

Challenges to Implementation. Though the coordinator at Doubleday was extremely dedicated and organized, she had a difficult time getting the cadre working, due in part to a great deal of staff turnover between the first two years of implementation. In an effort to make the cadre more inclusive, the coordinator elected to add a parent and a bus driver to the cadre. The cadre helped to finalize the discipline and referral process and created a regular meeting schedule to work directly on BPYS issues. Though the principal and the cadre worked together to finalize the discipline plan in the second year of implementation, the percentage of teachers reporting that the School rules regarding bullying have not been enforced increased from 9% to 15% (this percentage decreased back to 5% in the final year).

During the final year of implementation, the coordinator was given other time intensive tasks that greatly limited the time she had to devote to the BPYS program. She made a strong effort to incorporate the BPYS lessons into her other tasks, but was unable to monitor or support the staff as they implemented the program. In addition, the cadre was comprised of all new members during the last year of implementation and they lacked experience and direction when it came to facilitating the implementation of the program. Despite struggling to implement the BPYS lessons, the teachers felt that a caring community was developing at their school.

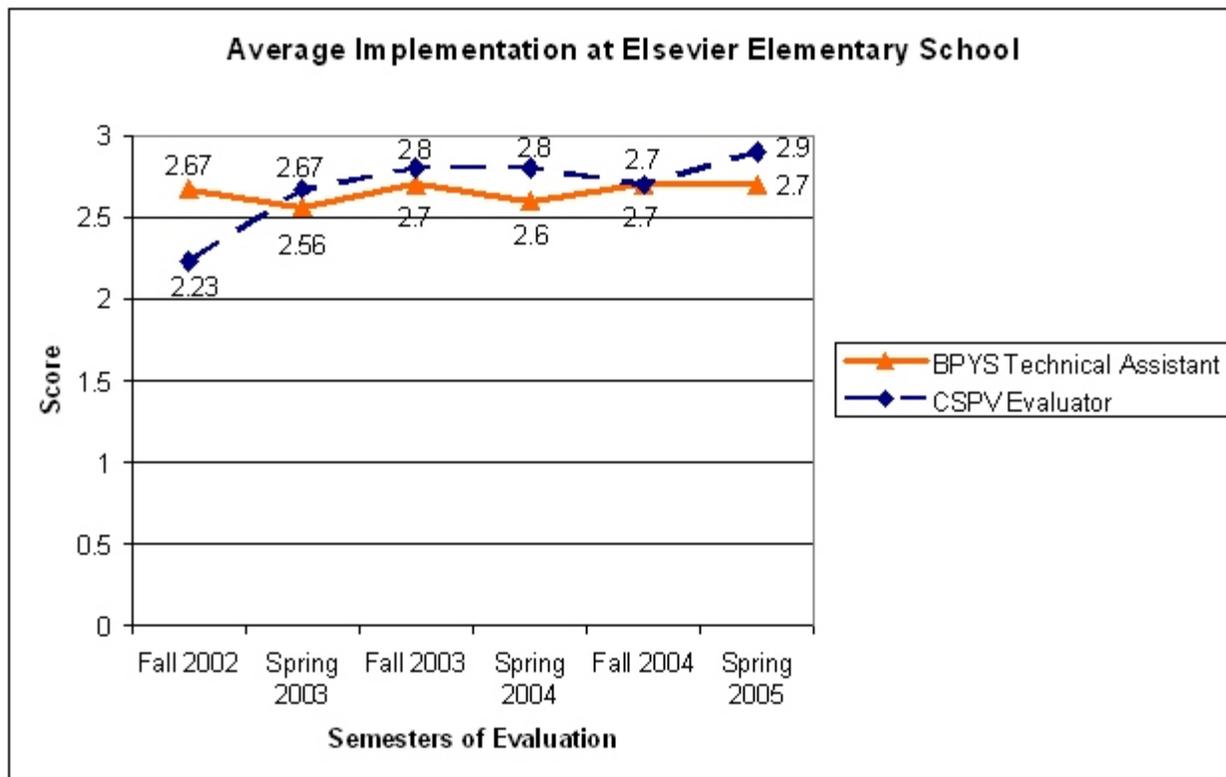


Another potent barrier to successful implementation at this school was difficulty providing the program to the high number of Spanish speaking students. Almost all of the teachers reported adapting the program to make the content more comfortable or pertinent to their students. Many of the lessons do not translate directly into Spanish, so these children were

either left out of the lesson or worked with differently. The publisher of the BPYS program is working on producing the BPYS lessons in Spanish, which will greatly help schools like Doubleday that have a substantial population of students whose primary language is Spanish. Lack of Spanish materials was also one of the reasons many teachers gave to explain student resistance to the curriculum. Resistance was reported by the most teachers in 2003, with 75% of the teachers reporting resistance from students, but this did not persist past the second year of implementation. Despite the fact that the school struggled to implement the program, 95% of the teachers said that they planned to continue to implement the BPYS program after the evaluation ended.

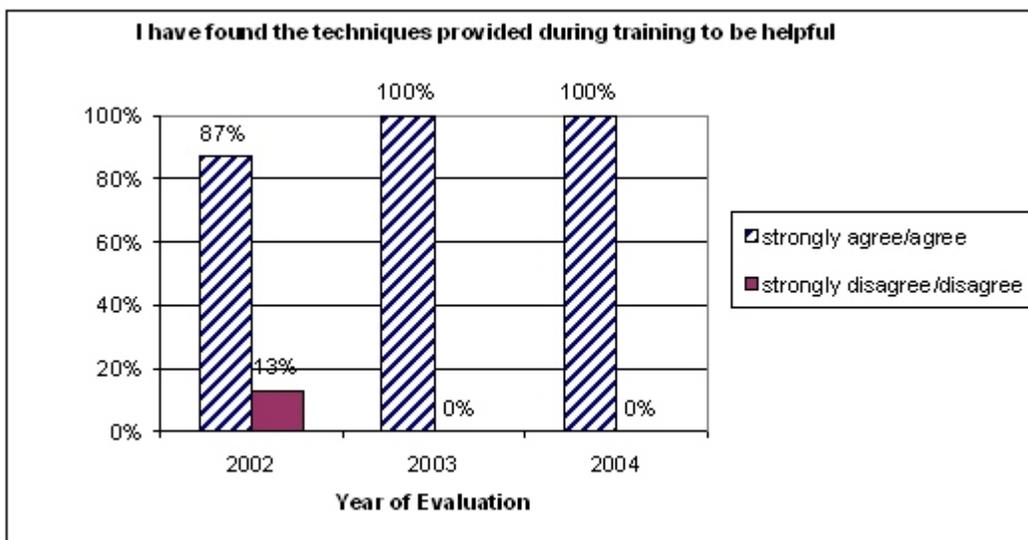
Treatment Elementary School 3 (Block 2)
Elsevier Elementary School

From the start, the staff at Elsevier Elementary has demonstrated a strong commitment to the Bully Proofing Your School (BPYS) Initiative. Before agreeing to implement the program, the principal polled her staff to make sure this was a program that they would want to implement. Throughout implementation, she continued to provide strong leadership to ensure that the program was implemented. Additionally, the school social worker contributed substantially to the implementation effort as the cadre leader.

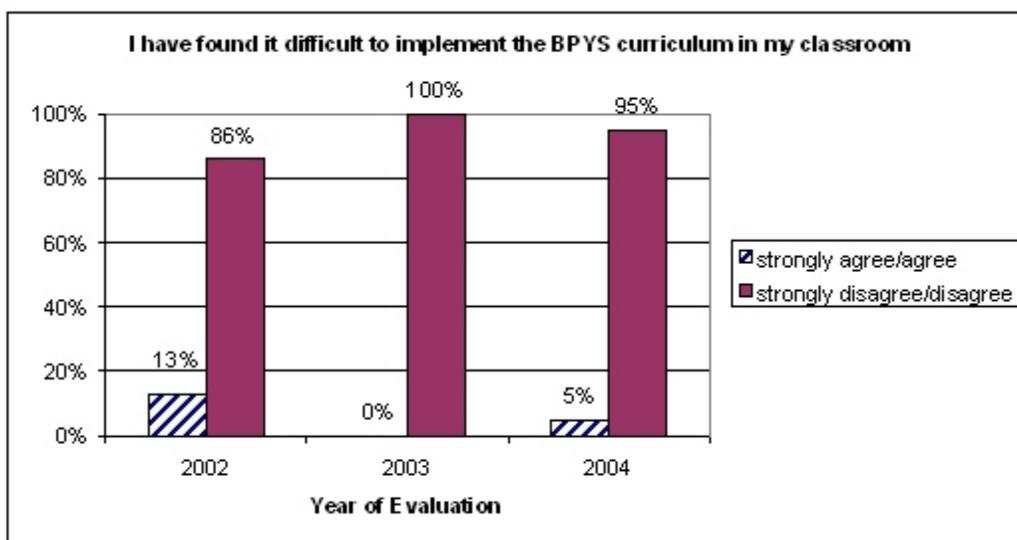


Throughout the program the staff and administrators at Elsevier seemed highly receptive to the training provided by the BPYS staff. On the teacher questionnaire, teachers were asked to give their opinion about the helpfulness of the BPYS training. Below is a graph of their responses over the course of the evaluation, which illustrates that teachers found the training extremely helpful (87% agreed in the first year and 100% agreed in the two following years).

Overall, the implementation at Elsevier Elementary School was excellent. The administration and staff immediately embraced the program and worked effectively to implement BPYS with high levels of fidelity. Because the program was consistently well-implemented and the fidelity improved gradually over time, this narrative will be less driven by a timeline and will focus instead on content.

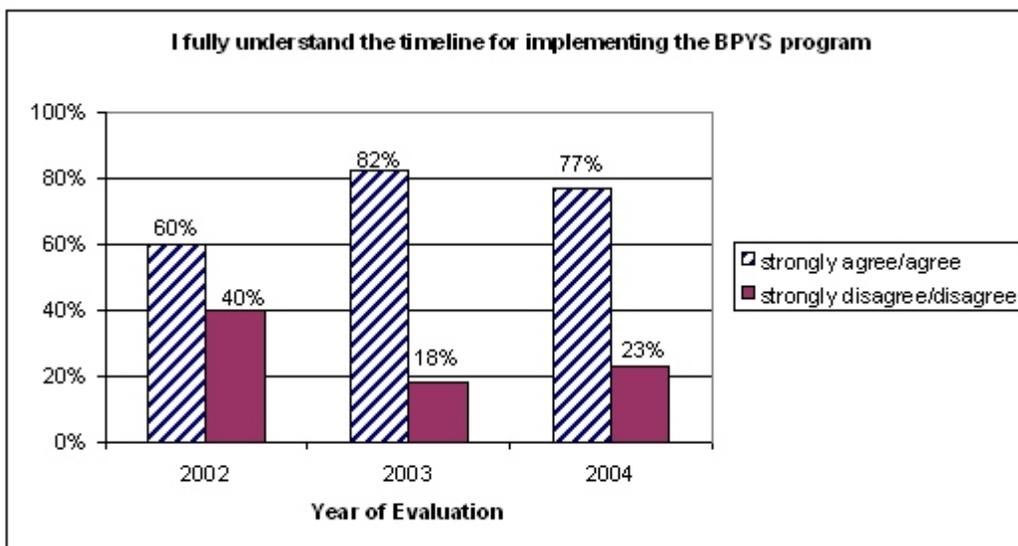
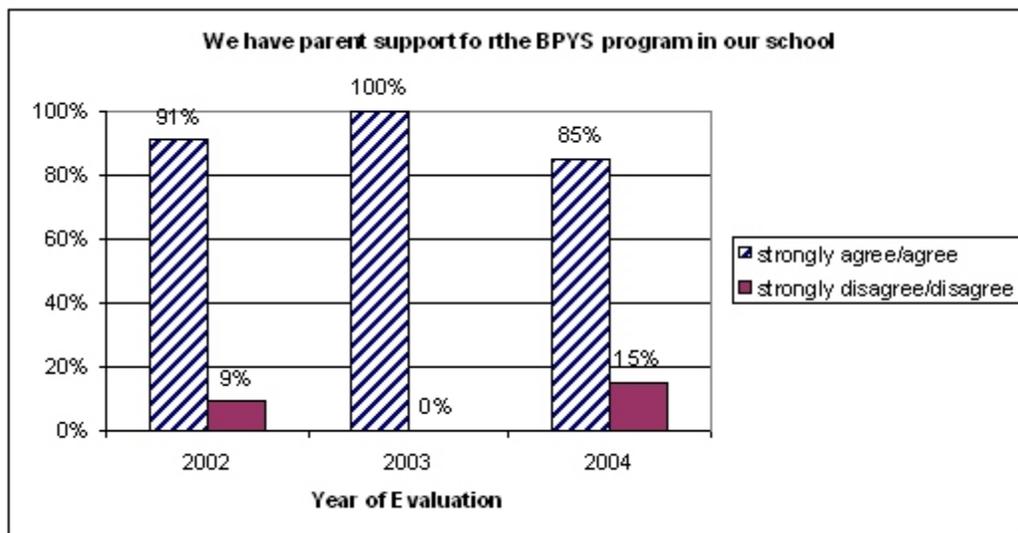


A cadre at Elsevier was already loosely formed prior to the beginning to the BPYS program due to the schools participation in other Safe Schools Initiatives. The group was fairly effective during the first year, but was reorganized for the second year so that it better represented teachers and para-professionals. The newly reorganized cadre was more active and effective in facilitating the implementation of the BPYS curriculum. The cadre scheduled weekly meetings, with the goal of providing consistent support to teachers. Results from the teacher questionnaire indicated that during the second year, teachers reported using the BPYS curriculum more regularly in the classroom. Teachers were asked if they felt it was difficult to implement the BPYS curriculum in the classroom. Below is a graph of their responses for each year of implementation, illustrating that the vast majority of the teachers did not find it difficult to implement the program. In the best year, there was complete agreement among the teachers on this point.

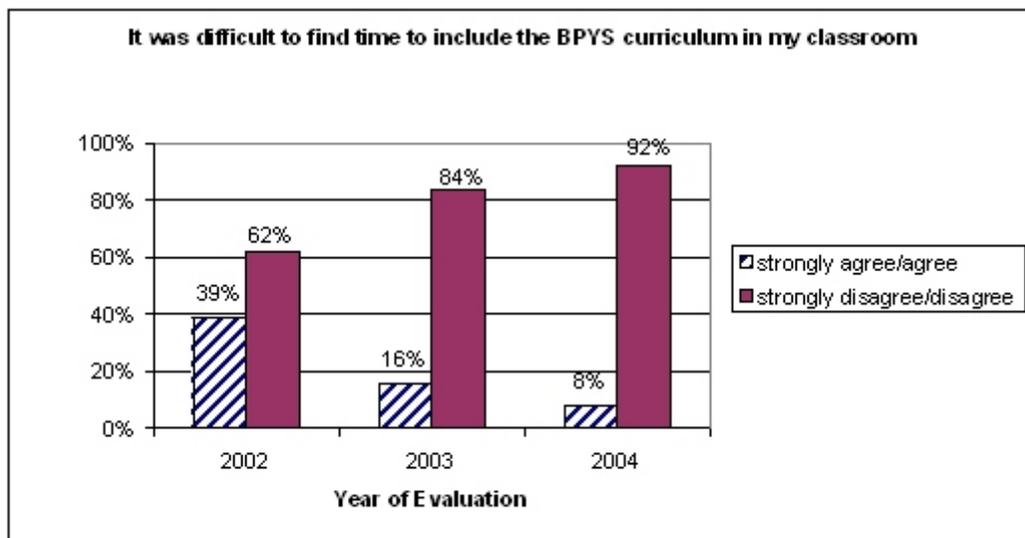


With the support of the cadre and the administration, the BPYS program became an integrated part of most of the classroom lessons. Over the course of the implementation, BPYS was seen less and less as an “add on” by teachers and more as an integral part of teaching. Teachers and administrators reported observing the students using the techniques and language of the BPYS program when resolving conflict. These reports were substantiated by the teacher questionnaire: In the last two years of implementation, 100% of teachers answering the questionnaire agreed with the statement, “A caring community is developing at our school”.

Another important component of Elsevier’s success was the support from students. 100% of the teachers responding to the teacher agreed with the statement, “I have found the students to be receptive to the BPYS curriculum.” In addition, most teachers reported having parent support for the program. Though the percentage of teachers reporting having support from parents is lower in the last year of implementation than in the first year, the percentage is still high.



Although overall Elsevier Elementary was given stellar ratings throughout the implementation of the BPYS program in their school, they did face several challenges to implementation. During the first year, many teachers did not understand the timeline for implementing the BPYS program in the first year. Only 60% agreed that they understood the timeline for implementation. Over time, more teachers gained an understanding of the timeline; however, by the last year of implementation, there were still 23% of teachers who were unclear about the timeline. Also, initially, 39% of the teachers said that it was difficult to find time to implement the BPYS program in their classroom sessions, but reports indicated that this difficulty decreased over time to only 8% in the final year of implementation.



Another challenge at Elsevier Elementary was scheduling consultation hours with the BPYS technical assistant. One potential explanation was that Elsevier was not experiencing problems with the BPYS program or its implementation, thus did not feel the need to have consultations with the technical assistant. The technical assistant did assist Elsevier with the staff training, back-to-school night, and the parent/stakeholder training.

The final challenge of note is that the teachers commented that the lessons were repetitive and that new ideas were needed (and the students noticed this repetitiveness). In addition, some teachers did not have the recommended literature, so they had to use other literature in its place. By the end of the implementation years, 83% of teachers reported that they found it necessary to adapt the BPYS curriculum to better suit their classroom. This included expanding the lessons into other areas of teaching, as well as designating part of the classroom as a “Peace Place”, where conflict can be worked out.

In conclusion, BPYS was well implemented by Elsevier Elementary School due to strong support by administration and staff. This sentiment is reflected in the following result: In the final year of conducting the teacher surveys, 100% of teachers indicated that they would continue to implement the BPYS program after the evaluation was completed and the technical support ended.

Part 4: ELEMENTARY SCHOOL OUTCOME EVALUATION

Results in this section are presented in five parts. First, the results are analyzed at the item level, examining treatment and comparison schools at baseline, in the three years of active program implementation, and in the post-implementation year, using all of the items considered in the evaluation, and adjusting the inferential statistics for multiple testing using a modified Bonferroni procedure, as described below. This provides the greatest detail on precisely where the program had or failed to have an impact. Second, the bivariate relationship between the intervention is presented for each of the multiple-item scales (described in the measurement section earlier) associated with the three main components of the program. Third, we use the average implementation scores presented in the previous section to see whether quality of implementation has an impact above and beyond the simple treatment-comparison contrast. Finally, we test the model presented in Figure 3, using the multiple-item scales associated with the major components of the program as outcome measures, but this time including additional controls for sociodemographic characteristics and the hypothesized intervening variables, peer group environment and one's own attitudes toward aggression and violence. As noted earlier, aggregate pretest characteristics and school characteristics are not directly included in the test model. Because these variables represent stable, aggregate characteristics of the school (in contrast to the process evaluation scores, which may vary considerably over time), aggregate pretest characteristics and school characteristics are collinear with the treatment-comparison distinction. Instead, we begin by testing for differences in aggregate pretest characteristics with respect to the variables used in this evaluation. To the extent that we are unable to reject the null hypothesis of no differences between treatment and comparison schools at baseline, we may conclude that any subsequent differences should be attributed to the intervention. Finally, we briefly consider the impact of BPYS on faculty and staff perceptions of school climate. It may be worth repeating here that the design of the data collection for the study is a repeated cross-sectional design; while there is overlap in respondents from one year to the next, this is not a longitudinal panel design, and hence techniques appropriate to longitudinal panel analysis are not applicable here.

Item-Level Results

Tables 10-12 present the item-level analysis of the impact of BPYS at the elementary school level. Table 10 compares all of the treatment schools with all of the comparison schools for each year of the study: the pre-implementation baseline year (2002), the three years of active implementation (2003-2005), and the post-implementation year (2006). Table 11 presents the same information for the first matched block of elementary schools (treatment schools Beacon and Doubleday, comparison schools Guilford and Harcourt), and Table 12 does the same for the second matched block (treatment school Elsevier and comparison school Ingram). Somers' d is used to assess the strength of the relationship between each of the items and the treatment/comparison distinction. Somers' d_{yx} and d_{xy} (generically Somers' d) are asymmetric measures designed for use when measurement is at least at the ordinal level (and a dichotomous variable like the treatment/comparison distinction can be treated as an ordinal variable), d_{yx} treating X as the predictor and Y as the dependent variable, and d_{xy} treating Y as the predictor and X as the dependent variable, having identical properties and nearly identical construction.

Table 10: Combined Treatment vs. Combined Comparison Schools, Individual Items

Variables	2002	2003	2004	2005	2006
* = Statistically significant at $\alpha=.05$ using Holm's sequential method for familywise statistical significance. These differences are statistically significant by the most conservative criteria used here.	Somers' d (p)/ Favorable for... (T=treatment, C=Comparison, - = Neither)	= Somers' d (p)/ Favorable for... (T=treatment, C=Comparison, - = Neither)			
A. School Climate					
1. I like school	-.001 (.975)/-	-.093 (.043)/T	-.136 (.001)//T *	-.049 (.251)/-	.108 (.014)/C
2. I look forward to going to school	.056 (.224)/-	-.065 (.161)/-	-.102 (.019)//T	-.050 (.249)/-	.054 (.228)/-
3. I try hard in school	.037 (.326)/-	-.058 (.126)/-	-.079 (.014)//T *	-.071 (.034)/T	.011 (.769)/-
4. My teacher tells me when I do good job	-.003 (.938)/-	-.075 (.096)/-	-.133 (.001)//T *	-.041 (.310)/-	.011 (.789)/-
5. My teacher listens to me...	.077 (.069)/(C)	-.052 (.226)/-	-.038 (.347)/-	.019 (.635)/-	.037 (.378)/-
6. My teacher cares about me	.031 (.446)/-	-.083 (.025)/T	-.108 (.004)/T *	-.077 (.039)T	-.063 (.112)/-
7. I like my teacher	.141 (.001)/C *	-.103 (.005)/T *	-.067 (.076)/(T)	-.089 (.018)/T *	.037 (.370)/-
8. Adults teach us not to pick on other students	.009 (.805)/-	-.127 (.000)/T *	-.118 (.000)/T *	-.092 (.002)/T *	-.039 (.232)/-
9. Adults try hard to prevent bullying	.026 (.503)/-	-.169 (.000)/T *	-.134 (.000)/T *	-.059 (.077)/(T)	-.024 (.494)/-
10. People here respect all races	-.043 (.358)/-	-.154 (.001)/T *	-.091 (.033)/T	-.165 (.000)/T *	-.091 (.037)/T
11 People of my race can succeed here	-.002 (.955)/-	.002 (.969)/-	-.023 (.547)/-	-.070 (.071)/(T)	-.013 (.736)/-
12. I feel lonely at school	.000 (.995)/-	.013 (.780)/-	.031 (.464)/-	.168 (.000)/T *	.014 (.737)/-
13. There is an adult I can trust	.024 (.518)/-	-.062 (.093)/(T)	-.073 (.038)/T	-.127 (.000)/T *	-.016 (.654)/-
14. I see graffiti here	-.020 (.663)/-	.151 (.001)/T *	.082 (.061)/(T)	-.088 (.045)/C	.049 (.283)/-
15. My school is clean	.094 (.039)/C	-.147 (.002)/T *	-.052 (.223)/-	-.045 (.277)/-	.011 (.804)/-
16. I like the way my school looks	.133 (.003)/C	-.079 (.089)/(T)	-.102 (.017)/T	-.016 (.695)/-	.058 (.184)/-
17. Students here know the rules here	-.007 (.877)/-	-.029 (.534)/-	-.122 (.005)/T *	-.075 (.073)/(T)	-.002 (.972)/-

Variables	2002	2003	2004	2005	2006
* = Statistically significant at $\alpha=.05$ using Holm's sequential method for familywise statistical significance. These differences are statistically significant by the most conservative criteria used here.	Somers' d (p)/ Favorable for... (T=treatment, C=Comparison, - = Neither)	= Somers' d (p)/ Favorable for... (T=treatment, C=Comparison, - = Neither)			
18. Rule breakers are treated the same	.115 (.013)/C	-.058 (.216)/-	-.069 (.107)/-	-.068 (.115)/-	.010 (.820)/-
19. Rules here are fair	.002 (.957)/-	-.104 (.019)/T	-.090 (.027)/T	-.045 (.261)/-	-.009 (.827)/-
20. Students help decide activities and rules	-.023 (.628)/-	-.086 (.079)/(T)	-.071 (.112)/-	-.101 (.021)/T *	-.008 (.864)/-
21. I care what teachers think of me	.009 (.830)/-	-.053 (.204)/-	-.039 (.329)/-	-.052 (.173)/-	.082 (.048)/C
22. I respect teachers here	.131 (.001)/C *	-.033 (.378)/-	-.017 (.618)/-	-.054 (.103)/-	.025 (.496)/-
23. I respect the principal here	.002 (.947)/-	-.023 (.498)/-	-.008 (.821)/-	-.044 (.164)/-	.146 (.000)/C *
B. School Safety: Attitudes and Aggressive Behavior (Perpetration, Victimization, and Witnessing)					
24. I feel safe at my school	.036 (.409)/-	-.052 (.247)/-	-.092 (.022)/T	-.145 (.000)/T *	-.119 (.004)/T *
25. I feel safe on the school bus	.001 (.987)/-	-.090 (.191)/-	-.068 (.325)/-	-.039 (.530)/-	.055 (.382)/-
26. I feel safe walking to school	-.143 (.029)/T	-.056 (.409)/-	-.174 (.003)/T *	.091 (.126)/-	-.027 (.682)/-
27. Ever stay away because unsafe at school	.037 (.180)/-	.082 (.004)//T *	.067 (.003)/T *	.093 (.000)/T *	.068 (.012)/T
28. Ever stay away unsafe on way to school	.033 (.170)/-	.045 (.099)/(T)	.054 (.020)/T	.079 (.001)/T *	.001 (.975)/-
29. I have a friend who cares about me	.013 (.726)/-	.000 (.999)/-	-.036 (.286)/-	-.049 (.161)/-	.030 (.416)/-
30. I get along with most kids here	.050 (.281)/-	-.019 (.687)/-	-.094 (.030)/T	-.086 (.043)/T	.038 (.391)/-
31. My friends think wrong to hit	.056 (.207)/-	-.068 (.137)/-	-.087 (.040)/T	-.140 (.001)/T *	-.011 (.789)/-
32. My friends think is OK to say mean things	.037 (.392)/-	.035 (.445)/-	.050 (.217)/-	.128 (.001)/T *	.025 (.553)/-
33. My friends think is OK to yell	.017 (.677)/-	-.008 (.849)/-	.070 (.070)/T	.144 (.000)/T *	.012 (.766)/-
34. My friends think is OK to push and shove	.053 (.191)/-	.090 (.034)//T	.042 (.282)/-	.174 (.000)/T *	.038 (.343)/-
35. My friends think sometimes must fight	.041 (.272)/-	-.005 (.907)/-	.025 (.487)/-	.105 (.002)/T *	.030 (.422)/-

Variables	2002	2003	2004	2005	2006
* = Statistically significant at $\alpha=.05$ using Holm's sequential method for familywise statistical significance. These differences are statistically significant by the most conservative criteria used here.	Somers' d (p)/ Favorable for... (T=treatment, C=Comparison, - = Neither)	= Somers' d (p)/ Favorable for... (T=treatment, C=Comparison, - = Neither)			
36. I saw someone physically attacked	.053 (.265)/-	-.046 (.333)/-	-.139 (.002)/T *	-.105 (.016)/T *	-.133 (.003)/T *
37. I saw someone teased in mean way	.009 (.842)/-	-.093 (.053)//T	-.098 (.028)/T	-.142 (.001)/T *	-.130 (.005)/T *
38. I saw someone threaten to hit	.021 (.650)/-	-.108 (.026)/T	-.041 (.352)/-	-.143 (.001)/T *	-.167 (.000)/T *
39. I pushed, shoved, hit, etc.	.096 (.032)/C	-.026 (.570)/-	-.054 (.176)/-	-.110 (.004)/T *	-.059 (.131)/-
40. I got into physical fight when angry	.073 (.049)/C	.031 (.406)/-	.005 (.887)/-	-.052 (.090)/(T)	.014 (.686)/-
41. I teased other students in a mean way	.071 (.100)/(C)	.086 (.057)/(C)	-.044 (.254)/-	-.129 (.000)/T *	-.009 (.815)/-
42. I lied about student to get student disliked	.017 (.624)/-	.051 (.145)/-	-.096 (.001)/T *	-.079 (.010)/T *	-.046 (.117)/-
43. I tried to exclude others from my group	.086 (.057)/(C)	.032 (.489)/-	-.083 (.036)/T	-.153 (.000)/T *	-.015 (.711)/-
44. I said mean things to get student disliked	.058 (.114)/-	.041 (.255)/-	-.038 (.215)/-	-.078 (.011)/T *	-.036 (.229)/-
45. I threatened to hit or hurt student	.098 (.008)/C	.046 (.214)/-	-.055 (.081)/(T)	-.075 (.011)/T *	-.035 (.246)/-
46. I got pushed, shoved, slapped, or kicked	-.028 (.561)/-	.033 (.504)/-	-.071 (.108)/-	-.179 (.000)/T *	-.054 (.235)/-
47. I got teased in a mean way	.012 (.795)/-	.023 (.647)/-	-.105 (.013)/T *	-.058 (.177)/-	-.045 (.319)/-
48. Student lied about me to get me disliked	-.031 (.509)/-	-.003 (.949)/-	-.132 (.002)/T *	-.098 (.020)/T *	-.103 (.020)/T
49. Student tried to keep me out of group	.066 (.156)/-	.060 (.221)/-	-.080 (.057)/(T)	.020 (.636)/-	-.089 (.039)/T
50. Student said mean things to get me disliked	-.014 (.749)/-	.001 (.976)/-	-.108 (.009)/T *	-.061 (.138)/-	-.060 (.154)/-
51. Student threatened to hit or hurt me	.036 (.429)/-	.050 (.282)/-	-.057 (.161)/-	-.107 (.008)/T *	-.125 (.002)/T *
52. ...threatened to hit or hurt me: lunchroom	.029 (.399)/-	.041 (.249)/-	-.031 (.278)/-	.018 (.550)/-	-.033 (.283)/-
53. ...threatened to hit or hurt me: playground	-.003 (.945)/-	.022 (.630)/-	-.043 (.291)/-	-.109 (.005)/T *	-.128 (.001)/T *
54. ...threatened to hit or hurt me: bathroom	-.003 (.921)/-	-.032 (.212)/-	-.030 (.261)/-	-.051 (.033)/T	-.052 (.061)/(T)

Variables	2002	2003	2004	2005	2006
* = Statistically significant at $\alpha=.05$ using Holm's sequential method for familywise statistical significance. These differences are statistically significant by the most conservative criteria used here.	Somers' d (p)/ Favorable for... (T=treatment, C=Comparison, - = Neither)	= Somers' d (p)/ Favorable for... (T=treatment, C=Comparison, - = Neither)			
55. ...threatened to hit or hurt me: hallway	.036 (.290)/-	-.006 (.860)/-	-.004 (.893)/-	-.046 (.084)/(T)	-.021 (.490)/-
56. ...threatened to hit or hurt me: bus	.047 (.368)/-	.015 (.780)/-	.063 (.166)/-	-.034 (.487)/-	-.104 (.017)/T
57. How many students get picked on	.022 (.640)/-	-.107 (.024)//T	-.123 (.006)/T *	-.129 (.003)/T *	-.086 (.051)/(T)
58. How many students pick on others	.001 (.976)/-	-.103 (.033)//T	-.082 (.074)/(T)	-.203 (.000)/T *	-.173 (.000)/T *
59. How many kids afraid of you b/c mean	.015 (.706)/-	-.062 (.126)/-	-.107 (.003)/T *	-.095 (.008)/T *	-.025 (.512)/-
60. How many kids do you pick on often	.092 (.020)/C	-.041 (.284)/-	-.093 (.007)/T *	-.128 (.000)/T *	-.032 (.374)/-
61. How many kids pick on you often	.019 (.681)/-	.023 (.642)/-	-.005 (.905)/-	-.049 (.253)/-	-.071 (.106)/-
62. How many kids do you fear b/c mean	-.069 (.112)/-	-.009 (.856)/-	-.019 (.645)/-	-.091 (.023)/T *	-.169 (.000)/T *
63. I think it is wrong to hit other people	.047 (.210)/-	-.028 (.479)/-	-.062 (.085)/(T)	-.108 (.002)/T *	-.017 (.637)/-
64. It is OK to push and shove if you are mad	.023 (.486)/-	.063 (.073)/(T)	.135 (.000)/T *	.069 (.040)/T	.048 (.149)/-
65. It is OK to say mean things if you are angry	.029 (.383)/-	.010 (.794)/-	.125 (.000)/T *	.115 (.000)/T *	.062 (.064)/(T)
66. It is OK to yell and say bad things to others	.017 (.608)/-	.057 (.100)/(T)	.086 (.004)/T *	.128 (.000)/T *	.017 (.596)/-
67. It is wrong to get into physical fights	.025 (.572)/-	-.128 (.004)//T *	-.063 (.133)/-	-.020 (.647)/-	-.072 (.088)/(T)
68. Sometimes must fight to get what you want	.001 (.985)/-	.127 (.001)//T *	.075 (.035)/T	.096 (.004)/T *	.060 (.104)/-
69. It's OK to hit if they hit you first	.128 (.004)/T	.119 (.010)//T	.100 (.020)/T	.137 (.001)/T *	.070 (.102)/-
C. Home and Family Environment					
70. My parents want me to get good grades	-.022 (.262)/-	.002 (.944)/-	.011 (.628)/-	-.017 (.338)/-	-.013 (.464)/-
71. I can tell my parents how I feel	-.072 (.053)/(T)	.008 (.856)/-	-.163 (.000)/T *	-.187 (.000)/T *	-.010 (.796)/-
72. I like to do things with my family	.009 (.728)/-	-.011 (.728)/-	-.022 (.408)/-	-.093 (.005)/T *	-.001 (.980)/-

Variables	2002	2003	2004	2005	2006
* = Statistically significant at $\alpha=.05$ using Holm's sequential method for familywise statistical significance. These differences are statistically significant by the most conservative criteria used here.	Somers' d (p)/ Favorable for... (T=treatment, C=Comparison, - = Neither)	= Somers' d (p)/ Favorable for... (T=treatment, C=Comparison, - = Neither)			
73. Parents know who I am with if I'm away	.036 (.291)/-	.005 (.884)/-	.001 (.975)/-	-.056 (.068)/(T)	-.041 (.180)/-
74. Parents limit how much TV I watch	-.042 (.379)	-.036 (.459) /-	-.005 (.908)/-	-.068 (.123)/-	.021 (.630)/-
75. Parents limit what kind of music I listen to	-.031 (.507)/-	-.042 (.392)/-	.038 (.388)/-	-.106 (.016)/T *	-.078 (.083)/(T)
76. Parents know who my friends are	-.029 (.374)/-	-.072 (.028)//T	-.085 (.009)/T *	-.022 (.528)/-	-.018 (.617)/-
77. Parents let me know if I do a good job	.015 (.668)/-	.018 (.628)/-	-.042 (.183)/-	-.032 (.326)/-	.002 (.944)/-
78. I share thoughts and feelings with parents	-.029 (.519)/-	.006 (.901)/-	-.142 (.000)/T *	-.156 (.000)/T *	-.007 (.866)/-
79. Most days I spend some time with parents	.014 (.737)/-	.012 (.787)/-	-.032 (.408)/-	-.069 (.079)/(T)	.097 (.017)/C
80. My parents always want me to do my best	-.008 (.668)/-	-.027 (.262)/-	-.005 (.813)/-	-.015 (.392)/-	.030 (.139)/-
81. There will always be people I can count on	-.002 (.955)/-	.052 (.096)/(T)	-.040 (.147)/-	-.085 (.000)/T *	.008 (.813)/-
82. Besides family there is an adult I can trust	-.058 (.098)/(T)	-.040 (.281)/-	-.016 (.656)/-	-.122 (.000)/T *	-.048 (.717)/-
83. I believe there is some good in everybody	.024 (.496)/-	-.026 (.477)/-	-.019 (.592)/-	-.010 (.785)/-	.073 (.056)/(C)

Table 11: Block 1 Treatment vs. Block 1 Comparison Schools, Individual Items

Variables	2002	2003	2004	2005	2006
* = Statistically significant at $\alpha=.05$ using Holm's sequential method for familywise statistical significance. These differences are statistically significant by the most conservative criteria used here.	Somers' d (p)/ Favorable for... (T=treatment, C=Comparison, - = Neither)				
A. School Climate					
1. I like school	.015 (.797)/-	-.028 (.632)/-	-.112 (.034)/T	-.026 (.623)/-	.238 (.000)/C *
2. I look forward to going to school	.080 (.189)/-	-.064 (.262)/-	-.028 (.609)/-	-.017 (.757)/-	.183 (.001)/C *
3. I try hard in school	-.013 (.784)/-	-.098 (.028)/T	-.072 (.084)/(T)	-.108 (.007)/T *	.045 (.653)/-
4. My teacher tells me when I do good job	.030 (.596)/-	-.074 (.180)/-	-.126 (.010)T	-.101 (.043)/T	.100 (.061)/(C)
5. My teacher listens to me...	.075 (.170)/-	-.021 (.689)/-	.000 (.992)/-	.025 (.618)/-	.119 (.023)/C
6. My teacher cares about me	.010 (.851)/-	-.075 (.116)/-	-.063 (.167)/-	-.121 (.010)/T *	.013 (.801)/-
7. I like my teacher	.138 (.008)/C	-.084 (.059)/(T)	-.031 (.496)/-	-.083 (.085)/(T)	.102 (.052)/(C)
8. Adults teach us not to pick on other students	-.034 (.463)/-	-.079 (.065)/(T)	-.101 (.009)T	-.065 (.088)/(T)	.038 (.373)/-
9. Adults try hard to prevent bullying	-.026 (.576)/-	-.139 (.001)/T *	-.118 (.002)T *	-.001 (.982)/-	.057 (.203)/-
10. People here respect all races	.031 (.612)/-	-.081 (.073)/(T)	-.016 (.764)/-	-.168 (.002)/T *	.005 (.927)/-
11 People of my race can succeed here	.017 (.780)/-	.027 (.638)/-	.008 (.873)/-	-.120 (.012)/T *	.071 (.162)/-
12. I feel lonely at school	.088 (.126)/-	-.001 (.980)/-	.045 (.388)/-	.205 (.000)/T *	-.004 (.947)/-
13. There is an adult I can trust	-.022 (.631)/-	-.053 (.251)/-	-.041 (.354)/-	-.124 (.003)/T *	.053 (.268)/-
14. I see graffiti here	.191 (.001)/T *	.258 (.000)/C *	.058 (.277)/-	-.133 (.019)/C	.070 (.213)/-
15. My school is clean	-.033 (.578)/-	-.152 (.007)/T	.000 (.995)/-	-.045 (.390)/-	.072 (.194)/-
16. I like the way my school looks	.048 (.402)/-	-.014 (.809)/-	-.075 (.167)/-	.021 (.692)/-	.164 (.003)/C *
17. Students here know the rules here	-.022 (.712)/-	.018 (.764)/-	-.052 (.339)/-	-.069 (.205)/-	.119 (.030)/C

Variables	2002	2003	2004	2005	2006
* = Statistically significant at $\alpha=.05$ using Holm's sequential method for familywise statistical significance. These differences are statistically significant by the most conservative criteria used here.	Somers' d (p)/ Favorable for... (T=treatment, C=Comparison, - = Neither)				
18. Rule breakers are treated the same	.097 (.107)/-	-.053 (.367)/-	.054 (.325)/-	-.051 (.354)/-	.126 (.024)/C
19. Rules here are fair	-.102 (.066)/(T)	-.097 (.074)(T)	-.052 (.315)/-	-.021 (.692)/-	.131 (.016)C
20. Students help decide activities and rules	.044 (.480)/-	-.041 (.500)/-	-.065 (.240)/-	-.042 (.456)/-	.068 (.229)/-
21. I care what teachers think of me	.016 (.765)/-	-.029 (.583)/-	-.066 (.177)/-	-.068 (.156)/-	.097 (.066)/(C)
22. I respect teachers here	.131 (.010)/C	.001 (.977)/-	-.006 (.896)/-	-.045 (.286)/-	.158 (.001)/C *
23. I respect the principal here	-.060 (.075)/(T)	.030 (.499)/-	.061 (.175)/-	.052 (.239)/-	.322 (.000)/C *
B. School Safety: Attitudes and Aggressive Behavior (Perpetration, Victimization, and Witnessing)					
24. I feel safe at my school	-.008 (.891)/-	-.006 (.910)/-	-.017 (.737)/-	-.123 (.013)/T *	-.038 (.486)/-
25. I feel safe on the school bus	-.063 (.408)/-	-.076 (.338)/-	.018 (.799)/-	-.072 (.331)/-	.080 (.280)/-
26. I feel safe walking to school	-.170 (.068)/(T)	.012 (.891)/-	-.153 (.041)/T	.080 (.291)/-	.038 (.654)/-
27. Ever stay away because unsafe at school	.088 (.007)/T	.065 (.079)(T)	.090 (.001)/T *	.078 (.032)/T	.074 (.042)/T
28. Ever stay away unsafe on way to school	.047 (.138)/-	.027 (.462)/-	.066 (.027)/T	.099 (.001)/T *	.025 (.430)/-
29. I have a friend who cares about me	-.012 (.811)/-	.012 (.802)/-	.002 (.967)/-	-.043 (.337)/-	.050 (.294)/-
30. I get along with most kids here	.129 (.031)/C	.018 (.765)/-	-.036 (.516)/-	-.069 (.213)/-	.118 (.033)/C
31. My friends think wrong to hit	.105 (.077)/(C)	.024 (.685)/-	-.065 (.216)/-	-.140 (.007)/T *	.032 (.550)/-
32. My friends think is OK to say mean things	-.011 (.857)/-	-.037 (.528)/-	.004 (.937)/-	.085 (.106)/-	-.039 (.465)/-
33. My friends think is OK to yell	.008 (.883)/-	-.108 (.061)/(C)	.040 (.417)/-	.182 (.000)/T *	-.062 (.228)/-
34. My friends think is OK to push and shove	.079 (.131)/-	.013 (.809)/-	.001 (.981)/-	.189 (.000)/T *	-.049 (.347)/-
35. My friends think sometimes must fight	.018 (.708)/-	-.070 (.190)/-	-.002 (.973)/-	.134 (.002)/T *	-.039 (.426)/-

Variables	2002	2003	2004	2005	2006
* = Statistically significant at $\alpha=.05$ using Holm's sequential method for familywise statistical significance. These differences are statistically significant by the most conservative criteria used here.	Somers' d (p)/ Favorable for... (T=treatment, C=Comparison, - = Neither)				
36. I saw someone physically attacked	.097 (.114)/-	-.025 (.674)/-	-.046 (.406)/-	-.143 (.010)/T *	.056 (.320)/-
37. I saw someone teased in mean way	.098 (.104)/-	-.132 (.027)/T	-.023 (.685)/-	-.172 (.002)/T *	-.046 (.424)/-
38. I saw someone threaten to hit	.042 (.483)/-	-.067 (.275)/-	.078 (.150)/-	-.186 (.001)/T *	.022 (.702)/-
39. I pushed, shoved, hit, etc.	.186 (.002)/C	.011 (.844)/-	-.030 (.550)/-	-.083 (.094)/(T)-	.059 (.241)/-
40. I got into physical fight when angry	.130 (.013)/C	.047 (.328)/-	.022 (.617)/-	-.047 (.253)/-	.107 (.021)/C
41. I teased other students in a mean way	.169 (.005)/C	.134 (.022)/C	-.009 (.855)/-	-.124 (.008)/T *	.027 (.582)/-
42. I lied about student to get student disliked	.012 (.791)/-	.073 (.113)/-	-.071 (.052)/(T)	-.110 (.005)/T *	-.062 (.097)/(T)
43. I tried to exclude others from my group	.148 (.015)/C	.115 (.055)/(C)	-.015 (.770)/-	-.138 (.005)/T *	.048 (.366)/-
44. I said mean things to get student disliked	.075 (.129)/-	.076 (.108)/-	-.015 (.703)/-	-.084 (.031)T	-.021 (.585)/-
45. I threatened to hit or hurt student	.170 (.002)/C *	.071 (.143)/-	-.048 (.226)/-	-.076 (.053)/(T)	.033 (.400)/-
46. I got pushed, shoved, slapped, or kicked	.039 (.555)/-	.061 (.324)/-	-.012 (.824)/-	-.171 (.002)/T *	.089 (.110)/-
47. I got teased in a mean way	.036 (.568)/-	.050 (.421)/-	-.071 (.183)/-	-.076 (.162)/-	-.003 (.961)/-
48. Student lied about me to get me disliked	-.015 (.816)/-	-.022 (.715)/-	-.092 (.088)/(T)	-.129 (.018)/T *	-.048 (.392)/-
49. Student tried to keep me out of group	.092 (.129)/-	.079 (.199)/-	-.069 (.186)/-	.017 (.750)/-	-.039 (.465)/-
50. Student said mean things to get me disliked	-.004 (.952)/-	.014 (.812)/-	-.083 (.114)/-	-.105 (.043)/T	-.018 (.725)/-
51. Student threatened to hit or hurt me	.063 (.305)/-	.074 (.202)/-	-.012 (.786)/-	-.105 (.040)/T	-.038 (.464)/-
52. ...threatened to hit or hurt me: lunchroom	.025 (.601)/-	.103 (.029)/C	-.003 (.942)/-	-.023 (.523)/-	-.023 (.529)/-
53. ...threatened to hit or hurt me: playground	.085 (.156)/-	.087 (.139)/-	.042 (.415)/-	-.088 (.076)/(T)	-.032 (.524)/-
54. ...threatened to hit or hurt me: bathroom	.047 (.292)/-	.005 (.889)/-	.031 (.400)/-	-.089 (.005)/T *	-.020 (.602)/-

Variables	2002	2003	2004	2005	2006
* = Statistically significant at $\alpha=.05$ using Holm's sequential method for familywise statistical significance. These differences are statistically significant by the most conservative criteria used here.	Somers' d (p)/ Favorable for... (T=treatment, C=Comparison, - = Neither)				
55. ...threatened to hit or hurt me: hallway	.061 (.189)/-	.041 (.341)/-	.041 (.283)/-	-.090 (.007)/T *	.017 (.665)/-
56. ...threatened to hit or hurt me: bus	.038 (.543)/-	.034 (.598)/-	.163 (.007)/C	-.046 (.446)/-	-.064 (.225)/-
57. How many students get picked on	.075 (.218)/-	-.054 (.362)/-	-.017 (.751)/-	-.189 (.001)/T *	.033 (.540)/-
58. How many students pick on others	.072 (.240)/-	-.059 (.338)/-	.073 (.197)/-	-.229 (.000)/T *	-.095 (.076)/(T)
59. How many kids afraid of you b/c mean	.060 (.279)/-	-.056 (.280)/-	-.122 (.008)/T	-.128 (.008)/T *	.009 (.849)/-
60. How many kids do you pick on often	.126 (.020)/C	-.002 (.971)/-	-.057 (.207)/-	-.128 (.002)/T *	.007 (.882)/-
61. How many kids pick on you often	.057 (.341)/-	-.008 (.897)/-	.075 (.162)/-	-.077 (.151)/-	.006 (.920)/-
62. How many kids do you fear b/c mean	-.029 (.630)/-	.005 (.938)/-	.006 (.908)/-	-.085 (.092)/(T)	-.144 (.004)/T *
63. I think it is wrong to hit other people	.108 (.041)/C	.028 (.569)/-	-.076 (.088)/(T)	-.115 (.011)/T *	.022 (.632)/-
64. It is OK to push and shove if you are mad	.013 (.776)/-	.007 (.874)/-	.121 (.003)/T *	.023 (.609)/-	-.011 (.806)/-
65. It is OK to say mean things if you are angry	.013 (.769)/-	-.043 (.382)/-	.120 (.003)/T *	.037 (.391)/-	.022 (.614)/-
66. It is OK to yell and say bad things to others	.018 (.688)/-	.017 (.707)/-	.076 (.003)/T *	.063 (.128)/-	-.026 (.524)/-
67. It is wrong to get into physical fights	.086 (.146)/-	-.154 (.006)/T	-.022 (.673)/-	.006 (.920)/-	-.011 (.843)/-
68. Sometimes must fight to get what you want	.003 (.951)/-	.055 (.260)/-	.024 (.599)/-	.056 (.210)/-	-.002 (.964)/-
69. It's OK to hit if they hit you first	.110 (.059)/(T)	.084 (.148)/-	.007 (.892)/-	.083 (.125)/-	-.053 (.333)/-
C. Home and Family Environment					
70. My parents want me to get good grades	-.039 (.376)/-	.000 (.998)/-	.020 (.475)/-	-.032 (.473)/-	-.027 (.207)/-
71. I can tell my parents how I feel	-.059 (.217)/-	.065 (.214)/-	-.129 (.003)/T *	-.225 (.000)/T *	-.025 (.581)/--
72. I like to do things with my family	-.020 (.511)/-	.029 (.464)/-	-.001 (.975)/-	-.071 (.047)/T	.021 (.575)/-

Variables	2002	2003	2004	2005	2006
* = Statistically significant at $\alpha=.05$ using Holm's sequential method for familywise statistical significance. These differences are statistically significant by the most conservative criteria used here.	Somers' d (p)/ Favorable for... (T=treatment, C=Comparison, - = Neither)				
73. Parents know who I am with if I'm away	.060 (.192)/-	-.013 (.758)/-	.015 (.699)/-	-.057 (.173)/-	-.053 (.190)/-
74. Parents limit how much TV I watch	.051 (.412)/-	.045 (.460)/-	.010 (.862)/-	-.042 (.461)/-	.080 (.146)/-
75. Parents limit what kind of music I listen to	.045 (.455)/-	-.003 (.960)/-	.004 (.944)/-	-.099 (.082)/(T)	-.057 (.295)/-
76. Parents know who my friends are	-.037 (.406)/-	-.034 (.422)/-	-.054 (.189)/-	-.012 (.788)/-	-.026 (.562)/-
77. Parents let me know if I do a good job	.049 (.295)/-	.049 (.297)/-	-.036 (.370)/-	-.075 (.067)/(T)	.027 (.553)/-
78. I share thoughts and feelings with parents	.041 (.479)/-	.085 (.138)/-	-.086 (.094)/(T)	-.174 (.001)/T	.018 (.729)/-
79. Most days I spend some time with parents	.039 (.476)/-	.019 (.730)/-	-.032 (.509)/-	-.058 (.252)/-	.081 (.102)/-
80. My parents always want me to do my best	-.045 (.021)/T	-.040 (.171)/-	-.006 (.832)/-	-.017 (.464)/-	.011 (.0659)/-
81. There will always be people I can count on	-.037 (.283)/-	.026 (.466)/-	-.025 (.465)/-	-.093 (.002)T	.013 (.744)/-
82. Besides family there is an adult I can trust	-.069 (.133)/-	-.032 (.500)/-	.044 (.333)/-	-.087 (.042)T	-.023 (.616)/-
83. I believe there is some good in everybody	-.018 (.660)/-	.017 (.700)/-	.028 (.510)/-	.014 (.756)/-	.099 (.041)/C

Table 12: Block 2 Treatment vs. Block 2 Comparison Schools, Individual Items

Variables	2002	2003	2004	2005	2006
* = Statistically significant at $\alpha=.05$ using Holm's sequential method for familywise statistical significance. These differences are statistically significant by the most conservative criteria used here.	Somers' d (p)/ Favorable for... (T=treatment, C=Comparison, - = Neither)				
A. School Climate					
1. I like school	-.020 (.777)/-	-.211 (.006)/T *	-.182 (.014)/T *	-.083 (.230)/-	-.127 (.073)/(T)
2. I look forward to going to school	.030 (.678)/-	-.065 (.412)/-	-.234 (.001)/T *	-.101 (.145)/-	-.190 (.009)/T *
3. I try hard in school	.116 (.041)/C	.020 (.766)/-	-.107 (.074)/(T)	-.013 (.831)/-	-.007 (.906)/-
4. My teacher tells me when I do good job	-.074 (.280)/-	-.072 (.341)/-	-.132 (.026)/T	.048 (.480)/-	-.141 (.046)/T
5. My teacher listens to me...	.045 (.495)/-	-.097 (.194)/-	-.119 (.086)/(T)	.007 (.911)/-	-.107 (.117)/-
6. My teacher cares about me	.037 (.569)/-	-.108 (.150)/-	-.198 (.002)/T *	-.006 (.920)/-	-.198 (.002)/T *
7. I like my teacher	.120 (.068)/(C)	-.140 (.038)/T	-.146 (.028)/T	-.098 (.102)/-	-.076 (.225)/-
8. Adults teach us not to pick on other students	.064 (.290)/-	-.207 (.000)/T *	-.150 (.003)/T *	-.133 (.007)/T *	-.177 (.001)/T *
9. Adults try hard to prevent bullying	.096 (.126)/-	-.214 (.001)/T *	-.168 (.004)/T *	-.154 (.005)/T *	-.175 (.002)/T *
10. People here respect all races	-.139 (.047)/T	-.241 (.002)/T *	-.225 (.001)/T *	-.157 (.013)/T	-.268 (.000)/T *
11 People of my race can succeed here	-.008 (.903)/-	-.038 (.581)/-	-.077 (.212)/-	.004 (.949)/-	-.176 (.004)/T *
12. I feel lonely at school	-.116 (.094)/(C)	.041 (.599)/-	.009 (.902)/-	.111 (.079)/(T)	.046 (.505)/-
13. There is an adult I can trust	.091 (.121)/-	-.078 (.207)/-	-.131 (.023)/T	-.133 (.018)/T	-.143 (.008)/T *
14. I see graffiti here	-.343 (.000)/C *	-.062 (.438)/-	.129 (.081)/(T)	-.015 (.823)/-	.014 (.859)/-
15. My school is clean	.273 (.000)/C *	-.128 (.104)/-	-.147 (.039)/T	-.048 (.471)/-	-.098 (.173)/-
16. I like the way my school looks	.253 (.000)/C *	-.195 (.009)/T *	-.153 (.027)/T	-.073 (.261)/-	-.136 (.052)/(T)
17. Students here know the rules here	.013 (.856)/-	-.116 (.136)/-	-.257 (.000)/T *	-.084 (.202)/-	-.222 (.001)/T *

Variables	2002	2003	2004	2005	2006
* = Statistically significant at $\alpha=.05$ using Holm's sequential method for familywise statistical significance. These differences are statistically significant by the most conservative criteria used here.	Somers' d (p)/ Favorable for... (T=treatment, C=Comparison, - = Neither)				
18. Rule breakers are treated the same	.110 (.126)/-	-.067 (.402)/-	-.296 (.000)/T *	-.095 (.170)/-	-.217 (.003)/T *
19. Rules here are fair	.151 (.024)/C	-.114 (.134)/-	-.162 (.015)/T *	-.084 (.184)/-	-.259 (.000)/T *
20. Students help decide activities and rules	-.114 (.112)/-	-.177 (.026)/T	-.076 (.306)/-	-.186 (.006)/T *	-.144 (.050)/T
21. I care what teachers think of me	-.007 (.905)/-	-.095 (.158)/-	.015 (.830)/-	-.027 (.664)/-	.057 (.389)/-
22. I respect teachers here	.132 (.024)/C	-.089 (.164)/-	-.036 (.516)/-	-.068 (.198)/-	-.210 (.000)/T *
23. I respect the principal here	.071 (.202)/-	-.118 (.025)/T	-.127 (.007)/T *	-.200 (.000)/T *	-.166 (.000)/T *
B. School Safety: Attitudes and Aggressive Behavior (Perpetration, Victimization, and Witnessing)					
24. I feel safe at my school	.100 (.140)/-	-.130 (.088)/(T)	-.240 (.000)/T *	-.157 (.003)/T *	-.264 (.000)/T *
25. I feel safe on the school bus	.176 (.159)/-	-.130 (.349)/-	-.236 (.012)/T *	.033 (.777)/-	.001 (.995)/-
26. I feel safe walking to school	-.137 (.140)/-	-.152 (.148)/-	-.205 (.022)/T	.116 (.238)/-	-.132 (.189)/-
27. Ever stay away because unsafe at school	-.032 (.499)/-	.115 (.009)/T *	.025 (.535)/-	.116 (.001)/T *	.057 (.135)/-
28. Ever stay away unsafe on way to school	.015 (.689)/-	.081 (.035)/T	.028 (.425)/-	.047 (.215)/-	-.042 (.345)/-
29. I have a friend who cares about me	.060 (.267)/-	-.023 (.713)/-	-.106 (.047)/T	-.057 (.293)/-	-.003 (.957)/-
30. I get along with most kids here	-.043 (.538)/-	-.093 (.220)/-	-.199 (.005)/T *	-.108 (.099)/(T)	-.114 (.120)/-
31. My friends think wrong to hit	-.005 (.945)/-	-.228 (.001)/T *	-.125 (.075)/(T)	-.139 (.028)/T	-.087 (.191)/-
32. My friends think is OK to say mean things	.101 (.124)/-	.158 (.023)/T	.139 (.042)/T	.193 (.001)/T *	.136 (.039)/T
33. My friends think is OK to yell	.025 (.677)/-	.169 (.006)/T *	.126 (.047)/T	.087 (.125)/-	.144 (.019)/T *
34. My friends think is OK to push and shove	.010 (.872)/-	.223 (.000)/T *	.117 (.062)/(T)	.151 (.006)/T *	.196 (.001)/T *
35. My friends think sometimes must fight	.071 (.210)/-	.104 (.078)/(T)	.076 (.216)/-	.059 (.276)/-	.157 (.003)/T *

Variables	2002	2003	2004	2005	2006
* = Statistically significant at $\alpha=.05$ using Holm's sequential method for familywise statistical significance. These differences are statistically significant by the most conservative criteria used here.	Somers' d (p)/ Favorable for... (T=treatment, C=Comparison, - = Neither)				
36. I saw someone physically attacked	.006 (.940)/-	-.085 (.280)/-	-.303 (.000)/T *	-.048 (.489)/-	-.444 (.000)/T *
37. I saw someone teased in mean way	-.103 (.167)/-	-.029 (.720)/-	-.231 (.002)/T *	-.100 (.160)/-	-.285 (.000)/T *
38. I saw someone threaten to hit	.009 (.906)/-	-.183 (.018)/T	-.257 (.001)/T *	-.080 (.262)/-	-.479 (.000)/T *
39. I pushed, shoved, hit, etc.	-.014 (.830)/-	-.090 (.190)/-	-.096 (.144)/-	-.153 (.011)/T	-.267 (.000)/T *
40. I got into physical fight when angry	.009 (.855)/-	.002 (.980)/-	-.024 (.655)/-	-.058 (.188)/-	-.150 (.002)/T *
41. I teased other students in a mean way	-.036 (.543)/-	-.003 (.969)/-	-.103 (.091)/(T)	-.137 (.014)/T	-.068 (.246)/-
42. I lied about student to get student disliked	.035 (.480)/-	.006 (.905)/-	-.141 (.002)/T *	-.030 (.532)/-	-.016 (.725)/-
43. I tried to exclude others from my group	.005 (.944)/-	-.112 (.099)/(T)	-.201 (.001)/T *	-.175 (.003)/T *	-.129 (.052)/(T)
44. I said mean things to get student disliked	.039 (.481)/-	-.026 (.616)/-	-.081 (.102)/-	-.068 (.168)/-	-.062 (.190)/-
45. I threatened to hit or hurt student	.025 (.587)/-	-.005 (.923)/-	-.067 (.198)/-	-.072 (.101)/-	-.172 (.001)/T *
46. I got pushed, shoved, slapped, or kicked	-.119 (.097)(T)	-.019 (.816)/-	-.176 (.017)/T *	-.194 (.004)/T *	-.314 (.000)/T *
47. I got teased in a mean way	-.011 (.852)/-	-.027 (.745)/-	-.166 (.016)/T *	-.030 (.672)/-	-.116 (.132)/-
48. Student lied about me to get me disliked	-.051 (.482)/-	.030 (.701)/-	-.208 (.003)/T *	-.051 (.455)/-	-.204 (.006)/T *
49. Student tried to keep me out of group	.023 (.746)/-	.024 (.762)/-	-.100 (.153)/-	.023 (.732)/-	-.176 (.016)/T *
50. Student said mean things to get me disliked	-.028 (.691)/-	-.019 (.804)/-	-.151 (.025)/T	.008 (.905)/-	-.133 (.060)/(T)
51. Student threatened to hit or hurt me	.029 (.661)/-	.002 (.982)/-	-.136 (.041)T	-.094 (.087)/(T)	-.283 (.000)/T *
52. ...threatened to hit or hurt me: lunchroom	.056 (.238)/-	-.065 (.214)/-	-.080 (.065)/(T)	.082 (.110)/-	-.054 (.321)/-
53. ...threatened to hit or hurt me: playground	-.108 (.095)/(T)	-.096 (.199)/-	-.204 (.002)/T *	-.142 (.024)/T	-.304 (.000)/T *
54. ...threatened to hit or hurt me: bathroom	-.054 (.072)/(T)	-.098 (.006)/T *	-.142 (.000)/T *	.007 (.848)/-	-.105 (.001)/T *

Variables	2002	2003	2004	2005	2006
* = Statistically significant at $\alpha=.05$ using Holm's sequential method for familywise statistical significance. These differences are statistically significant by the most conservative criteria used here.	Somers' d (p)/ Favorable for... (T=treatment, C=Comparison, - = Neither)				
55. ...threatened to hit or hurt me: hallway	.005 (.921)/-	-.092 (.040)/T	-.085 (.033)/T	.021 (.642)/-	-.090 (.064)/(T)
56. ...threatened to hit or hurt me: bus	.068 (.463)/-	-.061 (.516)/-	-.171 (.017)/T *	.006 (.941)/-	-.196 (.009)/T *
57. How many students get picked on	-.039 (.595)/-	-.209 (.006)/T *	-.322 (.000)/T *	-.034 (.625)/-	-.298 (.000)/T *
58. How many students pick on others	-.088 (.243)/-	-.186 (.015)/T	-.355 (.000)/T *	-.166 (.013)/T	-.311 (.000)/T *
59. How many kids afraid of you b/c mean	-.032 (.581)/-	-.075 (.239)/-	-.074 (.198)/-	-.041 (.431)/-	-.089 (.126)/-
60. How many kids do you pick on often	.058 (.315)/-	-.113 (.038)/T	-.155 (.003)/T *	-.129 (.011)/T	-.099 (.073)/(T)
61. How many kids pick on you often	-.020 (.771)/-	.073 (.355)/-	-.152 (.030)/T	-.005 (.947)/-	-.211 (.004)/T *
62. How many kids do you fear b/c mean	-.103 (.107)/-	-.033 (.666)/-	-.064 (.334)/-	-.100 (.137)/-	-.217 (.002)/T *
63. I think it is wrong to hit other people	-.027 (.606)/-	-.130 (.037)/T	-.038 (.538)/-	-.097 (.082)/(T)	-.092 (.108)/-
64. It is OK to push and shove if you are mad	.028 (.553)/-	.164 (.001)/T *	.160 (.001)/T *	.139 (.002)/T *	.146 (.003)/T *
65. It is OK to say mean things if you are angry	.046 (.353)/-	.109 (.056)/(T)	.133 (.013)/T *	.241 (.000)/T *	.129 (.009)/T *
66. It is OK to yell and say bad things to others	.005 (.918)/-	.132 (.009)/T *	.103 (.039)/T	.230 (.000)T *	.091 (.063)/(T)
67. It is wrong to get into physical fights	-.026 (.674)/-	-.087 (.239)/-	-.136 (.050)/T	-.058 (.387)/-	-.180 (.007)T *
68. Sometimes must fight to get what you want	-.010 (.855)/-	.250 (.000)/T *	.170 (.003)/T *	.156 (.002)/T *	.168 (.002)/T *
69. It's OK to hit if they hit you first	.138 (.034)/T	.190 (.013)/T	.263 (.000)/T *	.223 (.001)/T *	.283 (.000)/T *
C. Home and Family Environment					
70. My parents want me to get good grades	-.021 (.494)/-	.005 (.911)/-	-.006 (.853)/-	-.018 (.526)/-	.013 (.652)/-
71. I can tell my parents how I feel	-.105 (.077)/(T)	-.098 (.142)/-	-.226 (.000)/T *	-.091 (.083)(T)	.020 (.752)/-
72. I like to do things with my family	.041 (.386)/-	-.081 (.117)/-	-.060 (.157)/-	-.081 (.045)/T	-.039 (.354)//-

Variables	2002	2003	2004	2005	2006
* = Statistically significant at $\alpha=.05$ using Holm's sequential method for familywise statistical significance. These differences are statistically significant by the most conservative criteria used here.	Somers' d (p)/ Favorable for... (T=treatment, C=Comparison, - = Neither)				
73. Parents know who I am with if I'm away	-.001 (.987)/-	.037 (.519)/-	-.023 (.615)/-	-.053 (.224)/-	-.022 (.631)/-
74. Parents limit how much TV I watch	-.172 (.019)/T *	-.192 (.017)/T	-.025 (.742)/-	-.109 (.122)/-	-.086 (.253)/-
75. Parents limit what kind of music I listen to	-.139 (.051)/(T)	-.110 (.169)/-	.100 (.170)/-	-.118 (.093)(T)	-.106 (.162)/-
76. Parents know who my friends are	-.010 (.838)/-	-.140 (.007)/T *	-.145 (.006)/T *	-.036 (.500)/-	-.004 (.945)/-
77. Parents let me know if I do a good job	-.036 (.501)/-	-.038 (.525)/-	-.053 (.308)/-	.035 (.502)/-	-.039 (.456)/-
78. I share thoughts and feelings with parents	-.130 (.058)/(T)	-.133 (.073)/(T)	-.246 (.000)/T *	-.125 (.045)/T	-.053 (.445)/-
79. Most days I spend some time with parents	-.017 (.791)/-	-.002 (.977)/-	-.034 (.598)/-	-.085 (.162)/-	.123 (.078)/(C)
80. My parents always want me to do my best	.039 (.295)/-	-.003 (.945)/-	-.004 (.918)/-	-.012 (.657)/-	.065 (.073)/(C)
81. There will always be people I can count on	.050 (.283)/-	.101 (.071)/(C)	-.067 (.139)/-	-.071 (.064)/(T)	-.003 (.952)/-
82. Besides family there is an adult I can trust	-.043 (.444)/-	-.055 (.340)/-	-.124 (.029)/T	-.178 (.001)/T *	-.092 (.090)/(T)
83. I believe there is some good in everybody	.071 (.225)/-	-.097 (.130)/-	-.107 (.073)/(T)	-.046 (.427)/-	.028 (.655)/-

Somers' d_{yx} and d_{xy} can be interpreted as the proportional reduction in error (PRE) which occurs under the following conditions (Loether and McTavish 1993:224). Error without using the predictor is calculated as the error that occurs by predicting concordance or discordance at random for pairs not tied on the predictor (note the loss of information resulting from exclusion of pairs tied on the predictor). Errors with the predictor are the errors that occur by always predicting concordance (if the relationship is positive; or discordance if the relationship is negative) between the predictor and the dependent variable, again for pairs not tied on the predictor. Costner (1965) asserts that any measure that includes ties of any kind in its pool is not properly a PRE measure, because a tie cannot clearly be counted as either a correct or an erroneous prediction, and only these two categories are permissible for PRE measures. Loether and McTavish (1993) counter that if one takes the position that the pool of potential errors should include all those instances for which a prediction is likely to be made, then it is reasonable to include ties, and it is this latter position that is adopted here. Somers' d , then, can be described as the proportional reduction in error of concordance *for pairs untied on the predictor*. In the present context, this means proportional reduction in error for pairs that involve treatment versus comparison schools, which is ideal for our present purposes.

As a general guideline, it may be best to regard values of Somers' d greater than .100 as being of interest, and values of Somers' d less than .100 as being of little or no substantive significance. Further description of Somers' d may be found in a number of elementary statistics texts (for a clear presentation of its computation, see in particular Wright 1979; for a more recent treatment, see Walker and Madden 2005). Unlike other ordinal measures of association, Somers' d makes a distinction between the predictor and the dependent variable (it is an asymmetric measure), so Somers' d is particularly appropriate for data like those in the present study in which we are comparing ordinal rankings on outcome measures for two groups defined by an intervention. It provides not only an associated test for statistical significance (the same as for Kendall's tau and gamma, two widely used symmetric ordinal measures of association), but also a PRE measure of the strength of the relationship. Results favoring the treatment schools are indicated by a T, those favoring the comparison schools by a C. Letters C and T in parentheses, (C) or (T), indicate differences not significant at the .05 level, but marginally significant at the .10 level.

In addition to Somers' d , the estimated statistical significance level (p) of Somers' d is also included. There is a question whether inferential statistics, tests of statistical significance, are appropriate for data such as those in the present study, since the schools were not really selected in a random sampling procedure. Strictly speaking, then, inferential tests are not applicable here. The arguments that have been made for the application of inferential tests to nonprobability samples are that, in spite of the fact that there is no time-and-space-bounded population to which the inference is being drawn, (1) we can use the statistical significance levels in the same way that we might use some other criterion (like the .100 value for Somers' d , as suggested above) to distinguish between relationships we consider more important and relationships we consider less important, and (2) even a nonprobability sample *in space* can be regarded as a random sample *in time*, that is, one can infer to the same sample one has analyzed *at a different point in time*. The latter is common in practice; we assume that our results today will apply to our decisions about program adoption and implementation tomorrow. Here,

significance levels are used in both of the above senses, but the results would not differ substantially if we used some other criterion (for example, Somers' $d \geq .100$) instead. Given that we are paying attention to statistical significance, however, it is important that we also pay attention to the issue of repeated testing. If one uses the conventional $p \leq .05$ cutoff for statistical significance, multiple tests mean that using the same .05 criterion for each test results in a greater than .05 probability of rejecting the null hypothesis when it is true. In order to maintain an overall or *familywise* probability of .05 for rejecting the null hypothesis when it is true, it is necessary to take into account the fact that in 100 tests, even if the data are entirely random, one can expect to get a statistically significant result in about 5 tests out of the 100. Here we use Simes' modified Bonferroni procedure in conjunction with Holm's sequential Bonferroni test (Simes 1986) to adjust for multiple testing. Again, one may disregard the inferential statistics if one wishes; the results based solely on the magnitude of Somers' d are practically the same. Relationships involving Somers' d which are significant based on this criterion are indicated by an asterisk (*). Although the correspondence is not perfect, these will also be relationships associated with a high value of Somers' d .

For the elementary schools, as indicated in Table 10, there are only two differences between treatment and comparison schools at baseline on the items measured in the surveys. Students in the comparison schools at baseline were more likely to indicate that they liked their teachers (item 7) and that they respected their teachers (item 22). The 9 other differences that would have been regarded as statistically significant had the Bonferroni adjustment not been made mostly favored the comparison over the treatment schools. One difference not indicated in the comparison of the survey items is that the comparison schools were larger than the treatment schools, but the size of the class or school does not appear to be related to bullying behavior (Lawrence 2007). Overall, it appears that the treatment and comparison schools were well and closely matched at baseline.

In the first implementation year (2003), the treatment schools differentiated themselves from the comparison schools with students indicating that in the treatment schools students were more likely to notice teachers and other adults trying to prevent bullying (items 8 and 9), consistent with implementation of the BPYS curriculum, that there was greater respect for people of all races (consistent with a component of the BPYS curriculum), and that they were less likely to stay away from school because it was unsafe. There were also some changes in perceptions of the condition of the school, and students in treatment schools were more likely to indicate that it was wrong to get into physical fights and to disagree that it is sometimes necessary to fight to get what you want. In the second year of implementation (2004), many of these differences persisted, and students were generally more positive about the school climate in treatment than in comparison schools. In addition, students in treatment schools were less likely to see someone physically attacked (item 36) and to engage in relational aggression by lying about another student to get them disliked (item 42); to be victims of relational aggression (items 47, 48, 50); and treatment school students reported less picking on other students by themselves and others (items 57, 59, 60). In the third and final year of implementation, treatment school students reported generally better school climate, better school safety, and (see the third and fourth pages of Table 10) better attitudes and behaviors on the vast majority of the

outcome measures, even using the very conservative Bonferroni adjustment for statistical significance. Much of this, however, dropped off in the fifth (2006), post-implementation year.

Examination of Tables 11 and 12 indicates that there were substantial differences in the experiences of the two blocks of schools. For Block 1 (Chapman, Doubleday, Guilford, and Harcourt), the differences between treatment and comparison schools emerged later, were not as widespread across the different items, and by the post-implementation year, school climate differences actually favored the comparison schools. For Block 2 (Elsevier and Ingram), by contrast, the differences favoring the treatment schools emerged earlier, and not only persisted but actually appear to be strongest in the post-implementation year. This difference between the two blocks appears to be directly related to the quality of the implementation of the program in the different schools. As discussed earlier, both Chapman and Doubleday had problems in implementation, in contrast to Elsevier, where the program was implemented well practically from the start. Overall, it appears that BPYS, if implemented faithfully, can have a quick, pervasive, and enduring positive effect; but even if there are problems in implementation, it can have positive effects, at least during the period of active involvement by program staff.

Program Components and Hypotheses: Composite Scale Outcomes

Table 13 presents the results for the composite scales described in the previous section. As in the previous tables, results are presented for each year separately. Because the scale scores can be treated as being measured at the interval level, Pearson's r (instead of Somers' d as in the previous tables) is used here to measure the strength of the relationship between the program and the outcome. Statistical significance of the differences between the treatment and comparison schools is assessed using the test of statistical significance for Pearson's r (equivalent to a t -test for group differences or an ANOVA F test for a oneway analysis of variance with a single dichotomous factor). The modified Bonferroni procedure used in the previous tables is applied in the same way (separately by year) in Table 13. We also explored the use of robust standard errors that adjusted for clustering of students within schools. It is worth repeating, however, that the use of *any* inferential statistics in the present context may be questioned, and it is the strength of the relationship (the magnitude of Pearson's r) that should be emphasized; however, the results for the modified Bonferroni procedure correspond well with the strength of the relationship, highlighting those relationships strong enough to be of greatest interest here.

It is expected that in the baseline year, there will be no statistically significant differences between treatment and comparison schools, and as indicated for the outcomes for the year 2002, all but one of the baseline year differences is not statistically significant at the .05 level. Relational aggression perpetration appears to be significantly correlated with treatment at baseline ($p=.007$); but applying the same modified Bonferroni procedure used in the previous tables, the critical $\alpha=.006$, which is less than the attained significance level of .007, so we cannot reject the null hypothesis of no difference at baseline for the variables in Table 13. The results for the composite scales in Table 13 generally parallel the results for the individual items in Table 10. Most of the impact of the program occurs in the final year of implementation, and much of it is dissipated by the post-implementation year. In terms of the hypotheses related to the major program components:

Table 13: Bivariate Analysis of Program Impact - Elementary Schools

Outcomes: Pearson's r (p) (Treatment=1, Comparison=2)	Year				
	2002	2003	2004	2005	2006
Bullying discouraged	-.012 (.749)	-.140* (.001)	-.150* (.000)	-.089 (.018)	-.027 (.480)
Witnessed bullying	-.022 (.572)	.077 (.059)	.105* (.006)	.118* (.002)	.145* (.000)
Physical aggression perpetration	-.063 (.101)	.044 (.283)	.102* (.008)	.116* (.002)	.047 (.216)
Physical aggression victimization	.005 (.900)	-.009 (.833)	.052 (.176)	.109* (.004)	.101 (.008)
Relational aggression perpetration	-.103 (.007)	-.066 (.105)	.080 (.037)	.134* (.000)	.052 (.177)
Relational aggression victimization	-.027 (.479)	-.028 (.491)	.109* (.004)	.051 (.174)	.067 (.078)
Perceived school safety	-.069 (.073)	-.059 (.145)	-.085 (.026)	-.037 (.324)	-.002 (.954)
Peer environment (perceived peer attitudes toward aggression)	.016 (.674)	.067 (.102)	.052 (.171)	.152* (.000)	.036 (.348)
Own attitude toward aggression	.024 (.531)	.134* (.001)	.132* (.001)	.114* (.002)	.093 (.015)
<p>* Statistically significant at a $\alpha=.05$ (familywise, i.e., across all comparisons adjusting for nonindependent repeated testing) using Simes' modified Bonferroni procedure in conjunction with Holm's sequential Bonferroni test for the significance level of Pearson's r. Using this test identifies 13 of the 15 relationships for which $r > .100$ as statistically significant. An alternative approach is to estimate the model with robust standard errors that adjust for clustering within schools, and this identifies 8 of the 15 relationships $r > .100$ as statistically significant. Neither method identifies any of the relationships $r < .100$ as statistically significant, and applying both would identify only two (relational aggression perpetration and peer environment in 2005) as statistically significant.</p>					
<p>Note: At the suggestion of a reviewer, data were analyzed to see whether the program effect differed by ethnicity. Based on the modified Bonferroni test, none of the interactions between ethnicity and program impact was statistically significant at $\alpha=.05$.</p>					

(1) Discouragement of bullying (corresponding to the first major component of the program) is significantly higher for treatment than for comparison schools in the first and second year of implementation (2003 and 2004) in both Table 10 and Table 13; results are mixed for the third year of implementation in Table 10, and in Table 13, the difference is statistically significant at the conventional .05 cutoff, but once we make the sequential Bonferroni adjustment, we just barely fail to reject the null hypothesis (adjusted $\alpha=.007$, $p=.008$ for this comparison). As indicated in Table 10, under the condition of better implementation of the program, the impact of the program on discouragement of bullying persisted into the post-implementation year as well. ***The hypothesis that BPYS implementation results in students' recognizing that bullying is being discouraged appears to be supported, more so when BPYS is better implemented.***

(2) In both Table 10 and Table 13, most of the impact of BPYS on witnessing bullying, physical aggression perpetration and victimization, and relational aggression perpetration and victimization occurs mostly in the second and third years of program implementation (2004 and 2005). The impact of BPYS on witnessing bullying actually appears more persistent in Table 13 (composite scales) than in Table 10 (separate items), with a reduction in witnessing bullying that is statistically significant, even after the modified Bonferroni adjustment, in the second and third years of program implementation and also in the post-implementation year. Physical aggression and victimization, and relational aggression perpetration, all appear to be higher in comparison than in treatment schools in 2005, but much of the effect has dissipated by the post-implementation year (note that physical aggression victimization is higher in comparison than in treatment schools in the post-treatment year, with $r=.101$, but this does not quite meet the sequential Bonferroni criterion (adjusted $\alpha=.007$, $p=.008$ for this comparison). Once again, from Table 10, it is evident that the effects start earlier and persist longer with better implementation. ***The hypothesis that BPYS reduces bullying and related behaviors appears to be supported, particularly for (a) bullying itself (as indicated by the more persistent effect on this variable in Table 13) and (b) for better implementation of the program.***

(3) In both Table 10, the results regarding the impact of BPYS on perceived school safety are mixed, and these mixed results are reflected in the fact that we are unable to reject the null hypothesis of no impact of BPYS on perceived school safety in any of the five years (pre-implementation, three years implementation, one year post-implementation). Part of the problem may be the inclusion of different locations (at school, walking to school, on the school bus) and, in addition to feelings of safety, specific behavior, whether the student avoided school because of feeling unsafe (again with different locations, at school and traveling to school) in the same composite scale. Reliability on this scale was marginal, with Cronbach's $\alpha=.62$, and item-level results in Tables 10-12 suggest that the best results are obtained for the single item, "I feel safe at my school," the item over which BPYS is most likely to have an impact that is undiluted by extraneous factors (safety on the bus or in the neighborhoods surrounding or on the way to school). If we focus on this single item instead of the scale, it appears that BPYS has the intended effect, particularly where well implemented. ***The hypothesis that BPYS results in perceptions of increased safety at school is at least weakly supported, primarily (a) at school, rather than on the way to or from school, and (b) where BPYS is well implemented.***

(4) Finally, with regard to the intervening variables, the impact of BPYS appears to be stronger on one's own attitudes toward physical and relational aggression than on the perceived attitudes of one's friends toward physical and relational aggression (the peer environment). The impact on peer environment in Table 13 is statistically significant only in the final year of implementation; the impact on one's own attitudes is evident in all three years of implementation (and in the post-implementation year, $r=.093$ but $p=.015$ falls short of the adjusted $\alpha=.007$ for this comparison). One may speculate that this result may be in part a result of a lag between the change in one's own attitudes, which is immediately perceived by the student, and the change in the attitudes of friends, which the student may assume to be unchanged until there is concrete evidence of that change (and concrete evidence may be slow in coming). It is also the case that the impact appears, again, to be greater where the program is better implemented. In summary, ***BPYS appears to have a favorable impact on attitudes toward physical and relational aggression, and to a lesser extent on perceived peer environment, more so where it is well implemented.***

Quality of Implementation and Overall Impact

The discussion to this point has focused on whether BPYS successfully achieved its program goals; this and the next two sections provide additional context for the findings in the previous two sections. A recurrent theme in the previous two sections has been the impact of quality of implementation on the outcome of the intervention, but in the previous two sections, this has not been quantified, but only described in a qualitative way, by looking at differences in results between blocks of schools with stronger and weaker quality of implementation. Here, we present a summary of the impact of quality of implementation on the results. For this analysis, quality of implementation was coded as zero for all comparison schools in all years, indicating that they were not implementing BPYS; and it was also coded as zero for the treatment schools in the pre-implementation year, similarly indicating that they were not implementing the program in that year. For the three years in which the program was actively being implemented, the average implementation scores from the process evaluation were used. For each year of implementation, each treatment school was assigned the mean of the fall and spring BPYS and CSPV implementation scores, indicating the quality of implementation for that year. For the post-implementation year, each treatment school was assigned the mean implementation score over all three years of implementation. The reason for this approach is that we would expect the persisting effect of the implementation after active implementation has been discontinued to depend on how well it had been implemented during the period when the program providers were actively involved with the school. In other words, the effects of a better implemented intervention are more likely than the effects of a poorly implemented intervention to persist even after the active phase of the implementation is over. Focusing on the composite scales for the outcomes of the three major components of the program (peer environment and attitudes toward aggression are omitted here), the data for all five years (pre-implementation, three years of implementation, and post-implementation) were pooled to allow for greater variation in implementation. The results are presented in Table 14.

While statistical significance levels are presented in Table 14, their use is perhaps even less justified (see the earlier discussion of the use of inferential statistics in the present context)

here than elsewhere in this report, and we focus here on the descriptive rather than the inferential statistics: the magnitude of the correlation coefficient. In Table 14, for the last variable, perceived school safety, the magnitude of the correlation is the same for quality of implementation as for the simple treatment-comparison contrast. For all of the other outcomes, the magnitude of the correlation is larger when the score for quality of implementation is used instead of the simple treatment-comparison contrast. The effects themselves are small, but they confirm the more qualitative discussion earlier that suggested that quality of implementation was making a difference in the findings between the two blocks of treatment and comparison schools.

Table 14: Quality of Implementation and BPYS Program Goals - Elementary Schools

Outcome Pearson's r (p) (Treatment=1, Comparison=2)	Quality of Implementation Score	Program Treatment vs. Comparison
Bullying discouraged	-.092 (.000)	-.088 (.000)
Witnessed bullying	.118 (.000)	.094 (.000)
Physical aggression perpetration	.052 (.002)	.023 (.168)
Physical aggression victimization	.054 (.001)	.035 (.037)
Relational aggression perpetration	.071 (.000)	.048 (.004)
Relational aggression victimization	.070 (.000)	.060 (.000)
Perceived school safety	-.053 (.002)	-.053 (.002)

Multivariate Analysis of the Impact of BPYS

To further understand the broader context in which BPYS affects perceived discouragement of bullying, bullying and related aggressive behavior, and perceived school safety, Table 15 presents an analysis based on the model in Figure 3, and places the results of the BPYS intervention in that broader context. In Table 15, the explained variance (R^2) for each of the outcome variables for each year is provided, along with standardized regression coefficients for each of the predictors of each of the outcomes, with the outcomes listed by year in the leftmost column of the table and the predictors arrayed across the top. Levels of statistical significance are indicated by asterisks, and the same comments regarding statistical significance testing in a nonprobability sample as indicated previously apply here. As a general guideline, standardized regression coefficients greater than .100 are of more interest than standardized coefficients less than .100, and using this criterion produces substantive conclusions similar to those that would be obtained using statistical significance. As a practical matter, the asterisks associated with the significance levels make it easier to visually spot patterns of strong relationships in the data.

Table 15: Predictors of Program Outcomes (Standardized Regression Coefficients): Elementary School

Dependent Variable	R ²	Sex	White	Other	Class in School	Grades (A-F)	Family Bonding	Treatment/ Comparison	Peer Attitudes	Attitudes re Aggression
<u>Physical aggression victim</u>										
Year 1	.097	.002	.186**	.082	-.066	-.080*	-.056	.033	.250**	-.035
Year 2	.092	-.014	.201**	.197**	-.054	-.085*	-.101*	.017	.178**	-.023
Year 3	.098	.019	.089*	.219**	.025	-.057	-.133**	.063	.188**	-.046
Year 4	.113	.020	.079	.032	-.111**	-.021	-.122**	.077*	.297**	-.098*
Year 5	.089	-.041	.154**	.124**	-.110**	-.060	-.065	.133**	.229**	-.063
<u>Relational aggression victim</u>										
Year 1	.060	-.083*	.112*	.027	-.026	-.107**	-.018	-.008	.190**	-.006
Year 2	.091	-.092*	.144**	.141**	-.026	-.073	-.151**	-.019	.181**	-.031
Year 3	.070	-.085*	.086	.113*	.031	.033	-.068	.117**	.230**	-.099*
Year 4	.093	-.073	.069	.038	-.056	-.011	-.073	.009	.269**	.008
Year 5	.085	-.110**	.149**	.127**	-.095*	-.061	-.044	.093*	.221**	-.017
<u>Physical aggression perpetrator</u>										
Year 1	.390	.095**	-.060	-.084*	.013	-.086**	-.036	-.095**	.207**	.414**
Year 2	.425	.008	.068	.004	.055	-.031	-.083*	-.033	.167**	.492**
Year 3	.433	.046	-.012	.057	.041	-.062*	-.033	.025	.190**	.475**
Year 4	.494	.026	-.023	.010	-.053	-.070*	-.112**	.029	.217**	.492**
Year 5	.418	.040	.004	.038	.008	-.060	-.047	-.011	.214**	.462**
<u>Relational aggression perpetrator</u>										
Year 1	.211	-.038	.039	.011	.021	-.084*	-.021	-.100**	.191**	.297**
Year 2	.246	-.011	.022	.015	.057	.020	-.134**	-.125**	.210**	.266**
Year 3	.247	-.034	.011	.030	.017	-.028	-.112**	.023	.179**	.311**
Year 4	.355	-.034	.029	.023	-.032	-.011	-.130**	.056	.189**	.407**
Year 5	.283	-.035	.035	.063	.050	-.066	-.081*	.009	.178**	.358**
<u>Witnessed aggression/bullying</u>										
Year 1	.127	.066	.093*	.019	.136**	-.024	-.003	-.012	.211**	.086*
Year 2	.168	.010	.111*	.112*	.173**	.036	-.034	.077	.151**	.192**
Year 3	.184	-.045	.137**	.170**	.157**	-.009	-.033	.096**	.227**	.134**
Year 4	.214	.037	.075	.058	.045	.047	-.129**	.053	.302**	.115**
Year 5	.139	.022	.079	.079	-.002	.021	-.018	.141**	.270**	.094*

Dependent Variable	R ²	Sex	White	Other	Class in School	Grades (A-F)	Family Bonding	Treatment/ Comparison	Peer Attitudes	Attitudes re Aggression
<u>Feel safe at school</u>										
Year 1	.042	.075	.066	.122**	.047	.104**	.092*	-.024	-.020	-.045
Year 2	.049	.001	.069	.070	.101*	.021	.163**	-.031	-.056	.002
Year 3	.054	.070	-.013	.013	.073	.095*	.103*	-.087*	-.137**	.028
Year 4	.062	.022	-.084	-.048	.111**	.021	.129**	-.021	-.149**	-.018
Year 5	.034	.047	-.047	-.032	.050	.061	.064	-.003	-.132**	-.012
<u>Bullying discouraged at school</u>										
Year 1	.068	-.108**	.030	.076	-.015	.042	.144**	-.001	-.064	-.058
Year 2	.141	-.099*	.054	.108*	.019	.012	.148**	-.109**	-.131**	-.115*
Year 3	.103	-.044	.033	.087*	-.091*	-.012	.159**	-.119**	-.089	-.043
Year 4	.113	-.105**	-.104*	-.029	.063*	-.056	.167**	-.035	-.149**	-.088
Year 5	.096	.035	-.069	-.045	.008	.016	.106**	-.025	-.147**	-.142**
<u>Own attitudes toward aggression</u>										
Year 1	.117	.220**	-.119**	-.108*	.099**	-.065	-.176**	.018	NA	NA
Year 2	.160	.175**	-.050	-.052	.220**	-.076	-.171**	.141**	NA	NA
Year 3	.178	.214**	-.106*	-.148**	.120**	-.063	-.256**	.099**	NA	NA
Year 4	.166	.178**	-.073	-.062	.131**	-.007	-.296**	.060	NA	NA
Year 5	.162	.172**	-.146**	-.129**	.110**	-.043	-.289**	.060	NA	NA
<u>Friends attitudes toward aggression</u>										
Year 1	.126	.185**	-.021	-.029	.203**	-.056	-.181**	.012	NA	NA
Year 2	.171	.167**	.006	.008	.193**	.020	-.299**	.060	NA	NA
Year 3	.162	.177**	-.085*	-.091*	.213**	-.009	-.244**	.023	NA	NA
Year 4	.153	.120**	-.011	.041	.122**	-.013	-.299**	.110**	NA	NA
Year 5	.157	.120**	-.041	-.014	.140**	-.039	-.333**	.038	NA	NA
* p ≤ .050 ** p ≤ .010	Year 1 n = 679 Year 2 n = 583 Year 3 n = 670 Year 4 n = 686 Year 5 n = 674									

Looking first at physical and relational aggression victimization, there are differences by ethnicity (as noted earlier, a potentially unreliable classification), with white and other respondents reporting higher victimization rates than Latinos, the reference category. The relationship of victimization to class in school and grades in school is not consistent, but to the extent that it exists, victimization appears to be more prevalent among students in earlier years of school and students with lower grades (the prevalent pattern does not appear to be picking on the kids with better grades). Most importantly, friends attitudes toward aggression and violence are strongly and consistently related to victimization and perpetration of aggression and violence. Net of other influences on victimization, the impact of the BPYS intervention appears to be relatively weak, and only appears in the second year of implementation and later. This point will be discussed in further detail below, but it is consistent with the item-specific findings presented above.

Physical and relational aggression perpetration and witnessing aggression and bullying in the school have similar patterns of relationships. All three are driven by peer attitudes and one's own attitudes toward aggression and violence, with those students having attitudes more favorable to aggression and violence more likely to perpetrate and witness (perhaps as a result of their own perpetration) physical and relational aggression and bullying. Other variables in the model are less strongly related to these three outcomes. It does appear that in the first two years, relational aggression is higher in the treatment than in the comparison schools, controlling for the other variables in the model, but this relationship disappears in year 3 and does not reemerge. Family bonding appears to be a protective factor against relational aggression, but not against physical aggression or bullying more generally. The impact of BPYS is evident in reduced rates of witnessing aggression or bullying after the second year of implementation, but BPYS appears to have no direct effect on perpetration of relational or physical aggression.

BPYS also appears to have little impact on school safety, but more detailed analysis indicates that this is a function of how the school safety scale was constructed. The school safety scale included items about feeling safe at school, on the school bus, and on the way to school, plus whether the student had ever stayed away from school because they felt unsafe at school or on the way to school. Eliminating the second and third items from the scale (feel safe on the bus, feel safe on the way to school) and treating them as a separate scale indicated that BPYS had no impact on feelings of safety on the school bus or on the way to school, and no impact (as expected) on feelings of safety at school in the pre-implementation baseline year, but for all subsequent years, treatment school students indicated greater feelings of safety *at school*: for year 2, the standardized regression coefficient $b = -.075$ ($p=.067$); for year 3 $b = -.130$ ($p=.000$); for year 4 $b = -.102$ ($p=.006$); and for year 5 $b = -.082$ ($p=.034$). Thus feelings of school safety were no different in the baseline year, increased during the implementation period, then declined (but still favored the treatment schools) in the post-implementation year. Note that family bonding and peer attitudes also appear to have an impact on feelings of safety at school.

Discouragement of bullying at school appears to be most evident for the treatment schools in the second and third years of implementation, but controlling for the other variables in the model are less evident in the last two years. With regard to one's friends' and one's own attitudes toward aggression and violence, these appear to be strongly related to family bonding, gender (with males being more accepting of violence and aggression than females), and class in school (with older students more accepting of violence and aggression). BPYS shows less impact here than in the item-specific analysis.

BPYS and Faculty and Staff Perceptions of School Climate

The previous analyses involving the students address the principal goals of BPYS, but there is also an interest in the perceptions of the school climate on the part of faculty and staff. Originally, we had anticipated examining different facets of school climate for the faculty and staff, but a factor analysis indicated that what we had thought would be different dimensions of the faculty/staff perception of the school climate actually loaded onto a single dimension, and the eigenvalues and corresponding scree plot clearly indicated a single factor solution (one large eigenvalue followed by several eigenvalues close to each other and close to one in magnitude). We have therefore chosen to summarize the changes over time in the treatment and comparison schools in the faculty and staff perceptions of school climate in Figure 4 below. In Figure 4, more as a matter of convenience than for any other reason, the summary index of school climate is based on the factor score coefficients, instead of simply standardizing and adding the items to create the scale, as was done with the composite scales used to evaluate the major components of BPYS. The result here would not be substantially different were we to use the same procedure as for the other composite scales, but here there is less of a concern with replicability of results (the low return rates for the faculty and staff surveys already compromises the generalizability of these results) and more with simple description.

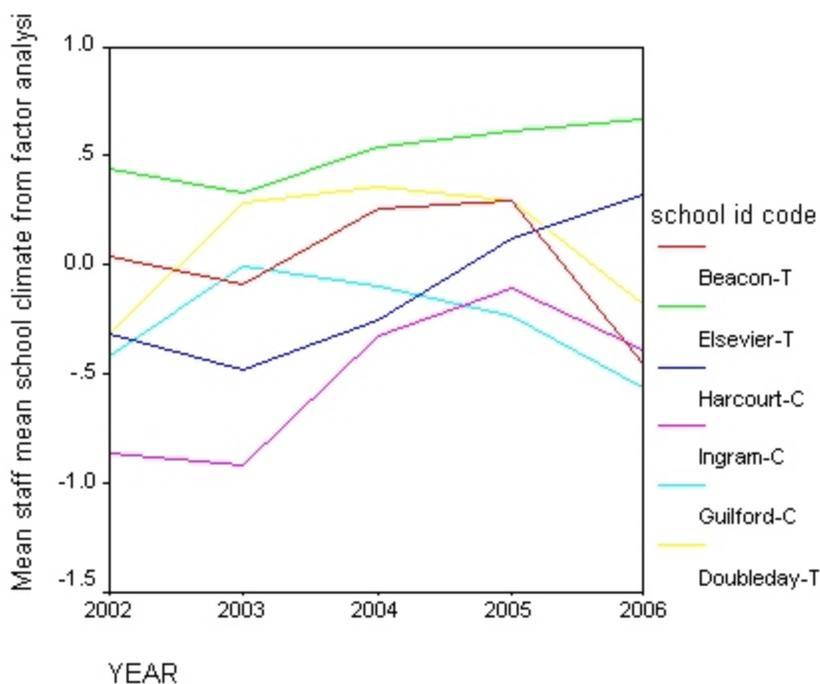


Figure 4: Changes in Faculty/Staff Perception of School Climate over Time - Elementary School

In Figure 4, note first that, although no differences were evident at baseline between the treatment and comparison schools based on the student surveys, here it appears that the three treatment schools (Beacon, Doubleday, and Elsevier) start out with higher (more favorable) scores on perceived school climate. The lowest score, by a considerable margin, is for one of the comparison schools (Ingram). To some extent this may be a matter of capacity building, or it may be related to the anticipation of the new intervention, or it may be attributable to factors which have nothing to do with the program. Second, note that Elsevier, the school which did best in implementing BPYS at the elementary school level, consistently has the highest score for faculty/staff perceptions of school climate. This may have been part of the reason for the more effective implementation at Elsevier across the five years of the study. In contrast, the other two treatment schools show a general increase in faculty/staff perceptions of school climate until the final year of implementation, then a sharp decline. The comparison schools also show a mix of patterns, with Harcourt and Ingram showing a general increase and Guilford an increase followed by a decline (with little overall change from the first to the fifth year of the study). There is certainly little evidence here that BPYS had any impact on faculty/staff perceptions of school climate. This and the fact that, for faculty and staff, school climate seems to be unidimensional, suggest that faculty and staff perceptions of school climate are driven by other considerations than the intervention itself. Further analysis, beyond the scope of the current report, may be of interest in further exploring whether faculty and staff perceptions of school climate are linked, at the composite scale or at the item-specific level, to implementation quality and to outcomes, but for the present, the limited faculty/staff school climate results presented here do not substantially affect the conclusions regarding the effectiveness of the program at the elementary school level.

Conclusion: The Impact of BPYS at the Elementary School Level

At the elementary school level, schools were well matched at pretest, allowing us to attribute subsequent differences between treatment and comparison schools to the impact of the intervention. As expected, based on the program goals and on past research on the intervention, BPYS did show evidence of achieving the goals stated for its three major components: it did appear to increase students' awareness of adults' discouragement of bullying; did appear to reduce bullying and related aggressive behaviors; and, at least weakly, did appear to increase the perception of the school as a safe place. The impact on bullying and other aggressive behaviors appears to be mediated at least in part through the impact of the program on students' attitudes toward aggression and, to perhaps a lesser extent, on the peer environment. The effects are quicker to materialize and more persistent over time with a strong rather than weak fidelity in program implementation.

Part 5: MIDDLE SCHOOL OUTCOME EVALUATION

As in the previous section, results in this section are presented in five parts. First, again, the results are analyzed at the item level, examining treatment and comparison schools at baseline, in the three years of active program implementation, and in the post-implementation year, using all of the items considered in the evaluation, and adjusting the inferential statistics for multiple testing using the modified Bonferroni procedure described in the previous section, to provide a conservative test of program impact and the greatest detail on where the program had or failed to have an impact. Second, the bivariate relationship between the intervention is presented for each of the multiple-item scales associated with the three main components of the program. Third, we use the average implementation scores presented in the process evaluation section to see whether quality of implementation has an impact above and beyond the simple treatment-comparison contrast. Finally, we again test the model presented in Figure 3, this time at the middle school level, using the multiple-item scales associated with the major components of the program as outcome measures plus additional controls for sociodemographic characteristics and the hypothesized intervening variables, peer group environment and one's own attitudes toward aggression and violence. For the same reasons as in the previous section, aggregate pretest characteristics and school characteristics are not explicitly included in the model because they are collinear with the treatment-comparison distinction, but we consider baseline (pre-implementation year) differences in school characteristics as potential influences on the findings. Finally, we briefly consider the impact of BPYS on faculty and staff perceptions of school climate at the middle school level.

Item-Level Results

Table 16 presents results for the middle school comparison, parallel to the elementary school comparison, of the BPYS intervention. The first thing to note in Table 16 is that there are several differences favoring the treatment schools in the baseline year, even with the conservative Bonferroni adjustment. In particular, indicators of school climate appear to be better for the treatment than for the comparison schools. This could indicate that the treatment schools were already on a favorable trajectory, or that they were more prepared to take advantage of whatever program was available to reduce bullying. If we take all of the differences favoring the treatment schools that existed in the baseline year, and eliminate from consideration any differences favoring the treatment schools *on those same items* in subsequent years, there still appear to be differences favoring the treatment schools in later years (as indicated by the asterisks in the columns for years 2-5; asterisks were eliminated from rows containing items on which there appeared to be an initial advantage for the treatment schools). It does appear that differences favoring the treatment schools emerge for several indicators of school climate in years 2-5; that there is some improvement in friends' attitudes toward aggression and violence in year 5, above and beyond any initial advantage in the baseline year; and that there are some, but inconsistent, improvements in witnessing bullying or aggression and in one's own attitudes and perpetration of bullying and aggression, particularly in the last (post-implementation) year; but because of the poor initial match at baseline, one cannot attribute these differences to the program with the same confidence as with the better-matched elementary school comparison.

Table 16: Middle Schools, Individual Items, Categorical Responses

Variables	2002	2003	2004	2005	2006
* = Statistically significant at $\alpha=.05$ using Holm's sequential method for familywise statistical significance. These differences are statistically significant by the most conservative criteria used here.	Somers' d (p) or Pearson's r (p)/ Favorable for... (T=treatment, C=Comparison, - = Neither)	Somers' d (p) or Pearson's r (p)/ Favorable for... (T=treatment, C=Comparison, - = Neither)	Somers' d (p) or Pearson's r (p)/ Favorable for... (T=treatment, C=Comparison, - = Neither)	Somers' d (p) or Pearson's r (p)/ Favorable for... (T=treatment, C=Comparison, - = Neither)	Somers' d (p) or Pearson's r (p)/ Favorable for... (T=treatment, C=Comparison, - = Neither)
A. School Climate					
1. I like school	d=-.088 (.172)	d=-.185 (.002) T *	d=-.239 (.000) T *	d=-.263 (.000) T *	d=-.346 (.000) T *
2. I look forward to going to school	d=-.059 (.380)	d=-.135 (.028) T	d=-.192 (.001) T *	d=-.274 (.000) T *	d=-.371 (.000) T *
3. I try hard in school	d=-.088 (.196)	d=-.040 (.499)	d=-.175 (.002) T *	d=-.137 (.019) T	d=-.158 (.005) T *
9. My teacher tells me when I do good job	d=-.251 (.000) T *	d=-.276 (.000) T	d=-.307 (.000) T	d=-.208 (.000) T	d=-.357 (.000) T
11. My teacher listens to me...	d=-.094 (.165)	d=-.167 (.007) T	d=-.315 (.000) T *	d=-.257 (.000) T *	d=-.280 (.000) T *
12. I have a teacher who cares about me	d=-.144 (.032) T	d=-.170 (.006) T	d=-.257 (.000) T	d=-.263 (.000) T	d=-.393 (.000) T
13. Adults teach us not to pick on oth. students	d=-.060 (.361)	d=-.321 (.000) T *	d=-.381 (.000) T *	d=-.380 (.000) T *	d=-.326 (.000) T *
14. Adults try hard to prevent bullying	d=-.083 (.195)	d=-.298 (.000) T *	d=-.290 (.000) T *	d=-.267 (.000) T *	d=-.338 (.000) T *
15. I like my teachers	d=-.114 (.082) (T)	d=-.139 (.023) T	d=-.320 (.000) T	d=-.233 (.000) T	d=-.376 (.000) T
16. People here respect all races	d=-.334 (.000) T *	d=-.354 (.000) T	d=-.334 (.000) T	d=-.300 (.000) T	d=-.247 (.000) T
17. People of my race can succeed here	d=-.086 (.170)	d=-.154 (.007) T	d=-.184 (.001) T *	d=-.131 (.017)T	d=-.021 (.699)
18. I feel lonely at school	d=-.043 (.502)	d=.009 (.877)	d=.009 (.878)	d=.011 (.849)	d=-.003 (.962)
19. I see graffiti here	d=.121 (.058) (T)	d=.087 (.164) X	d=.053 (.371) X	d=.127 (.025) T	d=.131 (.027) T
21. My school building is clean	d=-.237 (.000) T *	d=-.202 (.001) T	d=-.253 (.000) T	d=-.447 (.000) T	d=-.454 (.000) T

Variables	2002	2003	2004	2005	2006
* = Statistically significant at $\alpha=.05$ using Holm's sequential method for familywise statistical significance. These differences are statistically significant by the most conservative criteria used here.	Somers' d (p) or Pearson's r (p)/ Favorable for... (T=treatment, C=Comparison, - = Neither)	Somers' d (p) or Pearson's r (p)/ Favorable for... (T=treatment, C=Comparison, - = Neither)	Somers' d (p) or Pearson's r (p)/ Favorable for... (T=treatment, C=Comparison, - = Neither)	Somers' d (p) or Pearson's r (p)/ Favorable for... (T=treatment, C=Comparison, - = Neither)	Somers' d (p) or Pearson's r (p)/ Favorable for... (T=treatment, C=Comparison, - = Neither)
22. I like the way my school looks	d=-.357 (.000) T *	-.379 (.000) T	d=-.356 (.000) T	d=-.380 (.000) T	d=-.346 (.000) T
23. Students here obey the rules	d=-.305 (.000) T *	d=-.142 (.022) T	d=-.370 (.000) T	d=-.341 (.000) T	d=-.352 (.000) T
25. Rule breakers are treated the same	d=-.060 (.380)	d=-.118 (.063) (T)	d=-.040 (.507)	d=-.027 (.662)	d=.014 (.817)
26. Administrators respond appropriately-rules	d=-.114 (.080) (T)	d=-.097 (.118) X	d=-.059 (.322) X	d=-.119 (.041) T	d=-.128 (.026) T
27. I help decide activities and rules	d=-.271 (.000) T *	d=-.293 (.000) T	d=-.245 (.000) T	-.314 (.000) T	d=-.249 (.000) T
32. I care what teachers think of me	d=-.016 (.808)	d=-.092 (.149)	d=-.209 (.000) T *	d=-.112 (.057) (T)	d=-.254 (.000) T *
33. I respect teachers here	d=.034 (.600)	d=-.053 (.380)	d=-.072 (.210)	d=-.086 (.127)	d=-.164 (.003) T *
34. I respect the principal here	d=.067 (.280)	d=.082 (.149)	d=.099 (.075) (C)	d=.038 (.495)	d=-.014 (.806)
B. School Safety: Attitudes and Aggressive Behavior (Perpetration, Victimization, and Witnessing)					
35. I feel safe at my school	d=-.165 (.009) T	d=-.235 (.000) T	d=-.194 (.001) T	d=-.265 (.000) T	d=-.266 (.000) T
36. I feel safe on the school bus	d=-.144 (.059) (T)	d=-.125 (.083) (T)	d=-.170 (.013) T	d=-.274 (.000) T	d=-.187 (.006) T
37. I feel safe walking to school	d=-.144 (.059) (T)	d=-.208 (.003) T	d=-.220 (.001) T	d=-.320 (.000) T	d=-.255 (.000) T
38. Ever stay away because unsafe at school	r=.075 (.192)	r=.130 (.021) T	r=.062 (.248)	r=.015 (.770)	r=.124 (.017) T *
39. Ever stay away unsafe on way to school	r=-.039 (.533)	r=.005 (.925)	r=.076 (.150)	r=.058 (.237)	r=.077 (.143)
40. I have a friend who cares about me	d=-.144 (.012) T	d=-.116 (.028) T	d=-.103 (.057) (T)	d=-.017 (.759) X	d=-.094 (.073) (T)
42. My friends think wrong to hit	d=-.049 (.478)	d=-.076 (.238)	d=-.116 (.053) (T)	d=-.026 (.659)	d=-.129 (.030) T

Variables	2002	2003	2004	2005	2006
* = Statistically significant at $\alpha=.05$ using Holm's sequential method for familywise statistical significance. These differences are statistically significant by the most conservative criteria used here.	Somers' d (p) or Pearson's r (p)/ Favorable for... (T=treatment, C=Comparison, - = Neither)	Somers' d (p) or Pearson's r (p)/ Favorable for... (T=treatment, C=Comparison, - = Neither)	Somers' d (p) or Pearson's r (p)/ Favorable for... (T=treatment, C=Comparison, - = Neither)	Somers' d (p) or Pearson's r (p)/ Favorable for... (T=treatment, C=Comparison, - = Neither)	Somers' d (p) or Pearson's r (p)/ Favorable for... (T=treatment, C=Comparison, - = Neither)
43. Friends think OK to yell/say mean things	d=.095 (.163)	d=.028 (.665)	d=.082 (.170)	d=.060 (.305)	d=.166 (.004) T *
44. My friends think is OK to push and shove	d=.067 (.321)	d=.010 (.878)	d=.133 (.026) T	d=.127 (.028) T	d=.241 (.000) T *
45. My friends think OK to fight	d=.108 (.102)	d=-.037 (.559)	d=.101 (.086) (T)	d=.171 (.003) T *	d=.193 (.001) T *
46. Friends think wrong to call mean names	d=-.120 (.077) (T)	d=.042 (.512) X	d=-.079 (.199) X	d=-.052 (.381) X	d=-.157 (.008) T
47. Friends think wrong to get in physical fight	d=-.055 (.430)	d=-.033 (.604)	d=-.192 (.001) T *	d=-.138 (.021) T	d=-.130 (.029) T
48. Friends think OK to hit if hit first	d=.216 (.001) T *	d=.193 (.002) T	d=.154 (.009) T	d=.167 (.004) T	d=.157 (.007) T
49. Friends think OK to take out anger on other	d=.039 (.552)	d=.018 (.772)	d=.125 (.032) T	d=.142 (.011) T	d=.139 (.014) T *
54. I saw other students in a fight	r=.040 (.507)	r=.199 (.000) T *	r=.350 (.000) T *	r=.135 (.014) T	r=.210 (.000) T *
55. I saw other student get physically attacked	r=.070 (.262)	r=.019 (.739)	r=.094 (.087) (T)	r=.004 (.939)	r=.119 (.028) T
56. I saw other student get harassed	r=.167 (.006) T	r=.107 (.065) (T)	r=.097 (.082) (T)	r=.002 (.974) X	r=.231 (.000) T
57. I saw someone threaten to hit	r=.042 (.484)	r=.069 (.172)	r=.210 (.000) T *	r=.165 (.002) T *	r=.155 (.004) T *
58. I saw student with gun at school	r=.127 (.007) T	r=.137 (.014) T	r=.123 (.014) T	r=.121 (.008)) T	r=.199 (.000) T
59. Saw student with weapon besides gun	r=.066 (.259)	r=.249 (.000) T *	r=-.001 (.981)	r=.144 (.005) T *	r=.195 (.000) T *
60. I encouraged other students to fight	r=.050 (.396)	r=-.041 (.472)	r=.116 (.028) T	r=.088 (.052) (T)	r=.111 (.034) T
61. I pushed, shoved, hit, etc.	r=.005 (.928)	r=-.118 (.041) C	r=-.049 (.370)	r=.054 (.306)	r=.049 (.365)
62. I got into physical fight to get something	-.074 (.251)	r=-.035 (.547)	r=.056 (.291)	r=.032 (.541)	r=.123 (.017) T *

Variables	2002	2003	2004	2005	2006
* = Statistically significant at $\alpha=.05$ using Holm's sequential method for familywise statistical significance. These differences are statistically significant by the most conservative criteria used here.	Somers' d (p) or Pearson's r (p)/ Favorable for... (T=treatment, C=Comparison, - = Neither)	Somers' d (p) or Pearson's r (p)/ Favorable for... (T=treatment, C=Comparison, - = Neither)	Somers' d (p) or Pearson's r (p)/ Favorable for... (T=treatment, C=Comparison, - = Neither)	Somers' d (p) or Pearson's r (p)/ Favorable for... (T=treatment, C=Comparison, - = Neither)	Somers' d (p) or Pearson's r (p)/ Favorable for... (T=treatment, C=Comparison, - = Neither)
64. I acted cold or gave silent treatment	r=.047 (.435)	r=-.013 (.824)	r=.131 (.016) T	r=.053 (.329)	r=.011 (.846)
65. I harassed another student	r=.008 (.898)	r=-.032 (.584)	r=.065 (.230)	r=-.165 (.004) C	r=.142 (.007) T *
66. I tried to exclude others from my group	r=.037 (.541)	r=-.053 (.362)	r=-.038 (.410)	r=.016 (.763)	r=.006 (.917)
67. I threatened to hit or hurt another student	r=.060 (.314)	r=-.102 (.077) (C)	r=-.048 (.375)	r=.065 (.213)	r=.027 (.618)
68. I was mean when I was angry	r=.061 (.316)	r=.039 (.499)	r=-.044 (.423)	r=.133 (.011) T	r=.062 (.249)
69. I said bad things to hurt reputation	r=-.028 (.646)	r=.050 (.382)	r=.064 (.232)	r=.055 (.294)	r=.065 (.226)
70. I carried a gun to school	r=-.019 (.763)	r=.125 (.023) T	r=.103 (.043) T	r=.048 (.326)	r=.079 (.123)
71. I ganged up on someone	r=-.047 (.446)	r=-.034 (.557)	r=-.003 (.950)	r=.018 (.342)	r=.136 (.009) T *
72. Another student encouraged me to fight	r=.017 (.783)	r=.035 (.549)	r=.077 (.153)	r=.096 (.067) (T)	r=.112 (.036) T
73. Another student physically attacked me	r=.102 (.090) (T)	r=.058 (.626)	r=-.093 (.086) (C)	r=-.043 (.418)	r=.010 (.854)
74. I was harassed by another student	r=.057 (.338)	r=.048 (.407)	r=-.003 (.963)	r=-.098 (.071) (C)	r=.052 (.335)
75. Another student threatened to hurt me	r=.067 (.262)	r=.023 (.692)	r=.048 (.373)	r=.008 (.880)	r=.067 (.217)
76. Classmate cold/gave me silent treatment	r=-.026 (.669)	r=-.033 (.567)	r=.018 (.740)	r=-.025 (.645)	r=-.055 (.310)
77. Classmate kept me out of their group	r=-.022 (.717)	r=-.048 (.404)	r=-.151 (.006) C *	r=-.110 (.044) C	r=-.122 (.024) C
78. Classmate said bad things to hurt my rep	r=-.019 (.759)	r=.020 (.726)	r=-.104 (.057) (C)	r=-.015 (.780)	r=.076 (.158)
79. Students ganged up against me	r=.013 (.832)	r=-.040 (.486)	r=-.075 (.175)	r=.017 (.753)	r=.011 (.840)

Variables	2002	2003	2004	2005	2006
* = Statistically significant at $\alpha=.05$ using Holm's sequential method for familywise statistical significance. These differences are statistically significant by the most conservative criteria used here.	Somers' d (p) or Pearson's r (p)/ Favorable for... (T=treatment, C=Comparison, - = Neither)	Somers' d (p) or Pearson's r (p)/ Favorable for... (T=treatment, C=Comparison, - = Neither)	Somers' d (p) or Pearson's r (p)/ Favorable for... (T=treatment, C=Comparison, - = Neither)	Somers' d (p) or Pearson's r (p)/ Favorable for... (T=treatment, C=Comparison, - = Neither)	Somers' d (p) or Pearson's r (p)/ Favorable for... (T=treatment, C=Comparison, - = Neither)
80. I was in physical fight at school	r=.103 (.074) (T)	r=-.040 (.487)	r=.048 (.377)	r=.047 (.377)	r=.063 (.241)
81. I was threatened with weapon	r=.096 (.043) T	r=.081 (.154) X	r=-.028 (.260) X	r=.124 (.004) T	r=.113 (.027) T
82. I was injured in fight at school	r=-.004 (.945)	r=.007 (.902)	r=.002 (.909)	r=.084 (.071) (T)	r=.102 (.047) T
83. How many students get picked on	d=.132 (.050) T	d=.106 (.089) (T)	d=.200 (.001) T	d=.058 (.318) X	d=.140 (.016) T
84. How many students pick on others	d=.148 (.030) T	d=.117 (.027) T	d=.196 (.001) T	d=.103 (.082) (T)	d=.097 (.103) X
85. How many kids afraid of you b/c mean	d=.035 (.591)	d=.012 (.846)	d=.087 (.129)	d=.107 (.059) (T)	d=.061 (.267)
86. How many kids do you pick on often	d=.130 (.053) (T)	d=-.094 (.121)	d=-.002 (.970)	d=-.077 (.170)	d=-.027 (.622)
87. How many kids pick on you often	d=.100 (.138)	d=-.056 (.358)	d=-.119 (.042) C	d=-.171 (.003) C *	d=-.115 (.040) C
88. How many kids do you fear b/c mean	d=.043 (.497)	d=-.021 (.714)	d=-.045 (.390)	d=.024 (.624)	d=-.033 (.497)
100. I think it is wrong to hit other people	d=-.003 (.969)	d=-.017 (.787)	d=-.061 (.310)	d=-.105 (.075) (T)	d=-.124 (.035) T
101. OK to yell or say mean things to others	d=.040 (.544)	d=.052 (.409)	d=.068 (.241)	d=.055 (.347)	d=.124 (.031) T
102. It is OK to push and shove if you are mad	d=.030 (.648)	d=.059 (.331)	d=.033 (.566)	d=.128 (.023) T	d=.159 (.004) T *
103. It is wrong to call others mean names	d=.058 (.408)	d=-.027 (.677)	d=-.040 (.513)	d=-.118 (.047) T	d=-.157 (.007) T *
104. It is OK to take out anger on others	d=-.032 (.613)	d=-.004 (.949)	d=.118 (.037) T	d=.103 (.067) (T)	d=.142 (.010) T *
105. It is OK to fight to get what you want	d=-.043 (.472)	d=-.056 (.338)	d=.053 (.337)	d=.086 (.121)	d=.130 (.017) T *
106. It is OK to hit if they hit your first	d=.174 (.012) T	d=.107 (.089) (T)	d=.116 (.051) (T)	d=.186 (.001) T t=3.200, 349	d=.132 (.023) T

Variables	2002	2003	2004	2005	2006
* = Statistically significant at $\alpha=.05$ using Holm's sequential method for familywise statistical significance. These differences are statistically significant by the most conservative criteria used here.	Somers' d (p) or Pearson's r (p)/ Favorable for... (T=treatment, C=Comparison, - = Neither)	Somers' d (p) or Pearson's r (p)/ Favorable for... (T=treatment, C=Comparison, - = Neither)	Somers' d (p) or Pearson's r (p)/ Favorable for... (T=treatment, C=Comparison, - = Neither)	Somers' d (p) or Pearson's r (p)/ Favorable for... (T=treatment, C=Comparison, - = Neither)	Somers' d (p) or Pearson's r (p)/ Favorable for... (T=treatment, C=Comparison, - = Neither)
C. Home and Family Environment					
107. My parents want me to get good grades	d=-.000 (1.000)	d=-.012 (.741)	d=-.010 (.791)	d=-.006 (.838)	d=-.084 (.109)
108. I can tell my parents how I feel	d=-.088 (.176)	d=-.109 (.065) (T)	d=-.100 (.087) (T)	d=-.029 (.554)	d=-.125 (.026) T
109. I like to do things with my family	d=-.111 (.053) (T)	d=-.073 (.203) X	d=-.100 (.066) (T)	d=-.026 (.628) X	d=-.120 (.024) T
112. Parents know who I am with if I'm away	d=-.063 (.274)	d=.029 (.602)	d=-.013 (.812)	d=-.049 (.323)	d=-.015 (.769)
113. Parents limit how much TV I watch	d=-.036 (.519)	d=-.042 (.504)	d=-.074 (.214)	d=.027 (.659)	d=-.104 (.081) (T)
114. Parents know who my friends are	d=-.023 (.683)	d=-.030 (.568)	d=-.012 (.814)	d=-.032 (.525)	d=-.024 (.610)
115. Parents let me know if I do a good job	d=-.152 (.014) T	d=-.169 (.003) T	d=-.125 (.023) T	d=-.038 (.478) X	d=-.093 (.071) (T)
116. Will always be people I can count on	d=-.074 (.171)	d=-.072 (.142)	d=-.095 (.055) (T)	d=-.033 (.477)	d=-.071 (.175)
117. Besides family there is an adult I can trust	d=-.114 (.035) T	d=-.079 (.123) X	d=-.004 (.942) X	d=-.052 (.320) X	d=.018 (.718) X
118. I believe there is some good in everybody	d=-.007 (.907)	d=.053 (.373)	d=.097 (.092) (C)	d=.039 (.500)	d=-.104 (.070) (T)

X = Nonsignificant difference after an initial statistically significant difference or marginally significant difference favoring the treatment schools. Special note of this is made in this table because of the large number of differences favoring the treatment schools at baseline.

Program Components and Hypotheses: Composite Scale Outcomes

Table 17 presents the results for the composite scales described in the previous section. As in the previous tables, results are presented for each year separately. Here as in the previous section, Pearson's r is used to measure the strength of the relationship between the program and the outcome, and the test of statistical significance for Pearson's r is used to assess the statistical significance of the differences between treatment and comparison schools. The same modified Bonferroni procedure as was used in the previous tables is applied in the same way (separately by year) in Table 17, and use of robust standard errors was again explored as well (see note at bottom of Table 17). For the middle school sample, data are available on both prevalence (yes or no) and frequency (how many times; a natural logarithm transformation has been applied to reduce skewness) of physical and relational perpetration and victimization are presented. To maintain comparability between the elementary school and the middle school analysis, only data on prevalence are actually included in the description and in the Bonferroni correction, but the results would be the same if data on the frequency were used instead, and the frequency data are included to indicate this result in Table 17. As with the elementary school analysis, it is expected that in the baseline year, there will be no statistically significant differences between treatment and comparison schools, and in contrast to the item-level results in Table 16, once the modified Bonferroni adjustment is made, none of the differences in Table 17 for the baseline year (2002) is statistically significant (witnessing bullying has $p=.043$, but this falls short of statistical significance based on the adjusted $\alpha=.006$).

As in Table 16, there is strong evidence for the impact of BPYS on students' perceptions that bullying is discouraged at their school, and the impact lasts into the post-intervention year. The first component of the program, then, appears to be successful both here and at the item level. Evidence for the second component is weaker. None of the physical or relational aggression perpetration or victimization scales is significantly different between the treatment and comparison schools, and this is not entirely out of line with the results in Table 16. More out of line with the results in Table 16, and more favorable to the intervention, witnessing bullying at the middle school level appears to be lower in treatment than in comparison schools in the second year of implementation (2004) and the post-intervention year (2006); and the differences are in the right direction and would be significant for independent comparisons, but do not meet the criteria for significance using the modified Bonferroni adjustment for the first and third years of implementation (for 2003, $p=.012$ but adjusted $\alpha=.008$; and for 2005, $p=.048$ but adjusted $\alpha=.010$). With regard to the third major component of the program, in Table 16, initial differences in school safety led us not to draw any conclusions about the effectiveness of BPYS despite the significantly higher levels of perceived school safety in the treatment schools during the intervention and post-intervention years; but here, the difference in the baseline year is not statistically significant (even before the Bonferroni adjustment), so it seems reasonable to conclude, based on the results in Table 17, that BPYS does indeed increase perceived school safety at the middle school level. Finally, BPYS appears to have a favorable impact on both peer environment and one's own attitudes toward aggression in the final year of implementation (2005) and in the post-implementation year (2006), a result consistent with the item-level results in Table 16 for the post-implementation year, but better than suggested by the results in Table 16 for 2005. To summarize the results for the composite scales:

Table 17: Bivariate Analysis of Program Impact - Middle School

Outcomes: Pearson's r (p) (Treatment=1, Comparison=2)	Year				
	2002	2003	2004	2005	2006
Bullying discouraged	-.044 (.494)	-.318* (.000)	.198* (.000)	-.367* (.000)	-.311* (.000)
Witnessed bullying	.131 (.043)	.156 (.012)	.275* (.000)	.112 (.048)	.230* (.000)
Physical aggression perpetration (prevalence)	.040 (.619)	-.128 (.039)	.009 (.885)	.080 (.160)	.049 (.398)
Physical aggression perpetration (log frequency)	-.032 (.622)	-.057 (.361)	-.032 (.596)	.049 (.390)	.098 (.089)
Physical aggression victimization (prevalence)	.060 (.356)	.032 (.609)	-.022 (.709)	-.031 (.587)	.040 (.493)
Physical aggression victimization (log frequency)	.067 (.298)	.030 (.632)	-.033 (.582)	-.040 (.479)	.041 (.475)
Relational aggression perpetration (prevalence)	.019 (.766)	-.009 (.885)	.092 (.120)	.094 (.097)	.092 (.109)
Relational aggression perpetration (log frequency)	.056 (.868)	-.010 (.878)	.081 (.173)	.091 (.107)	.062 (.280)
Relational aggression victimization (prevalence)	.014 (.833)	.036 (.560)	-.053 (.377)	-.027 (.635)	-.003 (.954)
Relational aggression victimization (log frequency)	.047 (.729)	.033 (.600)	-.051 (.393)	-.031 (.587)	.008 (.886)
Perceived school safety	-.064 (.324)	-.186* (.003)	-.108 (.069)	-.163* (.004)	-.158* (.006)
Peer environment (perceived peer attitudes toward aggression)	.073 (.260)	.012 (.844)	.146 (.014)	.150* (.008)	.165* (.004)
Own attitude toward aggression	-.024 (.713)	-.007 (.907)	.081 (.173)	.174* (.002)	.170* (.003)

* Statistically significant at a $\alpha=.05$ (familywise, i.e., across all comparisons adjusting for nonindependent repeated testing) using Simes' modified Bonferroni procedure in conjunction with Holm's sequential Bonferroni test for the significance level of Pearson's r. Again the use of robust standard errors adjusting for clustering of students within schools was also explored. For both approaches, the same 13 of 14 relationships for which $r > .150$ were identified as statistically significant; and only 6 were identified as statistically significant using both. The modified Bonferroni approach identified none of the relationships for which $r < .150$ as statistically significant; but the robust standard error approach identified 5 relationships for which $r < .150$ (3 for which $r < .100$) as statistically significant.

Note: At the suggestion of a reviewer, data were analyzed to see whether the program effect differed by ethnicity. Based on the modified Bonferroni test, none of the interactions between ethnicity and program impact was statistically significant at $\alpha=.05$.

(1) *The hypothesis that BPYS implementation results in students' recognizing that bullying is being discouraged appears to be supported at the middle school level, as it was at the elementary school level.*

(2) *The hypothesis that BPYS reduces bullying and related behaviors appears to be supported for bullying but not for other behaviors, and even for bullying, this support is relatively weak; but the weakness for this support is because at the item level, treatment schools already appeared to have some advantage over comparison schools. The weakness in support for this hypothesis appears to be more a problem in the execution of the research (in particular, of the loss of treatment and comparison middle schools) than of the program.*

(3) *The hypothesis that BPYS results in perceptions of increased safety at school is at least weakly supported; the evidence for this hypothesis is good in the analysis of the composite school safety scale, but initial differences at the item level make us hesitant to draw conclusions on this hypothesis at the item level.*

(4) *BPYS appears to have a favorable impact on attitudes toward physical and relational aggression and on perceived peer environment in the late and post-implementation stages of the program.*

Quality of Implementation and Overall Impact

Moving now from the discussion of program outcomes to the context of those outcomes, we consider once again the effect of quality of implementation on program outcomes. The procedure for constructing a score for quality of implementation is the same as for the elementary schools. Quality of implementation was coded as zero for all comparison school in all years, indicating that they were not implementing BPYS in that school; and it was also coded as zero for the treatment schools in the pre-implementation year, similarly indicating that they were not implementing the program in that year. For the three years in which the program was actively being implemented, the average implementation scores from the process evaluation were used; and as before, for the post-implementation year, each treatment school was assigned the mean implementation score over all three years of implementation. Once again, we focus on the composite scales for the outcomes of the three major components of the program (peer environment and attitudes toward aggression are omitted here), and again the data for all five years (pre-implementation, three years of implementation, and post-implementation) were pooled to allow for greater variation in implementation. The results are presented in Table 18. Here again, the emphasis is on the descriptive rather than the inferential statistics.

Overall, these results are consistent with the results of the earlier analysis, but in contrast to the same analysis in the previous section, there appears to be no added value based on quality of implementation. Most of the comparisons of correlations actually favor the simple treatment-comparison contrast, rather than the quality of implementation score, as being more predictive of the outcomes. Two reasons why this might be the case come immediately to mind. First, with fewer schools at the middle school level, there is less variation in quality of program implementation. Second, the apparent advantage of the treatment-comparison contrast over the quality of implementation score as a predictor of program outcomes may be attributable to the

problem that plagues every aspect of the middle school analysis, the pre-existing differences in outcome measures, most evident in the item-level analysis, that favors the treatment middle schools from the outset. The fact that the treatment schools start with some apparent advantage over the comparison schools will be reflected in the treatment-comparison contrast but not in the quality of implementation score. Although the differences were not statistically significant in the baseline year for the composite scale outcomes, in the present context, this possibility can not be ruled out as an explanation for the results in Table 18.

Table 18: Quality of Implementation and BPYS Program Goals - Middle School

Outcome Pearson's r (p) (Treatment=1, Comparison=2)	Quality of Implementation Score	Program Treatment vs. Comparison
Bullying discouraged	.285 (.000)	.292 (.000)
Witnessed bullying	-.173 (.000)	-.185 (.000)
Physical aggression perpetration	-.056 (.024)	-.056 (.025)
Physical aggression victimization	.006 (.795)	.011 (.672)
Relational aggression perpetration	-.004 (.870)	-.018 (.468)
Relational aggression victimization	-.034 (.177)	-.028 (.266)
Perceived school safety	.140 (.000)	.145 (.000)

Multivariate Analysis of the Impact of BPYS

Table 19 presents the results of testing the model in Figure 3. In general, the results are similar to those in Table 15 for the elementary schools. The best predictors of victimization are peer attitudes (although this relationship diminishes in later years) and, for relational aggression, gender (with females being more likely to report being victims of relational aggression than males). Peer attitudes, but here not one's own attitudes, are also predictive of witnessing aggression and bullying. Peer attitudes and one's own attitudes are, as expected, the most consistent predictors of perpetration of physical aggression and, along with gender (with females being more likely perpetrators) of relational aggression. Similar results are obtained when frequency data are used; in Table 19, prevalence data are presented. Feelings of safety at school, peer environment, one's own attitudes toward aggression and violence, and also whether bullying is discouraged at school are all significantly related to family bonding. The relationship of family bonding to whether bullying is discouraged at school here and for the elementary schools may actually reflect an influence of BPYS on family bonding (one component of the program is family-directed). One's own and one's friends' attitudes toward aggression and violence are also related to class in school and school grades, with older students and students with lower grades having less unfavorable attitudes toward aggression and violence.

Table 19: Predictors of Program Outcomes (Standardized Regression Coefficients): Middle School

Dependent Variable	R ²	Sex	White	Other	Class in School	Grades (A-F)	Family Bonding	Treatment/Comparison	Peer Attitudes	Attitudes re Aggression
<u>Physical aggression victim</u>										
Year 1	.075	.078	.145*	.061	-.115	.061	-.069	.065	.211**	-.068
Year 2	.127	.116	.105	.025	-.111	-.063*	.065	.062	.246**	-.014
Year 3	.098	.029	.137*	.028	-.102	-.118	-.013	-.041	.245**	-.063
Year 4	.100	.105	.127*	.052	-.184**	-.025	-.108	-.038	.133	.058
Year 5	.028	-.005	-.012	.007	-.097	-.034	.009	.022	.152	.007
<u>Relational aggression victim</u>										
Year 1	.147	-.213**	.325**	.173*	-.155	-.216**	-.064	.021	.152*	-.016
Year 2	.159	-.174**	.065	.091	-.124*	-.158*	-.026	.024	.343**	-.076
Year 3	.093	-.177**	.106	.016	-.086	-.086	-.008	-.087	.180*	.095
Year 4	.086	-.203**	.053	.057	-.131*	-.002	-.133*	-.035	.125	.020
Year 5	.056	-.176**	-.013	.060	-.054	-.039	-.062	-.025	.148	-.006
<u>Physical aggression perpetrator</u>										
Year 1	.336	.132*	.036	-.038	-.050	-.193**	-.160**	.020	.113	.297**
Year 2	.366	.045	-.013	.034	-.090	-.137*	.036	-.120*	.189**	.410**
Year 3	.370	.130*	.028	.008	-.087	-.122*	-.077	-.060	.205**	.313**
Year 4	.354	.019	.008	-.059	-.119*	-.138**	-.091	-.023	.233**	.310**
Year 5	.297	-.080	-.171**	-.092	-.060	-.007	-.035	-.062	.302**	.246**
<u>Relational aggression perpetrator</u>										
Year 1	.349	-.249**	.062	.158**	-.076	-.001	-.202**	-.013	.219**	.315**
Year 2	.287	-.178**	-.014	-.007	-.083	-.057	-.169**	-.050	.233**	.271**
Year 3	.300	-.183**	.065	.029	-.101	-.068	-.048	.023	.169*	.401**
Year 4	.340	-.242**	.055	-.037	-.127**	-.060	-.012	-.008	.275**	.336**
Year 5	.267	-.230**	-.041	.023	.006	-.061	-.005	.003	.237**	.266**
<u>Witnessed aggression/bullying</u>										
Year 1	.112	-.147*	.167*	-.003	-.110	-.003	.023	.143*	.189*	.109
Year 2	.172	.041	.121	.084	.003	-.012	-.034	.175**	.326**	.036
Year 3	.173	-.077	.178**	.047	-.022	-.072	.059	.269**	.189*	.131
Year 4	.087	-.058	-.077	-.131*	-.142*	.078	-.087	.056	.136	.050
Year 5	.133	-.103	.046	.036	.072	.085	-.074	.180**	.196*	.041

Dependent Variable	R ²	Sex	White	Other	Class in School	Grades (A-F)	Family Bonding	Treatment/Comparison	Peer Attitudes	Attitudes re Aggression
<u>Feel safe at school</u>										
Year 1	.098	.029	-.024	.002	.230**	.074	.191**	.089	-.093	.026
Year 2	.230	.157**	.011	.084	.157**	.170**	.132*	.020	-.268**	-.059
Year 3	.205	-.027	-.116	.008	.219**	.055	.244**	.147**	-.106	-.103
Year 4	.119	-.065	-.014	-.037	.115*	-.025	.223**	.143*	-.186*	.038
Year 5	.091	.066	.038	-.039	.084	-.171**	.154*	.152**	-.257**	.106
<u>Bullying discouraged at school</u>										
Year 1	.132	.051	-.132	-.124	-.099	.062	.223**	-.025	-.059	-.098
Year 2	.275	-.050	-.093	-.047	-.049	.017	.183**	-.319**	-.086	-.227**
Year 3	.291	-.102	-.050	-.030	-.105*	.003	.242**	-.295**	-.110	-.091
Year 4	.288	.038	-.036	-.159**	-.071	.013	.188**	-.343**	-.256**	.008
Year 5	.305	.042	-.044	-.044	.002	-.088	.139**	-.231**	-.375**	-.061
<u>Own attitudes toward aggression</u>										
Year 1	.225	.172**	-.018	-.095	.075	-.113	-.369**	-.071	NA	NA
Year 2	.219	.211**	-.133*	-.058	.156**	-.167**	-.281**	-.037	NA	NA
Year 3	.367	.184**	-.088	-.121*	.108*	-.104*	-.471**	-.014	NA	NA
Year 4	.218	.177**	-.064	-.085	.125*	-.120*	-.320**	.133*	NA	NA
Year 5	.228	.103*	-.138*	.039	.150**	-.084	-.304**	.148**	NA	NA
<u>Friends attitudes toward aggression</u>										
Year 1	.197	.165**	.082	-.049	.169**	-.134*	-.268**	.043	NA	NA
Year 2	.149	.150*	-.012	-.093	.132*	-.073	-.290**	-.015	NA	NA
Year 3	.245	.204**	.045	-.018	.116*	-.219**	-.260**	.102	NA	NA
Year 4	.218	.129*	-.058	-.030	.087	-.067**	-.304**	.117*	NA	NA
Year 5	.280	.163**	-.084	-.012	.279**	-.150**	-.242**	.134**	NA	NA
* p ≤ .050 ** p ≤ .010	Year 1 n = 240 Year 2 n = 259 Year 3 n = 285 Year 4 n = 313 Year 5 n = 303									

As would be expected, there is no apparent impact of BPYS on whether bullying is discouraged at school in the pre-implementation year, but the effect is strong and statistically significant in subsequent years. As with the elementary schools, the impact of BPYS on perceptions of school safety are not as expected (in fact, overall, comparison schools report better perceptions of school safety), but again, this is because the scale includes both safety at school and safety on the bus or on the way to school. When safety on the bus or on the way to school are separated, school safety is perceived as being higher in the treatment than in the comparison schools.

If BPYS has an impact, at least some of that impact appears to be indirect, via the impact of BPYS on one's friends' and one's own attitudes toward aggression and violence, an impact which appears to occur in the later years of implementation for the middle schools. BPYS also appears to result in less witnessing of aggression or bullying.

Supplemental to these results, and not presented in detail here, we also examined the impact of BPYS on items not parallel to the elementary school items. Briefly, applying the modified Bonferroni criteria described above, there seemed to be little evidence of collateral effects of BPYS. In particular, there appeared to be little effect (and none was expected) of BPYS on substance use in the present study.

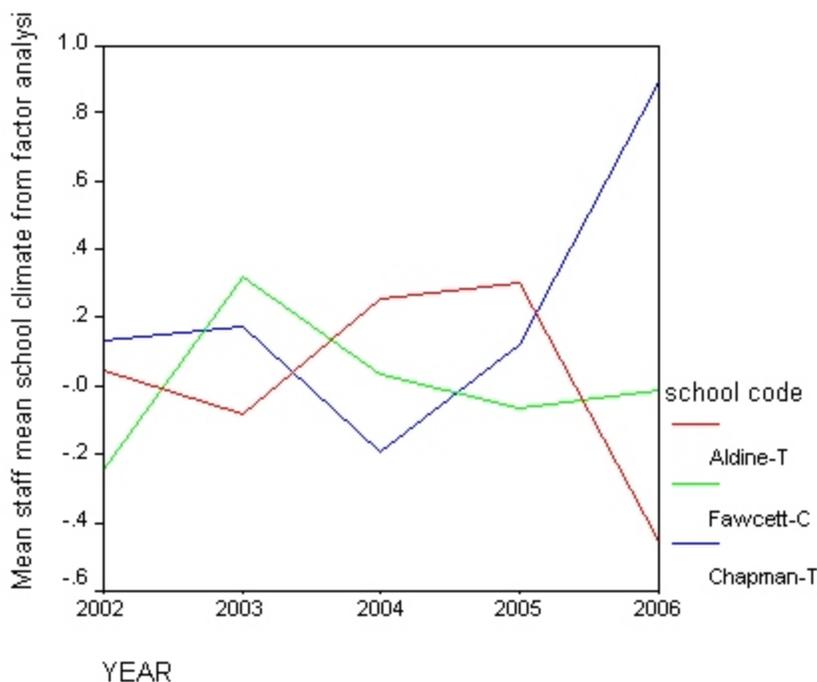


Figure 6: Changes in Faculty/Staff Perception of School Climate over Time - Middle School

BPYS and Faculty and Staff Perceptions of School Climate

Parallel to the examination of faculty/staff perceptions of school climate in the previous section, Figure 5 presents the trajectories of change over time in faculty/staff perceptions of the school climate at the middle school level, using the same factor analytic approach to the construction of the unidimensional faculty/staff school climate variable as before. As seen in Figure 5, the two treatment schools, Aldine and Chapman, appear to be more different from one another than either is from the comparison school, Fawcett. Fawcett begins with the lowest level of perceived school climate, increases slightly in 2003, declines almost as much in 2004, and ends up close to where it began by 2006. Aldine shows a slight decline in 2003, an increase in the second and third years of program implementation, then a very sharp decline leaving it well below its baseline level by 2006. Chapman begins with the highest score, fluctuates during the years of program implementation, then skyrockets in the post implementation year. More detail on the schools can be found in the process evaluation section, but here, as for the elementary schools, there appears to be no impact of BPYS on faculty/staff perceptions of school climate.

Conclusion: The Impact of BPYS at the Middle School Level

At the middle school level, there is reasonably strong evidence that the first major component of the program was successful in creating an atmosphere in which students knew that bullying was discouraged. Support for the effectiveness of the second program was weak, perhaps at least in part because of problems in executing the study, and in contrast to the more general effects of BPYS at the elementary school level, the effects of BPYS at the middle school level, to the extent to which those effects could be established, seemed to be more specific to bullying as opposed to physical and relational aggression more generally. Support for the third component of the program, aimed at creating a perception that the school is a safe place, appeared to be good in the analysis of the composite scale outcomes, but was problematic in the analysis of the item-level data.

It is frustrating, but all of these results must be qualified by noting that the match between treatment and comparison schools appears to have favored the treatment schools at the outset. Part of the problem is, as noted earlier, that one of the school districts that had initially agreed to participate in the evaluation dropped out early in the process. It would be most desirable to do an evaluation with a larger number of middle schools. At worst, it appears from the present results that the BPYS intervention did no harm at the middle school level; and at best, the differences between treatment and comparison middle schools on items on which they did not differ at baseline provide some evidence for the favorable impact of BPYS.

Part 6: CONCLUSION

Bully-Proofing Your School is a curriculum-based intervention designed to reduce aggressive, violent, and bullying behavior in the schools. Evidence from the present evaluation at the middle school level is limited by the weak match between treatment and comparison schools at baseline, but to the extent that it provides any evidence at all about the effectiveness of the program, it appears that the program at worst does no harm and at best may have beneficial effects on school climate, attitudes toward aggression and violence, and rates of perpetration of and victimization by relational and physical aggression. To reiterate, however, these results do not provide a sufficient basis for recommending the adoption of the program at the middle school level, not because the program itself appears to have failed, but because the research design was compromised during the evaluation process and as a result was not adequate to draw firm conclusions about the effectiveness of the program.

We can be more confident about our results at the elementary school level. The strong match, with practically no differences in the survey items between the treatment and comparison schools at baseline, may not be conclusive proof that the schools were equivalent, but it would be a huge coincidence if we happened to catch treatment and comparison schools on opposite trajectories just at the precise times those trends crossed. More plausible is that the treatment and comparison schools were comparable at baseline, and that the differences between the treatment and comparison schools after the baseline year, during and subsequent to implementation, were the result of the intervention.

These results indicate that BPYS does appear to have a favorable impact on school climate; on attitudes toward aggression and violence; and on perceived and directly experienced rates of perpetration of and victimization by relational and physical aggression. Based on the process evaluation and the comparison of the different treatment schools, it also appears that fidelity of program implementation is important (this is particularly evident at the elementary school level, and can not be discounted at the middle school level), with favorable results appearing earlier in the implementation process, being more pervasive across a broader range of specific outcomes, and persisting more strongly after implementation (once technical support from the provider has been withdrawn and the school is on its own to continue the program), when the implementation has been stronger at the outset. The results of the process evaluation indicate that when schools did not implement the program well, it was primarily because the principal was not fully engaged in the program and did not foster strong buy-in from the cadre and the school staff. The fact that several of the schools experienced difficulty in implementation suggests that in a real-world implementation, results may not be as favorable as those experienced in the elementary school with the best implementation, but at least while the program is actively being implemented, there do appear to be favorable results. The elementary school results suggest that BPYS is a promising program for implementation at the elementary school level, and that although it would be advisable to further evaluate the program, this study coupled with previous research on the program make it likely that future results would add further evidence that the program is effective in reducing bullying in the schools.

With regard to the criteria advanced by the Blueprints for Violence Prevention project, it appears from the present evaluation that the term “promising” may be appropriate in that context as well as more generally, at the elementary but not the middle school level. The requirement of a quasi-experimental design with adequate matching at baseline has been met, along with demonstratedly adequate levels of inter-rater reliability in the process evaluation, reliability of the composite scales used as outcomes, and consistency in the timing of the administration of the measurement instruments to program participants. At the elementary school level, there is clear evidence of success in meeting the objectives of the three major components of BPYS, including awareness that bullying was being discouraged, reductions in bullying and related aggressive behaviors, and perceptions of greater school safety; and in addition, BPYS appears to affect known risk factors for aggression and violence (peer environment and own attitudes toward aggression). At least some of these effects (particularly witnessing bullying, among the composite variables at the elementary school level) appear to be sustained even after the active involvement of the program provider in the intervention has been terminated. Still missing is an adequate multiple site replication with a similar research design and comparable results. At the middle school level, problems in matching treatment and comparison schools at baseline render the findings problematic, but the fact that the results, although flawed, are generally favorable to the program do suggest that further research on BPYS at the middle school level is warranted.

Recommendations for Future Research

One limitation of the present evaluation indicated above is the limited confidence we can place in conclusions regarding the middle schools. It would be desirable to design a study with a larger number of middle schools, matched as closely as possible on general school characteristics and on variable important to the outcome evaluation, to see whether BPYS can be recommended as a middle school intervention with the same degree of confidence as we are presently able to conclude that it has beneficial effects at the elementary school level.

Faculty and staff input provided valuable information regarding the process evaluation, particularly the difficulties in implementation and fidelity to program design, in the present study. Further insights may be obtained by more detailed examination of faculty and staff data.

Some analysis has been presented here regarding the impacts of family environment, peer group environment, and personal attitudes about aggression and violence as predictors of or influences on one's own experience as a victim or perpetrator of relational or physical aggression. While this is sufficient for the present purpose of evaluating BPYS in the context of family, school, and peer group climate, it would also be helpful to more fully examine the relationships among these variables and behavioral (victimization and perpetration of relational and physical aggression) outcomes. In particular, further examination of differences in the influences on physical and relational aggression for males and females, and possible interactions among family, school, and peer group environments, could add to our understanding of the etiology of physical and relational aggression in the school context, and could potentially provide insights that would allow further refinement of school-based interventions to reduce aggression and bullying. It would also be useful to examine in more detail whether the impact of BPYS itself varies with gender and with family, school, and peer group environment, above and beyond the use of these variables as controls in the present study.

Examination of substance use and abuse in the present evaluation was limited to whether BPYS might have any collateral effect on substance use and abuse. The data collected for this evaluation could also be used, however, to examine in more detail the relationships of substance use and abuse with family background, school climate, and peer group climate, replicating prior research in these areas; and also to examine the relationship of substance use and abuse to physical aggression, violence, and bullying in the school context, a subject that has not been as extensively addressed in presently existing research.

Except for the first recommendation above, all of these objectives could be pursued using data already collected for the present study.

REFERENCES

- Agresti, A. and Finlay, B. (1997). *Statistical Methods for the Social Sciences*, third edition. Upper Saddle River, NJ: Prentice Hall.
- Anderman, C., Cheadle, A., Curry, S., Diehr, P., Shultz, L., and Wagner, E. (1995). Selection bias related to parental consent in school-based survey research. *Evaluation Review* 19:663-674.
- Batsche, G.M. and Knoff, H.M. (1994). Bullies and their victims: understanding a pervasive problem in the schools. *School Psychology Review* 23:165-174.
- Battistich, Victor, Schaps, Eric, Watson, Marilyn, Solomon, Daniel, and Lewis, Catherine (2000). Effects of the Child Development Project on students' drug use and other problem behaviors. *Journal of Primary Prevention*, 21, 75-99.
- Berton, M.W. and Stabb, S.D. (1996). Exposure to violence and post-traumatic stress disorder in urban adolescents. *Adolescence* 31:489-498.
- Bifulco, R. (2002). Addressing self-selection bias in quasi-experimental evaluations of whole-school reform. *Evaluation Review* 26:545-572.
- Blumberg, M. (1979). Injury to victims of personal crimes: nature and extent. In W.H. Parsonage (editor), *Perspectives on Victimology*. Beverly Hills, CA: Sage, pp. 133-147.
- Bonds, M. and Stoker, S. (2000). *Bully-Proofing Your School: A Comprehensive Approach for Middle Schools*. Longmont, CO: Sopris West.
- Boney-McCoy, S. and Finkelhor, D. (1995). Psychosocial sequelae of violent victimization in a national youth sample. *Journal of Consulting and Clinical Psychology* 63:726-736.
- Botvin, Gilbert J., Baker, Eli, Dusenbury, Linda, Botvin, Elizabeth, and Diaz, Tracy (1995). Long-term follow-up results of a randomized drug abuse prevention trial in a white middle-class population. *Journal of the American Medical Association*, 273, 1106-1112.
- Boyd, L.H. and Iversen, G.R. (1979). *Contextual Analysis: Concepts and Statistical Techniques*. Belmont, CA: Wadsworth.
- Brown, S.R. and Melamed, L.E. (1990). *Experimental Design and Analysis*. Newbury Park, CA: Sage.
- Bureau of Justice Statistics. (1994). *Criminal Victimization in the United States, 1992*. Washington, DC: U.S. Department of Justice.

- Campbell, D.T. and Stanley, J.C. (1963). *Experimental and Quasi-Experimental Designs for Research*. Chicago: Rand McNally.
- Cook, T.D. and Campbell, D.T. (1979). *Quasi-Experimentation: Design and Analysis Issues for Field Settings*. Chicago: Rand McNally.
- Costner, H. L. (1965), Criteria for measures of association. *American Sociological Review* 30, 341-353.
- Crick, N.R. and Grotpeter, J.K. (1995). Relational aggression, gender, and social-psychological adjustment. *Child Development* 66:710-722.
- Crick, N.R. and Grotpeter, J.K. (1996). Children's maltreatment by peers: victims of relational aggression. *Development and Psychopathology* 8:367-380.
- DeVoe, K. F., Peter, K., Kaufman, P., Ruddy, S. A., Miller, A. K., Planty, M., Snyder, T. D., and Rand, M. R. (2003). *Indicators of School Crime and Safety: 2003*. Washington, DC: U.S. Department of Education and U.S. Department of Justice.
- DeVoe, K. F., Peter, K., Kaufman, P., Miller, A., Noonan, T. D., Snyder, T. D., and Baum, K. (2004). *Indicators of School Crime and Safety: 2004*. Washington, DC: U.S. Department of Education and U.S. Department of Justice.
- Eaton, D. K., Lowry, R., Brener, N. D., Grunbaum, J. A., and Kann, L. (2004). Passive versus active parental permission in school-based survey research: does the type of permission affect prevalence estimates of risk behaviors? *Evaluation Review* 28:564-577.
- Ellickson, P.L and Hawes, J.A. (1989). An assessment of active versus passive methods of obtaining parental consent. *Evaluation Review* 13:45-55.
- Elliott, D.S. (1994). Serious violent offenders: onset, developmental course, and termination. *Criminology* 32:701-722.
- Elliott, D.S., Ageton, S.S., and Canter, R.J. (1979). An integrated perspective on delinquent behavior. *Journal of Research on Crime and Delinquency* 16:3-27.
- Elliott, D.S., Huizinga, D., and Ageton, S.S. (1985). *Explaining Delinquency and Drug Use*. Newbury Park, CA: Sage.
- Elliott, D.S., Huizinga, D., and Menard, S. (1989). *Multiple Problem Youth: Delinquency, Substance Use, and Mental Health Problems*. New York: Springer-Verlag.
- Elliott, D.S. and Menard, S. (1996). Delinquent friends and delinquent behavior: temporal and developmental patterns. In J.D. Hawkins (ed.), *Delinquency and Crime: Current Theories*. Cambridge, UK: Cambridge University press, pp. 28-67.

- Esbensen, F., Deschenes, E.P., Vogel, R.E., West, J., Arboit, K. and Harris, L. (1996). Active parental consent in school-based research: an examination of ethical and methodological issues. *Evaluation Review* 20:737-753.
- Esbensen, F., Miller, M.H., Taylor, T.J., He, Ni, and Freng, A. (1997). Differential attrition rates and active parental consent. Paper presented at the 1997 Annual Meeting of the American Society of Criminology.
- Farrington, D.P. (1993). Have any individual, family, or neighbourhood influences on offending been demonstrated conclusively? In D.P. Farrington, R.J. Sampson, and P.O. Wikstrom (eds.) *Integrating Individual and Ecological Aspects of Crime*. Stockholm: National Council for Crime Prevention, pp. 7-37.
- Farrington, D.P. (1995). The development of offending and antisocial behaviour from childhood: key findings from the Cambridge Study in Delinquent Development. *Journal of Child Psychology and Psychiatry* 36:929-964.
- Fried, S. and Fried, P. (1996). *Bullies and Victims: Helping Your Child Survive the Schoolyard Battlefield*. New York: Evans.
- Garofalo, J., Siegel, L., and Laub, J. (1987). School-related victimization among adolescents: an analysis of National Crime Survey (NCS) narratives. *Journal of Quantitative Criminology* 3:321-338.
- Garrity, C., Baris, M., and Porter, W. (2000a). *Bully-Proofing Your Child: A Parent's Guide*. Longmont, CO: Sopris West.
- Garrity, C., Baris, M., and Porter, W. (2000b). *Bully-Proofing Your Child: First Aid for Hurt Feelings*. Longmont, CO: Sopris West.
- Garrity, C., Jens, K., Porter, W., Sager, N., and Short-Camilli, C. (2000) *Bully-Proofing Your School: A Comprehensive Approach for Elementary Schools*. Second edition. Longmont, CO: Sopris West.
- Grottpeter, J.K. and Crick, N.R. (1996). Relational aggression, overt aggression, and friendship. *Child Development* 62:2328-2338.
- Hardin, J. and Hilbe, J. (2001). *Generalized Linear Models and Extensions*. College Station, TX: Stata Press.
- Hardin, J.W. and Hilbe, J.M. (2003). *Generalized Estimating Equations*. Boca Raton, FL: Chapman and Hall/CRC.

- Henry, K.L., Smith, E.A., and Hopkins, A.M. (2002). The effect of active parental consent on the ability to generalize the results of an alcohol, tobacco, and other drug prevention trial to rural adolescents. *Evaluation Review* 26:645-655.
- Hoover, J.H., Oliver, R.L., and Hazler, R.J. (1992). Bullying: perceptions of adolescent victims in the Midwestern USA. *School Psychology International* 13:5-16.
- Hosmer, D.S. and Lemeshow, S. (2000). *Applied Logistic Regression*, second edition. New York: Wiley.
- Ireland, J.L. (2002). Official records of bullying incidents among young offenders: what can they tell us and how useful are they? *Journal of Adolescence* 25:669-679.
- Iversen, G.R. (1991). *Contextual Analysis*. Newbury Park, CA: Sage.
- Jackson, S. and Brashers, D.E. (1994). *Random Factors in ANOVA*. Thousand Oaks, CA: Sage
- Kaufman, P., Chen, X., Choy, S.P., Ruddy, S.A., Miller, A.K., Fleury, J.K., Chandler, C.A., Rand, M.R., Klaus, P., and Planty, M.G. (2000). *Indicators of School Crime and Safety, 2000*. Washington, DC: U.S. Department of Education and U.S. Department of Justice.
- Kearney, K.A., Hopkins, R.H., Mauss, A.L. and Weisheit, R.A. (1983). Sample bias resulting from a requirement for written parental consent. *Public Opinion Quarterly* 47:96-102.
- Kilpatrick, D.G., Saunders, B.E., Veronen, L.J., Best, C.L., and Von, J.M. (1987). Criminal victimization: lifetime prevalence, reporting to the police, and psychological impact. *Crime and Delinquency* 33:479-489.
- Klaus, P.A. (1994). The costs of crime to victims. *Crime Data Brief*. Washington, DC: U.S. Department of Justice.
- Laub, J.H. (1997). Patterns of criminal victimization in the United States. In R.C. Davis, A.J. Lurigio, and W.G. Skogan (editors), *Victims of Crime*, second edition. Thousand Oaks, CA: Sage, pp. 9-26.
- Lawrence, R. (2007). *School Crime and Juvenile Justice*. New York: Oxford University Press.
- Levy, P. and Lemeshow, S. (1999). *Sampling of Populations*, third edition. New York: Wiley.
- Lipsey, Mark (1999). Can intervention rehabilitate serious delinquents? *Annals, AAPSS*, 564, 142-166.
- Lipton, D., Martinson, R., and Wilks, J. *The Effectiveness of Correctional Treatment*. New York: Praeger.

- Little, T.D., Schnabel, K.U., and Baumert, J., eds. (2000). *Modeling Longitudinal and Multilevel Data: Practical Issues, Applied Approaches, and Specific Examples*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Loether, H. J. and McTavish, D. G. (1993). *Descriptive and Inferential Statistics: An Introduction*, Boston: Allyn and Bacon.
- Lurigio, A.J. (1987). Are all victims alike? The adverse, generalized, and differential impact of crime. *Crime and Delinquency* 33:452-467.
- Martinson, R. What works? Questions and answers about prison reform. *The Public Interest* 35:22-54.
- Menard, S. (1987). Short-term trends in crime and delinquency: a comparison of UCR, NCS, and self-report data. *Justice Quarterly* 4:455-474.
- Menard, S. (2000). The “normality” of repeat victimization from adolescence through early adulthood. *Justice Quarterly* 17:543-574.
- Menard, S. (2002a). Short and long term consequences of victimization. *Office of Juvenile Justice and Delinquency Prevention/Centers for Disease Control and Prevention Youth Violence Research Bulletin*. Washington, DC: Office of Juvenile Justice and Delinquency Prevention.
- Menard, S. (2002b). *Applied Logistic Regression Analysis*, second edition. Thousand Oaks, Ca: Sage.
- Menard, S. (2002c). *Longitudinal Research*, second edition. Thousand Oaks, CA: Sage.
- Menard, S. and Covey, H.C. (1988). UCR and NCS: comparisons over space and time. *Journal of Criminal Justice* 16:371-384.
- Menard, S. and Elliott, D.S. (1994). Delinquent bonding, moral beliefs, and illegal behavior: a three-wave panel model. *Justice quarterly* 11:173-188.
- Menard, S. and Huizinga, D. (1994). Changes in conventional attitudes and delinquent behavior in adolescence. *Youth and Society* 26:23-53.
- Mihalic, S. (2001). The importance of implementation fidelity. *Blueprints News*, 2, 1, 1-2. Office of Juvenile Justice and Delinquency Prevention.
- Mihalic, S., Irwin, K., Elliott, D., Fagan, A., and Hansen, D. (2001). Blueprints for Violence Prevention. *Juvenile Justice Bulletin*. Washington, DC: Office of Juvenile Justice and Delinquency Prevention.

- Miller, T.R., Cohen, M.A., and Wiersma, B. (1996). *Victim Costs and Consequences: A New Look*. Washington, DC: U.S. Department of Justice.
- Montgomery, I.M., Torbet, P.F., Malloy, D.A., Adamcik, L.P., Toner, M.J., and Andrews, J. (1994). *What Works: Promising Interventions in Juvenile Justice*. Pittsburgh, PA: National Center for Juvenile Justice.
- National Institute on Drug Abuse (1997). *Preventing Drug Use Among Children and Adolescents: A Research-Based Guide*. Washington, DC: National Clearinghouse for Alcohol and Drug Information.
- Natvig, G.K., Albrektsen, G., and Quarnstrøm, U. (2001). School-related stress experience as a risk factor for bullying behavior. *Journal of Youth and Adolescence* 30:561-575.
- Norris, F.H., Kaniasty, K., and Thompson, M.P. (1997). The psychological consequences of crime: findings from a longitudinal population-based study. In R.C. Davis, A.J. Lurigio, and W.G. Skogan (editors), *Victims of Crime*, second edition. Thousand Oaks, CA: Sage, pp. 146-166.
- O'Brien, R.M. (1985). *Crime and Victimization Data*. Beverly Hills, CA: Sage.
- O'Moore, M. and Kirkham, C. (2001). Self-esteem and its relationship to bullying behaviour. *Aggressive Behavior* 27:269-283.
- Olweus, D. (1992). Victimization by peers: antecedents and long-term outcomes. In K.H. Rubin and J.B. Asendorf (eds.), *Social Withdrawal, Inhibition, and Shyness in Childhood*. Hillsdale, NJ: Lawrence Erlbaum.
- Olweus, D. (1993). *Bullying at School: What We Know and What We Can Do*. Cambridge, MA: Blackwell Publishers, Inc.
- Olweus, D., Limber, S., & Mihalic, S. (1999). *Blueprints for Violence Prevention, Book Nine: Bullying Prevention Program*. Boulder, CO: Center for the Study and Prevention of Violence.
- Raudenbush, S.W. and Bryk, A.S. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods*. Second edition. Thousand Oaks, CA: Sage.
- Reichardt, C.S. and Mark, M.M. (1998). Quasi-experimentation. In L. Bickman and D.J. Rog (eds.), *Handbook of Applied Social Research Methods*. Thousand Oaks, CA: Sage, pp. 193-228.
- Resick, P.A. and Nishith, P. (1997). Sexual assault. In R.C. Davis, A.J. Lurigio, and W.G. Skogan (editors), *Victims of Crime*, second edition. Thousand Oaks, CA: Sage, pp. 27-52.

- Riecken, H.W., Boruch, R.F. (eds.), Campbell, D.T., Caplan, N., Glennan, T.K., Jr., Pratt, J.W., Rees, A., and Williams, W. (1974). *Social Experimentation: A Method for Planning and Evaluating Social Intervention*. New York: Academic Press.
- Rigby, K. (1998). Health effects of school bullying. *The Professional Reading Guide for Educational Administrators* 19:35-38.
- Roitberg, T. and Menard, S. (1995). Adolescent violence: a test of integrated theory. *Studies in Crime and Crime Prevention* 4:177-196.
- Rossi, P.H., Freeman, H.E., and Lipsey, M.W. (1999). *Evaluation: A Systematic Approach*. Sixth edition. Thousand Oaks, CA: Sage.
- Ruback, R.B. and Thompson, M.P. (2001). *Social and Psychological Consequences of Violent Victimization*. Thousand Oaks, CA: Sage.
- Severson, H.H. and Biglan, A. (1989). Rationale for the use of passive consent in smoking prevention research: politics, policy, and pragmatics. *Preventive Medicine* 18:267-279.
- Sherman, L.W., Gottfredson, D., MacKenzie, D., Eck, J., Reuter, P., and Bushaway, S. (1997). *Preventing Crime: What Works, What Doesn't, What's Promising*. Washington, DC: Office of Justice Programs.
- Simon, T., Mercy, J., and Perkins, C. (2001). Injuries from Violent Crime, 1992-98. *Bureau of Justice Statistics/Centers for Disease Control and Prevention Special Report*. Washington, DC: U.S. Department of Justice.
- Simes, R.J. (1986). An improved Bonferroni procedure for multiple tests of significance. *Biometrika* 73:751-754.
- Stevens, V., Van Oost, P, and De Bourdeaudhuij, I. (2001). Implementation process of the Flemish antibullying intervention and relation with program effectiveness. *Journal of School Psychology* 39:303-317.
- Stoolmiller, M. (1995). Using latent growth curve models to study developmental processes. In J.M. Gottman (ed.), *The Analysis of Change*. Mahwah, NJ: Lawrence Erlbaum, pp. 103-138.
- Tanner, L. (2002). AMA adopts anti-bully measure. Associated Press, June 19.
- Thompson, S.K. (2002). *Sampling*, second edition. New York: Wiley.
- Unger, J.B., Gallaher, P., Palmer, P.H., Baezconde-Garbanati, L., Trinidad, D.R., Cen, S., and Johnson, C.A. (2004). No news is bad news: characteristics of adolescents who provide

neither parental consent nor refusal for participation in survey research. *Evaluation Review* 28:52-63.

U.S. Department of Health and Human Services. (2001). *Youth Violence: A Report of the Surgeon General*. Rockville, MD: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Injury Prevention and Control; Substance Abuse and Mental Health Services Administration, Center for Mental Health Services; and National Institutes of Health, National Institute of Mental Health.

Walker, J. T. and Madden, S. (2005). *Statistics in Criminology and Criminal Justice: Analysis and Interpretation*, second edition. Boston: Jones and Bartlett.

Waller, I., Gauthier, L., Hicks, D., Sansfacon, D., and Salel, L. (1999). *100 Promising Crime Prevention Programs Across the World*. Montreal: International Centre for the Prevention of Crime.

Wright, S. R. (1979). *Quantitative Methods and Statistics: A Guide to Social Research*. Beverly Hills, CA: Sage.

Zeller, R.A. and Carmines, E.G. (1980). *Measurement in the Social Sciences: The Link Between Theory and Data*. Cambridge, UK: Cambridge University Press.