



# Evaluation of the Olweus Bully Prevention Program in an Urban School System in the USA

Albert D. Farrell<sup>1</sup> · Terri N. Sullivan<sup>1</sup> · Kevin S. Sutherland<sup>2</sup> · Rosalie Corona<sup>1</sup> · Saba Masho<sup>3</sup>

Published online: 14 June 2018  
© Society for Prevention Research 2018

## Abstract

This study evaluated the Olweus Bully Prevention Program (OBPP) in urban middle schools serving a mostly African American student population. Participants were 1791 students from three communities with high rates of crime and poverty. We evaluated the impact of the OBPP using a multiple-baseline experimental design in which we randomized the order and timing of intervention activities across three schools. We assessed the frequency of violence and victimization using self-report and teachers' ratings of students collected every 3 months over 5 years. Initiation of the OBPP was associated with reductions in teachers' ratings of students' frequency of aggression, with effects emerging in different years of implementation for different forms of aggression. Whereas reductions in teachers' ratings of students' verbal and relational aggression and victimization were evident during the second implementation year, reductions in physical aggression did not appear until the third year. Effects were consistent across gender and schools, with variability across grades for relational and verbal aggression and victimization. In contrast, there were no intervention effects on students' reports of their behavior. Positive outcomes for teachers', but not students' ratings, suggest the intervention's effects may have been limited to the school context. Variation in when effects emerged across outcomes suggests that changes in physical aggression may require more sustained intervention efforts. The intervention was also associated with increases in teachers' concerns about school safety problems, which may indicate that teachers were more attuned to recognizing problem behaviors following exposure to the OBPP.

**Keywords** Olweus Bully Prevention Program · Violence prevention · School intervention · Evaluation · Aggression

Although there has been significant progress toward developing effective interventions to reduce youth violence (Matjasko et al. 2012), much work remains to be done. Over the past several decades, researchers have moved from interventions that focus on individual-level factors toward more comprehensive approaches that also address factors within the broader environment (Matjasko et al. 2016). There has also been increasing recognition of the need for interventions that meet the needs of youth at

varying levels of risk. These include universal interventions that focus on all youth, selective interventions focusing on those at higher risk, and indicated interventions designed for youth already displaying aggression (Farrell and Vulin-Reynolds 2007). These trends are evident in violence prevention efforts developed for implementation in school settings. Schools are an obvious setting for such efforts because they can support the large-scale dissemination of programs and provide a natural setting for the development of socioemotional and behavioral competencies (Sullivan et al. 2016). They also provide an opportunity to create a positive environment that can discourage aggressive behavior and provide support for prosocial interactions among students. Although a variety of comprehensive school-based prevention programs have been developed, further work is needed to evaluate their impact, to determine the extent to which their effects can generalize across different populations, and to evaluate the time they require to produce effects.

The Olweus Bullying Prevention Program (OBPP; Olweus and Limber 2010) is an example of a school-based intervention designed to provide a comprehensive approach to

---

✉ Albert D. Farrell  
afarrell@vcu.edu

<sup>1</sup> Department of Psychology, Virginia Commonwealth University, VCU Box 842018, Richmond, VA 23284-2018, USA

<sup>2</sup> Department of Counseling and Special Education, Virginia Commonwealth University, Box 842020, Richmond, VA 23284-2020, USA

<sup>3</sup> Department of Family Medicine and Population Health, School of Medicine, Virginia Commonwealth University, Richmond, VA 23298-0212, USA

reducing aggression. Its goal is to promote a positive and responsive school climate characterized by low rates of aggression and high rates of positive, prosocial relations among and between students and teachers. School level components include a Bullying Prevention Coordinating Committee that conducts activities such as monitoring a school safety plan, identifying “hot spots” where aggressive behaviors are more likely to occur, identifying positive and negative consequences for behavior, training all staff, and coordinating school events related to the program (Limber 2011). A classroom component focuses on promoting student adherence to anti-bullying rules and includes classroom meetings to discuss bullying and related topics (Limber 2011). An individual component includes an “on-the-spot” intervention conducted by school staff who witness bullying behaviors and follow-up interventions with students involved in these incidents (Limber 2011).

The primary goal of the OBPP is to reduce bullying, defined as aggressive behavior carried out repeatedly and over time in a relationship in which there is an actual or perceived imbalance of power or strength (Olweus and Limber 2010). Although it focuses on bullying, it can also be considered a violence prevention program in that many acts of aggression, particularly those involving adolescents in school settings, fit within the definition of bullying. These include acts that result in physical harm, and other forms of aggression such as relational aggression that inflict harm by damaging social relations. Moreover, components of OBPP such as class meetings address risk and protective factors for aggression (e.g., managing emotions, developing positive relationships, and respect for others) and adult intervention address responses to aggressive behaviors and victimization. Given its focus, it is not surprising that the OBPP is included in reviews of more general violence prevention programs (e.g., <http://www.blueprintsprograms.com/>).

The OBPP was developed and first implemented in Norway in the mid-1980s (Limber 2011). Evaluations of the OBPP in Norway, particularly with students in grades four through seven, have found decreases in bullying perpetration and victimization (Olweus and Limber 2010). Other international evaluations have found decreases in self-reported victimization but mixed findings for self-reported bullying perpetration including both decreases (O’Moore and Minton 2005) and increases (Pepler et al. 1994). Although researchers have evaluated the OBPP in other countries, less is known about how well their findings can be replicated for schools in the USA. Limber (2011) noted challenges for schools in the USA that attempt to implement the OBPP. These include time constraints on teachers, differences in school structure, and challenges engaging the broader community. Studies evaluating the impact of the OBPP on bullying and victimization in elementary, middle, and high schools in the USA have reported mixed results. Schroeder et al. (2012) found reductions in student self-reports of bullying in a quasi-experimental

evaluation of the OBPP in 107 schools covering grades K through 12 from 49 Pennsylvania counties. Melton et al. (1998) found reductions in self-reports of bullying in 11 schools covering grades 4 through 8 in South Carolina during the first year of implementation, but did not find similar effects in the second year or when the OBPP was subsequently implemented in seven control schools. Black and Jackson (2007) found reductions in observations of bullying incidents across 4 years of implementing the OBPP in six elementary and middle schools, but results varied across schools and there were no effects on student reports of bullying. Effects have also differed by grade and gender. Bowllan (2011) found greater reductions in being excluded and bullied for seventh grade girls, but eighth grade girls reported increases in being bullied and bullying others, and there were no effects for boys.

There are several limitations of studies that have evaluated the impact of the OBPP in the USA. The OBPP requires implementation at the school level, which requires designs that focus on schools as the unit of analysis. This has led to a variety of quasi-experimental designs. For example, Bauer et al. (2007) compared outcomes in seven middle schools that implemented the OBPP in response to a state mandate to implement anti-bullying policies to three schools that chose other approaches. Black and Jackson (2007) evaluated the OBPP across 4 years of implementation in six intervention schools without including control schools. Others have used an age-cohort design (Olweus 2005) that compares outcomes for students attending a school prior to implementing the OBPP to outcomes for students in the same schools following implementation. Schroeder et al. (2012) used this design in their analysis of the OBPP in 107 schools, as did Bowllan (2011) in a study evaluating the impact of the OBPP in a single school. Most studies evaluating the OBPP in US schools have relied on the Olweus Bullying Questionnaire (Olweus 2005) to assess outcomes. This assesses the number of students classified as bullying or being bullied, but does not reflect the frequency of these behaviors, nor does it differentiate among forms of bullying or victimization (e.g., verbal versus physical aggression). Moreover, measures that require adolescents to identify themselves as bullying or being bullied may produce unreliable self-reports. This may particularly be an issue for Black and Hispanic youth and males who may be more willing to report experiencing specific bullying behaviors, but are less likely to report being bullied (Lai and Kao 2018). This has led researchers to recommend assessing the frequency of behaviors without labeling them as bullying (e.g., Bosworth et al. 1999). Finally, although there is evidence to suggest that interventions may take time to produce effects (e.g., Schroeder et al. 2012), few studies have evaluated effects over more than 1 or 2 years (e.g., see Black and Jackson 2007).

There is also a need to determine the extent to which the content of the OBPP is relevant for other populations. Bauer et al. (2007) noted that the OBPP was developed for a

homogeneous population in Norway and its translation to more diverse samples may be complicated. Several studies have focused on urban populations with diverse samples that include African American students (Bauer et al. 2007; Bowllan 2011), and one study included a sample that was 82% African American (Black and Jackson 2007). However, these studies did not report the extent to which the content of the intervention, in particular the topics of class meetings, was adapted to increase its relevance to these groups. Further work is also needed to examine its impact on early adolescents as this developmental timeframe is characterized by growth in areas of social cognition, emotional competencies, and identity development.

The purpose of this study was to evaluate the effectiveness of the OBPP in an urban school system serving a predominantly African American student population. It was designed to address several limitations of prior evaluations of the OBPP. We used a multiple-baseline experimental design that randomized the order and timing of implementing the intervention and evaluated outcomes based on student and teacher report of aggression, victimization, and overall school climate collected every 3 months over a 5-year period. Biglan et al. (2000) noted the benefits of multiple-baseline designs and described how they more effectively address threats to internal validity than quasi-experimental designs. We evaluated the impact of the OBPP across 19 waves of data collected over several years of implementation and investigated the consistency of effects across schools. We adapted the class meetings to incorporate topics relevant to our student population based on feedback from students, teachers, and administrators. We also assessed the fidelity of teachers' implementation of the class meetings. Finally, we collected data on teacher-reported school norms and safety. These outcomes are consistent with the expected impact of the OBPP, but have been included in few evaluation studies. Studies evaluating OBPP in Norway have found positive changes on measures of school climate including youth-report perceptions of school and prosocial relationships (e.g., Olweus and Alsaker 1991). In contrast, Bauer et al. (2007) assessed student perceptions of school safety, school engagement, and support from students and teachers, but did not find intervention effects for these variables.

Based on prior studies, we hypothesized that school-level implementation of the OBPP would decrease both student and teachers' ratings of students' frequency of aggression and victimization, and overall concerns related to school safety. We further hypothesized that there would be decreases in school norms supporting aggression, and increases in school norms supporting nonviolence. We expected effects to be consistent across schools. We also examined the consistency of effects across gender and grade, but did not believe there was a sufficient theoretical justification for formulating specific hypotheses about gender or grade differences.

## Methods

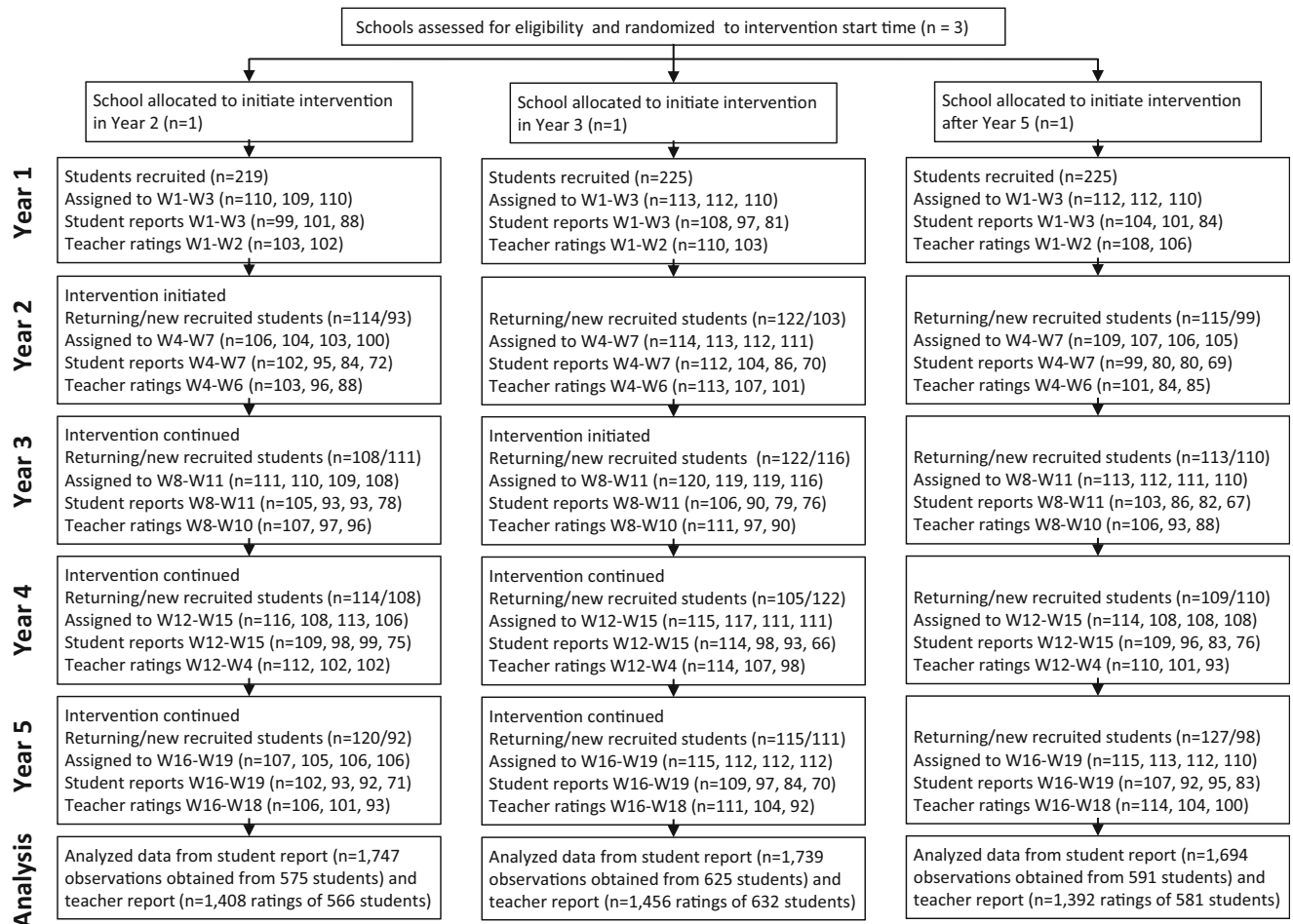
### Setting and Design

We evaluated the school-level effects of the OBPP using a multiple-baseline design that treated schools as the unit of analysis. We conducted this study in three urban public middle schools in the southeastern United States that served a predominantly African American student population from low-income families, most of whom (i.e., 74 to 85%) were eligible for the federal free lunch program. Enrollments at the start of the school year across the 5 years of the project ranged from 401 to 493 for school A, from 519 to 575 for school B, and from 419 to 610 for school C. We assessed outcomes every 3 months starting in February of 2011 and continuing through the summer of 2015. This provided 19 waves of student-report data and 14 waves for measures collected only during the school years (i.e., teachers' ratings and ratings of school climate). Based on randomization of the order and timing of initiating intervention activities in each school, we began the intervention at school A during the 2011–2012 school year, at school B during the 2012–2013 school year, and provided support for school C to implement the intervention after all data were collected (see Fig. 1). In each school, once we started the intervention, it was continued through the end of the project.

### Participants

We collected data from a random sample of students from each grade every 3 months to assess school-level effects. We also obtained teachers' ratings of participating students every 3 months during every school year. Because the design focused on school-level changes, it did not require four waves of data from every student every year, nor was this required for determining trajectories of school-level outcomes based on multilevel models. We therefore used a planned missingness design in which we randomized each participating student to complete two of the four waves each year to reduce participant fatigue and testing effects. The random assignment of students to waves prevents this pattern of missing data from biasing the analyses. During the project's first year, we randomly sampled about equal numbers of students from each grade from the rosters of each school for an initial sample of 669 students. Each subsequent year, we recruited a random sample of entering sixth graders and of new seventh and eighth graders to replace those who left the school. We obtained active parental permission and student assent from approximately 80% of those eligible each year. Across the 5 years of the project, we collected 19 waves of data on student ratings. A total of 1791 different students provided ratings, with 212 to 334 students participating at each wave ( $M = 275$ ; see Fig. 1).

The mean ages of participants on the date of the fall wave of data collection were 11.3 ( $SD = 0.56$ ), 12.3 ( $SD = 0.57$ ),



**Fig. 1** CONSORT figure showing random assignment of schools to the timing of starting the intervention, number of students assigned to participate at each wave (W1–W19), and number of students for whom student and teacher reports were actually obtained. Students were randomly assigned to participate at two of the four waves during each

year they attended a participating school. Teacher ratings were not collected at summer waves. Additional students were recruited each year to replace those who were promoted to the ninth grade or who left the school for other reasons

and 13.4 (SD = 0.58) years, for sixth, seventh, and eighth graders, respectively. The sample was 53% female. Fifteen percent identified their ethnicity as Hispanic or Latino. The majority (81%) identified as Black or African American as the sole category (73%) or as one of several categories (8%). Nine percent, most of whom (93%) had identified themselves as Hispanic or Latino, did not endorse any racial category. Of the remainder, 5% identified themselves as White, with 1% or less in each remaining category. Data at the first wave for each participant indicated that 41% lived with a single mother, 26% with both biological parents, 21% with a parent and stepparent, 6% with a relative without a parent, and 4% with their father without a mother or stepmother.

## Intervention

We hired a full-time youth development specialist to work with school staff on the implementation of the OBPP. Individual-level components included supervision of students

(e.g., monitoring of hot spots) by school staff, who conducted “on-the-spot” interventions when they witnessed bullying incidents, met with students involved in these incidents and with their parents and guardians as needed, and developed individual plans for the students.

Classroom-level components included helping students adhere to school-wide rules against bullying and conducting weekly class meetings. Core academic or elective teachers conducted 21 to 23 weekly meetings with students in each classroom and grade during each school year. Class meetings based on the OBPP addressed anti-bullying rules, the bullying circle, and other topics (e.g., cyberbullying, relational aggression, and respect for oneself and others). Other meeting topics (e.g., leadership and stress management) were added based on input from students gathered as part of a class meeting, from teachers who provided written feedback on the lessons, and from administrators who provided feedback after reviewing the lessons.

We collected observer ratings on the fidelity of implementation for 396 class meetings conducted during the 4 years of implementation at school A and for 212 conducted during the 3 years of implementation at school B. These represented a random sample of 20% of the teachers each week. Observers rated teachers' use (adherence) and quality of delivery (competence) of ten instructional behaviors (e.g., "Teacher provided feedback to students"). They also rated teachers' adherence to and quality of delivery of eight items representing components of class meetings (e.g., "Teacher explained the purpose of class meeting"). All items were rated on a 3-point scale. Anchor points were 1 = *Not at all*, 2 = *Somewhat*, and 3 = *Extensively* for adherence, and 1 = *Poor*, 2 = *Acceptable*, and 3 = *Excellent* for competence. Cronbach's alphas were .83 and .79 for instructional and procedural adherence, respectively and .89 and .90 for instructional and procedural competence, respectively.

Ratings of instructional and procedural adherence averaged 2.35 and 2.13, respectively in school A, and 2.37 and 2.06, respectively in school B. Ratings on the measures of instructional and procedural competence averaged 2.22 and 2.16, respectively in school A, and 2.30 and 2.17, respectively in school B. These mean ratings exceeded 2.00, which corresponds to anchor points of *somewhat* and *acceptable* for adherence and competence, respectively. There were no significant differences in fidelity ratings across schools with one exception. Ratings of procedural adherence were slightly higher at school A than those at school B (partial  $\eta^2 = .014$ ,  $p < .008$ ). At school A, there were small differences across school years for ratings of instructional adherence (partial  $\eta^2 = .031$ ,  $p = .006$ ), procedural adherence (partial  $\eta^2 = .036$ ,  $p < .004$ ), and procedural competence (partial  $\eta^2 = .026$ ,  $p < .015$ ), but not for instructional competence (partial  $\eta^2 = .010$ ,  $p = .262$ ). Tests of linear change revealed a trend toward increasing levels for instructional adherence ( $d = .35$  comparing the first and last year,  $p = .025$ ), but no consistent trend for the other measures. At school B, there were small to moderate variations across intervention years for instructional adherence (partial  $\eta^2 = .109$ ,  $p < .001$ ), procedural adherence (partial  $\eta^2 = .085$ ,  $p < .001$ ), instructional competence (partial  $\eta^2 = .098$ ,  $p < .001$ ), and procedural competence (partial  $\eta^2 = .083$ ,  $p < .001$ ). These were captured by significant linear trends for decreases in fidelity ratings across implementation years for procedural adherence, instructional competence, and procedural competence ( $d$ -coefficients comparing the first and last waves were  $-.43$ ,  $-.52$ , and  $-.57$ , respectively), but no consistent trend for instructional adherence.

School-level components included the formation of a Bullying Prevention Coordinating Committee that conducted initial and refresher trainings for school staff, addressed issues related to the school supervisory systems, and created a school-wide plan to institute negative and positive consequences for student behavior. Each school's committee held

8 to 12 meetings each year. Each fall, they planned and held student and parent kick-off events to introduce and review school rules against bullying with increased parent and family attendance each year.

We incorporated community-level components by presenting the OBPP to parents and community members at "Back to School" nights, organizing events such as a career day, and a family night targeting sixth grade families to provide information about the OBPP. We also incorporated an after-school youth leadership program (Fuentes et al. 2015). Participants were nominated by teachers who identified students who displayed leadership skills, but were not considered active school leaders. This program was consistent with recommendations to extend OBPP principles and anti-bullying messages into the community (Limber 2011). Student projects focused on school and community service that were consistent with the OBPP principles. We also included some lessons from the Youth Empowerment Solutions curriculum (Zimmerman et al. 2011) to foster cooperative learning, discussion, and conversations about social issues related to positive youth development and bullying prevention. Each group met 8 to 12 times, with an average of 9 students attending each meeting.

We supplemented the OBPP with family interventions. The original plan was to implement a self-directed version of *Staying Connected to Your Teen* (Haggerty et al. 2007) with families of youth with multiple incidents of truancy. We encountered numerous challenges engaging families in these communities in a self-directed program. We addressed this by using alternative formats ranging from providing telephone support to implementing the intervention in multiple family groups at the schools and in community settings. We also offered a Spanish version of *Parenting Wisely* (Gordon 2000) as an alternative for Spanish-speaking families. Despite extensive efforts, we received limited referrals for youth with truancy issues. We therefore expanded our criteria to include youth with disciplinary referrals, and ultimately to include youth self-referred by parents. At school A, 92 families participated in the family intervention, with 48 (52%) completing all lessons. At school B, 53 families participated in the family intervention, with 15 (28%) completing all lessons. Spanish-speaking families represented 37% of those who completed all the lessons.

## Measures of Aggression and Victimization

We obtained ratings of each participating student on the Problem Behavior Frequency Scale—Teacher Report Form (PBFS-TR; Farrell et al. 2018) from 155 core academic teachers who completed an average of 27 ratings ( $SD = 25.6$ ). The majority of teachers were women (69%); 61% were non-Hispanic Black, 28% non-Hispanic White, and 5% Hispanic or Latino. Teachers rated how frequently the student engaged in or experienced specific acts of physical (e.g., "Hit

or slapped someone”), verbal (e.g., “Put someone down to their face”), and relational (e.g., “Left another student out on purpose when it was time to do an activity”) aggression in the past 30 days using a 4-point scale, where 1 = *Never*, 2 = *Sometimes*, 3 = *Often*, and 4 = *Very often*. Farrell et al. (2018) evaluated the PBFS-TR in two samples of urban middle school students and found support for separate factors representing each form of aggression and victimization with strong measurement invariance over time, gender, grade, and participation in prevention programs. They also found support for concurrent validity of the PBFS-TR based on its correlations with student reports of aggression and victimization ( $r = .18$  to  $.24$ ) and with teachers’ ratings on other measures of related constructs ( $r = .41$  to  $.75$ ). Alpha coefficients for aggression and victimization scales in the present study ranged from  $.87$  to  $.94$ .

We assessed students’ self-report of their frequency of aggression and victimization on the adolescent report form of the PBFS (PBFS-AR; Farrell et al. 2016). Students rated how frequently they engaged in or experienced specific acts of aggression or victimization in the past 30 days on a 6-point frequency scale, 1 = *Never*, 2 = *1–2 times*, 3 = *3–5 times*, 4 = *6–9 times*, 5 = *10–19 times*, and 6 = *20 or more times*. In contrast to the teacher report measure, the PBFS-AR is not limited to behavior at school or to acts of aggression directed at or by peers. As such, it provides a broader measure of the impact of the intervention on aggression and victimization. Farrell et al. (2017) found support for distinct factors on the PBFS-AR representing physical aggression (e.g., “Shoved or pushed someone”), relational aggression (e.g., “Spread a false rumor about someone”), and overall victimization (including physical [e.g., “Someone hit you hard enough to hurt”], verbal [e.g., “Someone put you down to your face”], and relational [e.g., “Someone said they wouldn’t like you unless you did what he or she wanted”] victimization), with strong measurement invariance across gender and grades. Farrell et al. (2016) found support for the concurrent validity of the PBFS-AR based on its patterns of correlations with teachers’ ratings of students’ adjustment, and self-report measures of beliefs, values, and peer associations. We calculated scores for each scale by collapsing the three highest frequency categories and calculating the mean across items. Alphas in the present study were  $.77$ ,  $.70$ , and  $.90$  for physical aggression, relational aggression, and victimization, respectively.

## Measures of School Safety

Students who completed the student-report measures also completed the student form of the School Safety Problems scale (Henry et al. 2011) at each wave during the school year. This 10-item measure asks students to rate their perceptions of the frequency of threats to their safety at school on a 4-point scale ranging from 1 = *Not a problem* to 4 = *Serious problem*.

Items address students behavior (e.g., fighting, carrying weapons), teacher behavior (e.g., ignoring incidents where students threaten or tease other students, lack of adequate supervision of students), and other concerns (e.g., presence of unsafe areas in the school, gangs, and racial tension or racism). Henry et al. (2011) found support for a single factor representing student concerns about school safety. Alpha in the current study was  $.90$ .

We also obtained ratings of school safety concerns and school norms at the three waves during the school year from a random sample of about 25 teachers from each school representing core and elective from all three grades. We assessed school safety with the High Risk Behaviors scale from the teacher version of the School Safety Problems measure (Henry et al. 2011). This 11-item scale asks teachers to rate the extent to which they believe specific concerns are a problem at their school on a 4-point scale ranging from *Not a problem* to *Serious problem* based on items similar to those on the student form. The alpha in the current study was  $.88$ . We assessed teachers’ perceptions of student norms related to aggression and nonviolence using a teacher version of the School Norms for Aggression and Nonviolent Alternatives Scale (Henry et al. 2011). This measure asks teachers to rate their perceptions of how students in their school would view other students’ responses to specific situations. These include both aggressive (e.g., “Hit someone who said something mean”) and nonviolent (e.g., “Apologized to someone that he or she accidentally bumped into in the hall”) responses. Teachers rate student beliefs using a 3-point scale (1 = *Approve*, 2 = *Disapprove*, 3 = *Neutral*). Responses are averaged across items to create scores representing norms supporting aggression (alpha =  $.85$ ), and norms supporting nonviolence (alpha =  $.78$ ). The two scales are moderately negatively correlated ( $r = -.29$ ).

## Data Analysis

We log transformed scores on the measures of aggression and victimization to reduce their skewness and kurtosis and used linear transformations to provide scores with similar means and standard deviation as the original scores. We used SAS Proc Mixed to model outcomes as a function of an intercept, the project year (dummy-coded with year 1 as the reference), the time of year (modeled as linear within the school year with a dummy-coded variable representing the summer wave where applicable), the school (dummy-coded with school C as the reference), and the intervention year (dummy-coded with first year of implementation as the reference). We specified random intercepts at the student level for the overall intercept and project year and modeled the covariance structure as autoregressive. We used the Kenward-Rogers adjustment for degrees of freedom. We examined individual parameters to compare changes during each year of implementation

relative to baseline for schools receiving the intervention with changes in schools not receiving the intervention during that same period of time. We estimated effect size (i.e., Cohen’s *d*) by dividing the unstandardized parameters by the SD of the outcome measure. We calculated SDs by taking the square root of the sum of the level 1 and level 2 variance estimates. We expanded this model to incorporate school × intervention year interactions to determine if intervention effects differed across the two schools where it was implemented. Finally, we incorporated students’ grade and gender into models to investigate whether intervention effects differed across grades or gender. We examined individual interactions only when there was support based on a significant omnibus *F* test that included all interaction terms.

## Results

### Baseline Differences Across Schools

Analyses comparing teachers’ ratings in schools A and B to school C revealed small baseline differences (i.e., *ds* = .23 to .36 in absolute value) on four of the six measures of student behavior (see Table 1). There were also small, but significant baseline differences across schools on student self-ratings of their frequency of aggression (*ds* = .13 to .19 in absolute value), but no significant differences in their ratings of victimization (see Table 2). Compared with school C, students at school B reported a higher frequency of school safety problems (*d* = −.23), and teachers at the other two schools rated

school safety issues as less of a problem (*ds* = −.38 and −.35). There were no significant baseline differences in teachers’ ratings of school norms.

We examined baseline prevalence rates of aggression and victimization in the past 30 days based on the first wave of school year data collected from each student during year 1. On the PBFS-AR, 68% of students reported committing one or more act of physical aggression and 45% reported committing one or more act of relational aggression. For victimization, 51, 62, and 48% of students reported experiencing one or more act of physical, verbal, and relational victimization, respectively. Teachers identified 53, 67, and 45% of students as having committed one or more acts of physical, verbal, and relational aggression, respectively. They also identified 51, 60, and 48% of students who experienced one or more acts of physical, verbal, and relational victimization, respectively.

### Teachers’ Ratings of Students’ Aggression and Victimization

Coefficients in Table 1 reflect the overall patterns of change in teachers’ ratings of aggression and victimization within and across project years and the extent to which implementing the intervention altered these patterns. There were significant intervention effects on teachers’ ratings of students for all three forms of aggression, with effects emerging in different years for different constructs. There were no significant intervention effects for any outcome during the first implementation year (see coefficients for intervention year 1 in Table 1). Teachers’ ratings of students’ frequency of verbal and relational

**Table 1** Intercepts and effect size estimates (standard errors) from multilevel models predicting teacher outcome ratings of students’ behavior across project waves

| Effect              | Physical aggression | Verbal aggression | Relational aggression | Physical victimization | Verbal victimization | Relational victimization |
|---------------------|---------------------|-------------------|-----------------------|------------------------|----------------------|--------------------------|
| Intercept           | 1.37 (0.03)         | 1.47 (0.04)       | 1.22 (0.03)           | 1.36 (0.03)            | 1.54 (0.04)          | 1.27 (0.03)              |
| School A            | −0.20 (0.09)*       | 0.02 (0.09)       | 0.20 (0.09)*          | −0.31 (0.09)***        | −0.15 (0.08)         | 0.08 (0.09)              |
| School B            | 0.07 (0.07)         | 0.21 (0.07)**     | 0.36 (0.07)***        | −0.07 (0.07)           | −0.06 (0.07)         | 0.10 (0.07)              |
| Year 2              | −0.27 (0.06)***     | −0.09 (0.06)      | −0.18 (0.06)**        | −0.28 (0.06)***        | −0.22 (0.06)***      | −0.20 (0.06)**           |
| Year 3              | −0.03 (0.08)        | 0.07 (0.08)       | 0.08 (0.08)           | −0.10 (0.08)           | −0.10 (0.08)         | 0.00 (0.08)              |
| Year 4              | −0.04 (0.09)        | 0.14 (0.09)       | 0.18 (0.09)*          | −0.18 (0.08)*          | −0.06 (0.08)         | 0.09 (0.08)              |
| Year 5              | 0.03 (0.10)         | 0.21 (0.10)*      | 0.29 (0.10)**         | −0.17 (0.09)           | −0.03 (0.09)         | 0.15 (0.09)              |
| Linear slope        | 0.11 (0.02)***      | 0.14 (0.02)***    | 0.12 (0.02)***        | 0.08 (0.02)***         | 0.12 (0.02)***       | 0.09 (0.02)***           |
| Intervention year 1 | 0.01 (0.08)         | −0.05 (0.08)      | −0.13 (0.08)          | 0.03 (0.08)            | 0.06 (0.08)          | −0.04 (0.08)             |
| Intervention year 2 | −0.16 (0.09)        | −0.29 (0.10)**    | −0.46 (0.10)***       | −0.15 (0.09)           | −0.21 (0.09)*        | −0.41 (0.09)***          |
| Intervention year 3 | −0.24 (0.11)*       | −0.30 (0.11)**    | −0.50 (0.11)***       | −0.11 (0.10)           | −0.16 (0.10)         | −0.42 (0.10)***          |
| Intervention year 4 | −0.33 (0.14)*       | −0.38 (0.14)**    | −0.69 (0.14)***       | −0.11 (0.14)           | −0.30 (0.14)*        | −0.56 (0.14)***          |

Values are intercepts and effect size estimates (*d*-coefficients) with standard errors in parentheses. School A and school B effects represent baseline differences between the two schools that received the intervention and school C. Year 2 to year 5 effects represent changes relative to year 1 after controlling for intervention effects. Linear slope represents linear change within each school year intervention year 1 to intervention year 4 effects represent changes in each intervention year relative to the baseline

\**p* < .05; \*\**p* < .01; \*\*\**p* < .001

**Table 2** Coefficients from multilevel models predicting students' ratings of their behavior across project waves

| Effect              | Physical aggression | Relational aggression | Victimization   |
|---------------------|---------------------|-----------------------|-----------------|
| Intercept           | 1.49 (0.03)         | 1.31 (0.02)           | 1.41 (0.02)     |
| School A            | −0.17 (0.08)*       | −0.19 (0.08)*         | 0.03 (0.08)     |
| School B            | −0.09 (0.07)        | −0.15 (0.07)*         | −0.05 (0.07)    |
| Year 2              | −0.11 (0.06)*       | −0.19 (0.06)***       | −0.09 (0.06)    |
| Year 3              | −0.13 (0.07)        | −0.25 (0.07)***       | −0.22 (0.07)**  |
| Year 4              | −0.22 (0.08)**      | −0.36 (0.08)***       | −0.30 (0.08)*** |
| Year 5              | −0.34 (0.09)***     | −0.41 (0.08)***       | −0.43 (0.09)**  |
| Linear slope        | 0.02 (0.02)         | −0.05 (0.02)*         | −0.06 (0.02)**  |
| Summer              | −0.18 (0.05)***     | −0.08 (0.05)          | −0.09 (0.05)*   |
| Intervention year 1 | 0.05 (0.07)         | 0.14 (0.07)*          | −0.08 (0.07)    |
| Intervention year 2 | 0.10 (0.09)         | 0.13 (0.08)           | −0.08 (0.09)    |
| Intervention year 3 | 0.16 (0.10)         | 0.14 (0.09)           | −0.12 (0.10)    |
| Intervention year 4 | 0.15 (0.13)         | 0.13 (0.12)           | −0.2 (0.13)     |

Values are intercepts and effect size estimates ( $d$ -coefficients) with standard errors in parentheses. School A and school B effects represent baseline differences between the two schools that received the intervention and school C. Year 2 to year 5 effects represent changes relative to year 1 after controlling for intervention effects. Linear slope represents linear change within each school year intervention year 1 to intervention year 4 effects represent changes in each intervention year relative to the baseline

\* $p < .05$ ; \*\* $p < .01$ ; \*\*\* $p < .001$

aggression showed significant reductions in the intervention schools relative to the control condition beginning in year 2 of implementation ( $ds = -.29$  and  $-.46$ , respectively). These effects remained significant during the third and fourth implementation years for verbal aggression ( $ds = -.30$  and  $-.38$ , respectively) and relational aggression ( $ds = -.50$  and  $-.69$ , respectively). Significant reductions in teachers' ratings of physical aggression occurred during the third year of implementation ( $d = -.24$ ) and were maintained the following year ( $d = -.33$ ). Figure 2a illustrates the pattern for relational aggression, which shows a decrease following implementation of the intervention (i.e., in year 2 for school A and year 4 for school B) and a pattern of increases in school C, which did not receive the intervention.

We also found significant intervention effects on teachers' ratings of students' frequency of verbal and relational victimization. Effects emerged during the second year of implementation ( $ds = -.21$  and  $-.41$ , respectively), and remained significant during the third intervention year for relational victimization ( $d = -.42$ ) and during the fourth year for verbal and relational victimization ( $ds = -.30$  and  $-.56$ ). There were, however, no significant intervention effects on teachers' ratings of students' frequency of physical victimization. Figure 2b, which depicts the pattern for relational victimization, shows decreasing rates following implementation of the intervention in the two intervention schools, with a relatively consistent rate over time in the school where the intervention was not implemented.

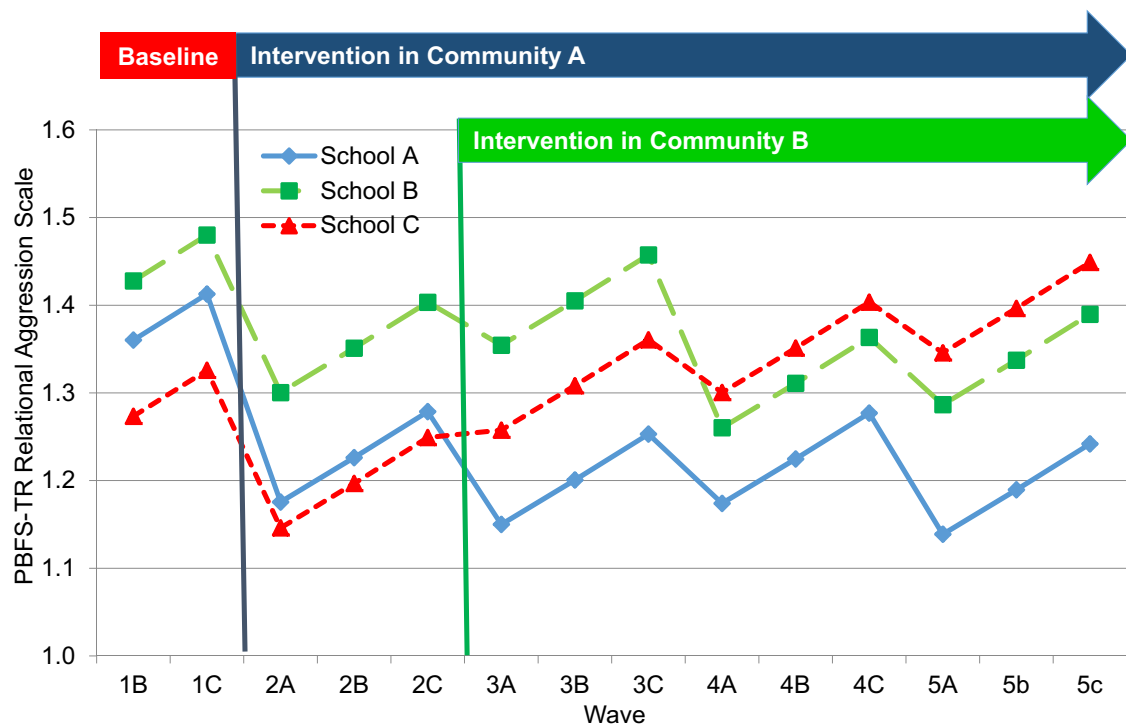
We expanded the main effects model to evaluate the consistency of intervention effects across schools, grades, and gender. There were no significant school by implementation year or gender by implementation year interactions for teachers' ratings of

aggression or victimization. This suggests that the intervention had consistent effects at both schools where it was implemented, and that effects did not vary by gender. We did, however, find differences in intervention effects across grades for teachers' ratings of verbal aggression ( $F[8,2353] = 1.99$ ,  $p = .044$ ), relational aggression ( $F[8,2422] = 4.56$ ,  $p < .001$ ), verbal victimization ( $F[8,2423] = 2.63$ ,  $p = .007$ ), and relational victimization ( $F[8,2452] = 6.53$ ,  $p < .001$ ). These indicated that the intervention varied in how many years it had to be implemented before effects emerged for students in specific grades. For example, reductions in teachers' ratings of relational aggression were obtained for seventh graders during the first year of implementation (see Fig. 3a). Although effects were not significant for sixth graders until the second year of implementation, effects were large when they did emerge ( $d = -.74$ ,  $p < .001$ ). The strongest effects for all grades were found during the fourth year of implementation. A similar pattern was found for teachers' ratings of relational victimization (see Fig. 3b) such that significant reductions emerged for seventh graders during the first year of implementation and large effects were obtained for sixth graders starting in the second year of implementation.

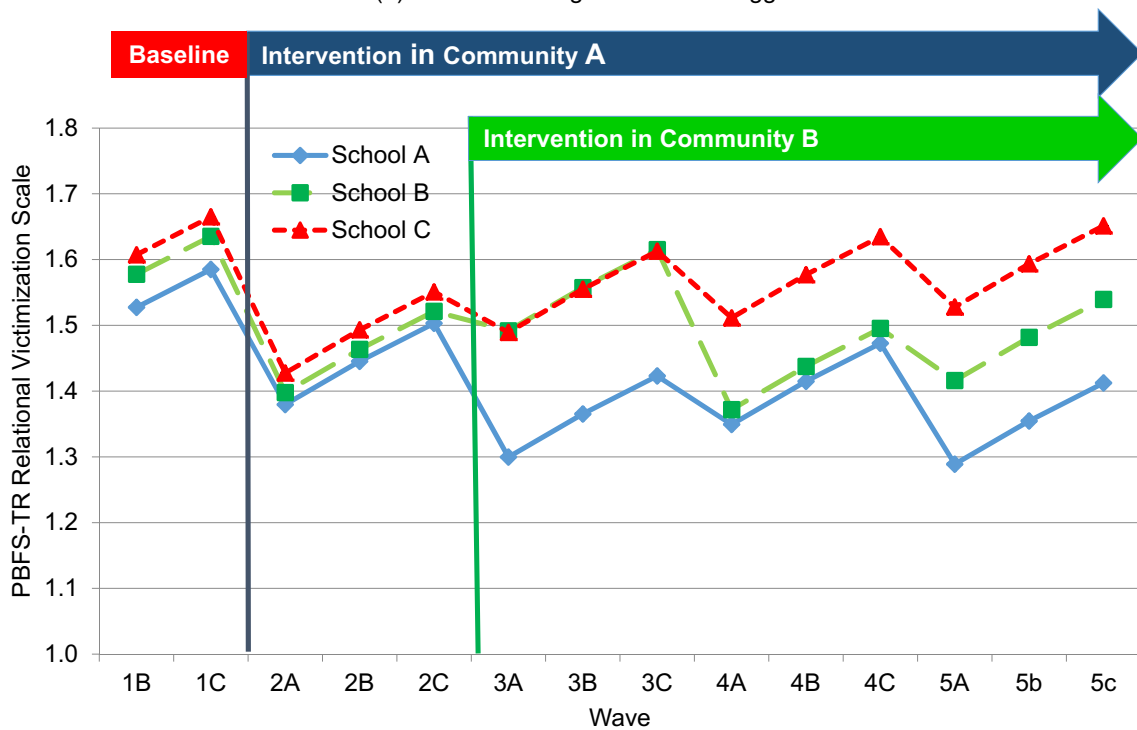
### Student Reports of Aggression and Victimization

Analyses of student-reported frequencies of their aggression and victimization did not reveal any significant intervention effects (see Table 2). Although there was an overall pattern of decreasing frequencies across project years, implementing the intervention did not alter this pattern. Analyses examining moderators did not reveal any significant differences in intervention effects across schools, gender, or grades.





(a) Teachers' ratings of relational aggression



(b) Teachers' ratings of relational victimization

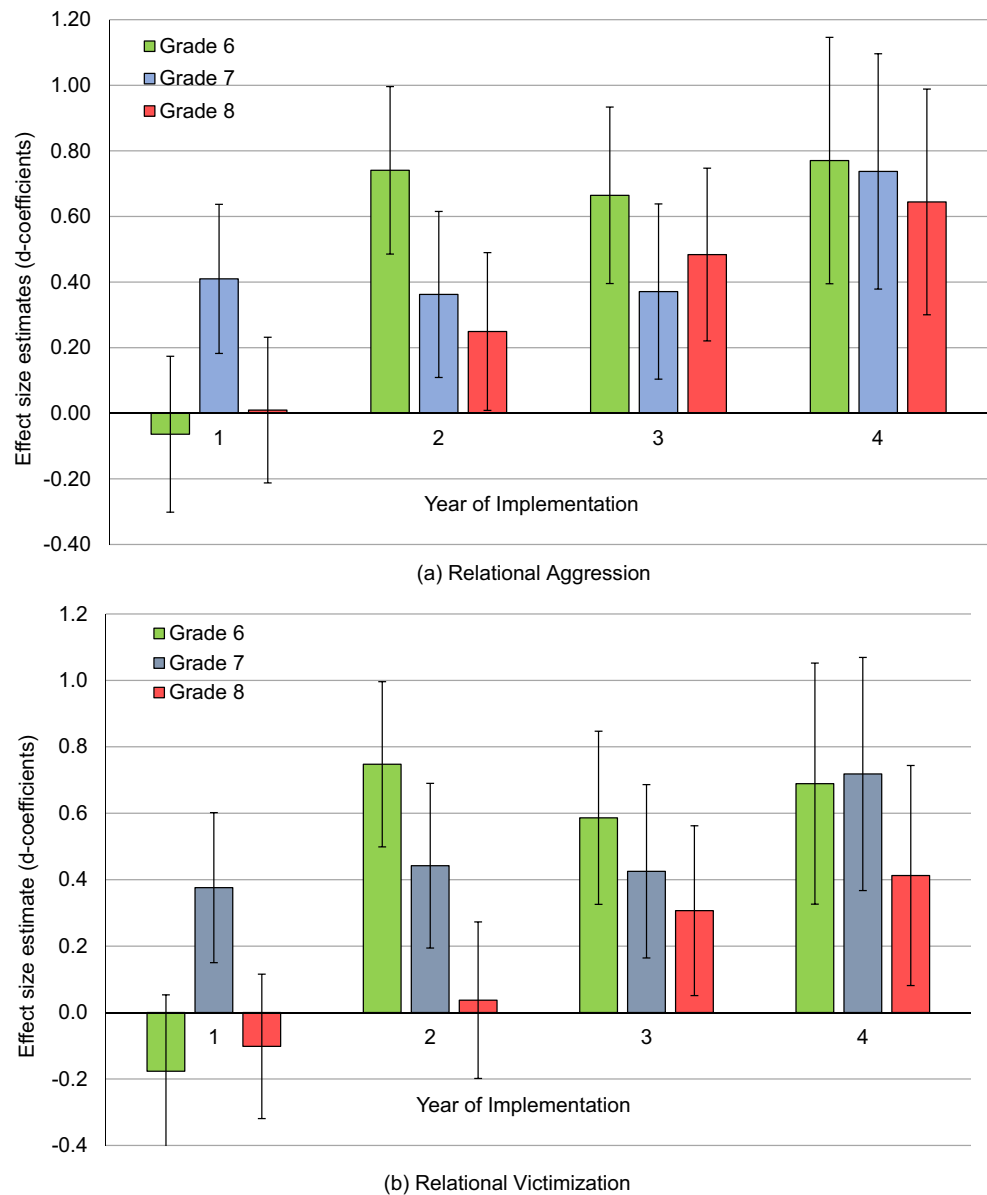
Fig. 2 Means by wave. Wave number indicates project year and letter indicates wave within year (A, Fall; B, Winter; C, Spring)

### Ratings of School Safety and Norms

There was a significant positive linear slope for teachers' ratings of school safety problems indicating increasing concerns within each school year, but a significant negative slope for

student ratings indicating decreasing concerns about safety problems (see Table 3). There was also a tendency for an overall decrease in both teacher and student ratings of safety problems across the years of the project relative to year 1. Our analyses revealed several significant intervention effects.

**Fig. 3** Effect size estimates (*d*-coefficients) and error bars (95% confidence intervals) for intervention effects on teachers' ratings of (a) aggression and (b) victimization by student grade and year of implementation. Coefficients were reversed such that positive coefficients represent beneficial effects (i.e., decreases)



Implementation of the OBPP was related to an increase in teachers' concern about the extent to which students' involvement in high-risk behavior was a serious problem in their school during the first year of intervention ( $d = .44$ ). This effect was maintained in subsequent years ( $ds = .66$  to  $1.22$ ). The intervention's effects on teachers' ratings of school safety problems did not differ across the two schools where the OBPP was implemented. Analyses of teachers' ratings of school norms did not reveal any intervention effects.

For student ratings, there was a significant intervention effect representing an increase in students' ratings of the frequency of school safety problems ( $d = .18$ ). This effect was small and was not evident in subsequent years. Effects on student ratings of school safety problems did not differ across schools, grades, or gender.

## Discussion

The goal of this study was to evaluate the effectiveness of the OBPP in urban middle schools in the USA that served a predominantly African American student population. We found that the impact of the OBPP varied across teacher and student reports and across outcomes. Reductions in teachers' ratings of different forms of students' aggression and victimization emerged during different years of implementation. These effects were consistent across schools and gender, but varied across grades for relational and verbal aggression and for victimization. In contrast to teachers' ratings, we did not find intervention effects on students' reports of their behavior. Analyses of school climate found that implementing the intervention was associated with increases in teachers' concerns

**Table 3** Coefficients from multilevel models predicting teacher and student ratings of school-level variables across project waves

| Effect              | School safety problems (T) | School safety problems (S) | Norms for aggression (T) | Norms for nonviolence (T) |
|---------------------|----------------------------|----------------------------|--------------------------|---------------------------|
| Intercept           | 1.09 (0.06)                | 1.83 (0.05)                | 1.16 (0.05)              | 1.08 (0.05)               |
| School A            | −0.78 (0.17)***            | 0.05 (0.09)                | −0.1 (0.18)              | 0.24 (0.18)               |
| School B            | −0.35 (0.14)*              | −0.16 (0.08)*              | 0.19 (0.15)              | 0.12 (0.15)               |
| Year 2              | −0.40 (0.12)**             | −0.23 (0.06)***            | 0.13 (0.13)              | 0.00 (0.13)               |
| Year 3              | −0.25 (0.16)               | −0.42 (0.08)***            | 0.47 (0.16)**            | −0.11 (0.16)              |
| Year 4              | −0.51 (0.18)**             | −0.46 (0.09)***            | 0.30 (0.19)              | −0.08 (0.18)              |
| Year 5              | −0.86 (0.22)***            | −1.01 (0.10)***            | 0.19 (0.22)              | 0.11 (0.22)               |
| Linear slope        | 0.10 (0.03)***             | −0.08 (0.02)***            | 0.06 (0.03)              | −0.06 (0.03)              |
| Intervention year 1 | 0.44 (0.15)**              | 0.18 (0.08)***             | −0.26 (0.16)             | −0.01 (0.16)              |
| Intervention year 2 | 0.66 (0.19)***             | 0.16 (0.10)                | −0.23 (0.20)             | −0.03 (0.20)              |
| Intervention year 3 | 1.05 (0.24)***             | 0.14 (0.11)                | −0.16 (0.25)             | −0.28 (0.25)              |
| Intervention year 4 | 1.22 (0.33)***             | 0.12 (0.14)                | 0.19 (0.34)              | −0.35 (0.34)              |

Values are intercepts and effect size estimates (*d*-coefficients) with standard errors in parentheses. School A and school B effects represent baseline differences between the two schools that received the intervention and school C. Year 2 to year 5 effects represent changes relative to year 1 after controlling for intervention effects. Linear slope represents linear change within each school year intervention year 1 to intervention year 4 effects represent changes in each intervention year relative to the baseline. ‘T’ indicates teacher report. ‘S’ indicates student-report measure

\* $p < .05$ ; \*\* $p < .01$ ; \*\*\* $p < .001$

about school safety problems, but no changes teacher ratings of norms or in students’ reports of school safety.

We found significant reductions in teachers’ ratings of students’ frequency of engaging in verbal, relational, and physical aggression, and on verbal and relational, but not physical victimization following implementation of the OBPP. Effects on verbal and relational aggression first appeared in the second year of implementation, but effects on physical aggression were not evident until the third year. These findings were consistent across schools and gender and were sustained across subsequent years of intervention. Variations in when effects emerged across outcomes suggest that changes in physical aggression require more sustained effort. This may be particularly true in urban schools where adolescents may feel compelled to respond aggressively to forms of provocation they feel cross the line (Farrell et al. 2010). This finding is consistent with several other studies that found lagged effects for the OBPP. For example, Black and Jackson (2007) found a gradual decline in bullying event density from 64 to 54% across the initial 3 years of intervention but a more pronounced decrease to 36% at the end of the fourth year. The increased effects over time in the present study emerged despite decreases in ratings of fidelity at one of the two intervention schools. This suggests that increased effects may reflect the cumulative impact of the intervention over time rather than an improvement in fidelity of implementing the class meetings. The stronger effect over time may reflect the fact that it

takes 3 years for students to experience the intervention during all 3 years of middle school.

Although we implemented the same intervention components across grades, we found differences in when intervention effects emerged for students in each grade. Intervention effects on teachers’ ratings of relational aggression and victimization were evident for seventh graders during the first year of implementation, but did not emerge for sixth and eighth graders until the second and third year of implementation, respectively. Bowllan (2011) found gender-specific grade differences reflecting positive intervention effects for seventh grade girls, nonsignificant effects for sixth grade girls, and increases in bullying and being bullied for eighth grade girls. However, their design did not assess the impact of additional years of implementing the OBPP. It may be that seventh grade is an optimum time for an intervention focused on school climate in that seventh graders have more social status than sixth graders, but are not as established in their patterns of behavior as eighth graders. The emergence of effects for sixth graders beginning in the second intervention year may reflect the impact of entering a school where upper classes have had prior exposure to the intervention. The increasing effect for eighth graders may reflect a dosage effect in that eighth grade students in the third and fourth year of implementation are more likely to have had exposure to the intervention during their prior two school years. Whatever the reason, our findings suggest that studies evaluating a single year of implementing violence prevention efforts may not fully assess

their potential, particularly if they do not assess for differential effects across grades.

In contrast to teachers' ratings, students' ratings did not reveal intervention effects on aggression or victimization. The absence of effects on student self-reports differs from a prior evaluation of the OBPP that found decreases in student reports of bullying perpetration (Schroeder et al. 2012). In contrast, Melton et al. (1998) also found initial effects during the first year of implementation that were not sustained over time or replicated when they implemented the intervention in control schools. Studies in the USA that evaluated the impact of the OBPP on student reports of victimization have had mixed findings. Whereas Black and Jackson (2007) found increases in student reports of being bullied, Melton et al. (1998) did not find significant effects on victimization. Bauer et al. (2007) found race and ethnic differences in the impact of the OBPP on student reports of relational and physical victimization such that decreases were only evident for White students. Their findings may explain the lack of significant effects in the present study, which focused on a primarily African American student population.

Our positive outcomes for teacher, but not student reports, suggest that the intervention's effects may have been limited to the school context. Prior evaluations of the OBPP have typically used the Olweus Bullying Questionnaire (Olweus 2005), which is restricted to bullying and victimization incidents at school. In contrast, the PBFS-AR assesses the frequency of aggression and victimization across contexts including school, home, neighborhood, and community. Although it includes acts of aggression that may not meet the more restrictive definition of bullying, it avoids concerns that students may underreport incidents labelled bullying because of the associated stigma (Bosworth et al. 1999; Lai and Kao 2018). Although it provides a more robust test of the impact of the OBPP on student behavior, the lack of significant findings for student-reported aggression and victimization was surprising and is at odds with effects on teachers' ratings of these student behaviors.

Our analysis of school-level outcomes did not show the expected pattern. Initiation of the intervention was associated with increases rather than decreases in teachers' ratings of school safety problems. It is important to note that our measure had teachers rate the extent to which specific issues were a problem at their school, not their frequency. It may be that implementing the intervention, which involved training teachers to identify and address bullying behaviors, increased their awareness of these behaviors or their concern about the extent to which they considered them a problem. With the exception of an initial increase during the first year of implementation, we did not find a similar pattern for student ratings, which were based on the frequency of school safety problems. The lack of effects on student ratings of climate is consistent with those of Bauer et al. (2007) who did not find significant intervention effects on middle school students'

ratings of classroom safety. The intervention also failed to produce the expected effects on teachers' ratings of student norms supporting aggression and nonviolence. This may reflect the fact that these norms are influenced by multiple contexts including home, school, neighborhood, and community. For example, parental messages supporting violence and non-violence are highly influential and may attenuate the capacity of violence prevention efforts limited to the school context to influence these attitudes (Farrell et al. 2010). Further work using more objective measures such as observational measures is needed to address this issue.

The current study is one of the few efforts evaluating the OBPP to assess the frequency of specific forms of aggression and victimization (e.g., relational, physical, and verbal) (for an exception, see Bauer et al. 2007) across multiple years of implementation. Our inclusion of a measure of teachers' ratings of individual students differs from other studies in the USA and internationally that have either relied exclusively on student reports of bullying (Bauer et al. 2007; Eslea and Smith 1998; Melton et al. 1998; Pepler et al. 1994) or had teachers rate the overall extent of bullying at their school (Bowllan 2011; O'Moore and Minton 2005; Olweus and Limber 2010; Schroeder et al. 2012). We believe aggregating teachers' ratings of the specific behaviors of individual students provides a more objective and less biased indicator of change than global ratings that represent general impressions of school climate. Our findings of differences in effects for specific forms of aggression and victimization over time also highlight the benefit of examining the impact of the OBPP on specific forms of aggression and victimization rather than on general measures of bullying.

## Limitations

Several issues related to the study's design and analyses should be acknowledged. Although our multiple-baseline design, which involved randomization of both the order and timing of interventions, offers significant advantages over designs that do not employ randomization (e.g., Biglan et al. 2000), it also has limitations. The presence of differences across schools at baseline is concerning. With some exceptions, we found slightly lower rates of problems in the two schools where we implemented the intervention. This reduces the possibility of regression to the mean. Intervention effects should also be considered within the context of changes in each school over time. For example, although the significant intervention effect for relational aggression reflected a decrease following initiation of the intervention in the two schools where it was implemented, it also reflected an increase in rates at the school where it was not implemented (see Fig. 2a). This finding is in contrast to the effects on relational victimization for which there were decreases in rates at the two intervention schools and a relatively stable rate in the

school where it was not implemented (see Fig. 2b). A significant strength of the multiple-baseline design is that it allowed us to examine the consistency of effects for the two schools where the intervention was initiated in different years, and to evaluate the extent to which effects were sustained over time. The consistency of intervention effects across the two schools suggests that the observed effects were not simply the result of increasing problems at the school that served as the control. Differences in the focus of student and teacher measures also complicates the interpretation of our findings. The significant findings on teachers' ratings, but not on student report, may indicate that program effects did not generalize outside of the school setting. Although it could also reflect bias in teachers' ratings, teachers' ratings of increased school safety problems and the lack of significant intervention effects on their ratings of student norms are not consistent with this premise. Our ability to detect intervention effects may have also been affected by limited variability in the frequency of aggression based on the past 30 days, particularly for physical aggression and victimization.

Several factors related to our implementation of OBPP should be considered in interpreting our findings. Although our use of observations to assess the fidelity of teachers' delivery of classroom meetings is a strength of our study, we did not assess fidelity of other components of the OBPP. We also tailored many of the class meetings to address issues relevant to our urban population. Although consistent with the key principles of the OBPP, our modified version might differ from other attempts to tailor the OBPP to meet the needs of other populations. The inclusion of multiple intervention components is also a limitation. We implemented the OBPP in conjunction with an attempt to implement a family-based intervention. However, challenges engaging families in the intervention resulted in a small percentage of families within the schools completing the family intervention. This makes it unlikely that it played a strong role in producing intervention effects at the school level. Our intervention strategy also included an after-school leadership program. Although this was designed to complement the OBPP and was only implemented with a subset of students, this may have also contributed to any intervention effects.

## Future Directions

Our findings highlight several directions for future research. First, more work is needed to develop comprehensive intervention approaches that integrate school and community-based services across multiple contexts (e.g., home, school, neighborhood, and community) to address the needs of youth for universal prevention and selective interventions (Matjasko et al. 2016). Given the delay in intervention effects for physical aggression, more work is needed to identify how the OBPP, as a universal prevention program, most effectively combines with selective

and indicated interventions to meet the needs of specific populations of youth, particularly those at greatest risk. For example, the OBPP includes an individual-level component with individual supports for students who perpetrated bullying or were victims. Further work is needed to determine the degree to which this dovetails with selective and indicated interventions (e.g., Crone et al. 2010) for students who are struggling with bullying behavior. Intervening early following disciplinary referrals may help reduce risk for aggression and physical victimization. It is also important to consider potentially effective combinations of the OBPP with other universal youth violence prevention programs that address individual-level skill development. More research is also needed to determine how best to equip teachers to implement the OBPP to the extent that their ability to effectively deliver and sustain this program requires general classroom management skills. Moreover, given its multiple components and time needed to implement them effectively, the OBPP requires ongoing administrative support to sustain it at the individual school and district levels (Olweus and Limber 2010). An important goal of the OBPP is to increase students' and teachers' awareness and ability to identify bullying behaviors that occur at school. However, this may also serve to heighten perceptions of a negative school climate and highlights the ongoing need to emphasize students' and teachers' positive and prosocial behavior to counterbalance this perception.

Our analyses of teacher ratings revealed lagged effects that across outcomes. Further research using designs that examine changes in intervention effects over time is needed to identify the factors that may be responsible for these findings. These delayed effects may be the result of better integration of the intervention into the school over time, or the impact of bringing students into a school where students in older grades have been exposed to multiple years of implementation. At the least, they emphasize the need for patience among both researchers and administrators who introduce a new intervention.

Implementing interventions in under-resourced schools and communities poses considerable challenges. School-based interventions in such settings may have limited impact if other components are not in place. Conceptual models of implementation (e.g., Han and Weiss 2005) have identified factors hypothesized to influence implementation of school-based programs at a variety of levels, including macro-level factors (e.g., policies, funding); school and program level factors (e.g., school culture and climate, administrative support and leadership); and teacher level factors (e.g., training, experience). For the current study, we employed a full-time youth prevention specialist who split his time across two schools to assist with implementation at the classroom and school level. We also monitored the fidelity of classroom meeting implementation using direct observations of teacher delivery and provided feedback if the fidelity of implementation for a particular component decreased (Goncy et al. 2015). Unfortunately, it is difficult to compare fidelity of implementation of the current study, particularly of the classroom

meeting component, with previous studies as they have either not measured classroom meeting fidelity (e.g., Schroeder et al. 2012) or have used self-report checklists rather than observations (e.g., Black and Jackson 2007). Nonetheless, unless administrative staff time and support is devoted to implementing the OBPP on a consistent basis, maintaining high fidelity for implementation may be challenging.

A key issue is determining how best to tailor the OBPP to address the needs of students in settings that differ from where it was initially developed. The OBPP provides considerable leeway for the focus of class lessons and other components compared with other more manualized interventions. We attempted to tailor the OBPP (e.g., the class meeting topics and anti-bullying activities) to fit the schools in our community. Although this produced some desirable effects based on teachers' ratings, these effects took time to emerge and similar effects were not found for student ratings. This underscores the need for further work to adapt other aspects of the OBPP (e.g., providing classroom management training and supports to teachers) to meet the needs of middle schools in the USA. Teachers and administrators in the US public school systems have limited time to implement prevention efforts. This may limit the support they can provide for this program. As Melton et al. (1998) noted, the task of implementing comprehensive violence prevention programs poses particular challenges for middle schools where bullying behaviors reach their peak and the school structure may present difficulties for implementing many of the core elements of the OBPP. Moreover, students attending under-resourced schools and living in poverty face a variety of environmental factors that increase their risk for aggression including peer, family, and neighborhood factors (Farrell et al. 2010).

**Acknowledgements** The authors are grateful to John Ferron for his input on this project's design and analysis, and to Anne Greene who served as project coordinator.

**Funding** This study was funded by the National Center for Injury Prevention and Control, Centers for Disease Control and Prevention, CDC Cooperative Agreement 5U01CE001956. The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention.

## Compliance with Ethical Standards

**Ethical Approval** All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards.

**Informed Consent** Informed consent was obtained from all individual participants in the study.

**Conflict of Interest** The authors declare that they have no conflict of interest.

## References

- Bauer, N. S., Lozano, P., & Rivara, F. P. (2007). The effectiveness of the Olweus Bullying Prevention Program in public middle schools: A controlled trial. *Journal of Adolescent Health, 40*, 266–274. <https://doi.org/10.1016/j.jadohealth.2006.10.005>.
- Biglan, A., Ary, D., & Wagenaar, A. C. (2000). The value of interrupted time-series experiments for community intervention research. *Prevention Science, 1*, 31–49. <https://doi.org/10.1023/A:1010024016308>.
- Black, S. A., & Jackson, E. (2007). Using bullying incident density to evaluate the Olweus Bullying Prevention Program. *School Psychology International, 28*, 623–638. <https://doi.org/10.1177/0143034307085662>.
- Bosworth, K., Espelage, D. L., & Simon, T. R. (1999). Factors associated with bullying behavior in middle school students. *Journal of Early Adolescence, 19*, 341–362. <https://doi.org/10.1177/0272431699019003003>.
- Bowllan, N. M. (2011). Implementation and evaluation of a comprehensive, school-wide bullying prevention program in an urban/suburban middle school. *Journal of School Health, 81*, 167–173. <https://doi.org/10.1111/j.1746-1561.2010.00576.x>.
- Crone, D. A., Hawken, L. S., & Horner, R. H. (2010). *Responding to problem behavior in schools: The behavior education program* (2nd ed.). New York: Guilford.
- Eslea, M., & Smith, P. K. (1998). The long-term effectiveness of anti-bullying work in primary schools. *Educational Research, 40*, 203–218. <https://doi.org/10.1080/0013188980400208>.
- Farrell, A. D., & Vulin-Reynolds, M. (2007). Violent behavior and the science of prevention. In D. Flannery, A. Vazonsyi, & I. Waldman (Eds.), *Cambridge handbook of violent behavior* (pp. 766–786). New York: Cambridge University Press. <https://doi.org/10.1017/CBO9780511816840.041>.
- Farrell, A. D., Mays, S., Bettencourt, A., Erwin, E. H., Vulin-Reynolds, M., & Allison, K. W. (2010). Environmental influences on fighting versus nonviolent behavior in peer situations: A qualitative study with urban African American adolescents. *American Journal of Community Psychology, 46*, 19–35. <https://doi.org/10.1007/s10464-010-9331-z>.
- Farrell, A. D., Sullivan, T. N., Goncy, E. A., & Le, A. T. H. (2016). Assessment of adolescents' victimization, aggression, and problem behaviors: Evaluation of the Problem Behavior Frequency Scale. *Psychological Assessment, 28*, 702–714. <https://doi.org/10.1037/pas0000225>.
- Farrell, A. D., Thompson, E. L., Sullivan, T. N., & Goncy, E. A. (2017). Assessment of adolescents' victimization, aggression, and problem behaviors: Evaluation of the Problem Behavior Frequency Scale - Adolescent Report. Manuscript submitted for publication.
- Farrell, A. D., Goncy, E. A., Sullivan, T. N., & Thompson, E. L. (2018). Evaluation of the problem behavior frequency scale—teacher report form for assessing behavior in a sample of urban adolescents. *Psychological Assessment*. <https://doi.org/10.1037/pas0000571>.
- Fuentes, V. E., Goncy, E. A., & Sutherland, K. S. (2015). Cross-cultural perspectives after participation in the YES program: A pilot study. *Journal of Youth Development, 10*, 64–73. <https://doi.org/10.5195/JYD.2015.8>.
- Goncy, L. A., Sutherland, K. S., Farrell, A. D., Sullivan, T. N., & Doyle, S. T. (2015). Measuring teacher implementation in delivery of a bullying prevention program: The impact of instructional and protocol adherence and competence on student responsiveness. *Prevention Science, 16*, 440–450. <https://doi.org/10.1007/s11121-014-0508-9>.
- Gordon, D. A. (2000). Parent training via CD-ROM: Using technology to disseminate effective prevention practices. *Journal of Primary*

- Prevention*, 21, 227–251. <https://doi.org/10.1023/A:1007035320118>.
- Haggerty, K. P., Skinner, M. L., MacKenzie, E. P., & Catalano, R. F. (2007). A randomized trial of parents who care: Effects on key outcomes at 24-month follow-up. *Prevention Science*, 8, 249–260. <https://doi.org/10.1007/s11121-007-0077-2>.
- Han, S. S., & Weiss, B. (2005). Sustainability of teacher implementation of school-based mental health programs. *Journal of Abnormal Child Psychology*, 33, 665–679. <https://doi.org/10.1007/s10802-005-7646-2>.
- Henry, D. B., Farrell, A. D., Schoeny, M. E., Tolan, P. H., & Dymnicki, A. (2011). Influence of school-level variables on aggression and associated attitudes during middle school. *Journal of School Psychology*, 49, 481–503. <https://doi.org/10.1016/j.jsp.2011.04.007>.
- Lai, T., & Kao, G. (2018). Hit, robbed, and put down (but not bullied): Underreporting of bullying by minority and male students. *Journal of Youth and Adolescence*, 47, 618–635. <https://doi.org/10.1007/s10964-017-0748-7>.
- Limber, S. P. (2011). Development, evaluation, and future directions of the Olweus Bullying Prevention Program. *Journal of School Violence*, 10, 71–87. <https://doi.org/10.1080/15388220.2010.519375>.
- Matjasko, J. L., Vivolo-Kantor, A. M., Massetti, G. M., Holland, K. M., Holt, M. K., & Cruz, J. D. (2012). A systematic meta-review of evaluations of youth violence prevention programs: Common and divergent findings from 25 years of meta-analyses and systematic reviews. *Aggression and Violent Behavior*, 17, 540–552. <https://doi.org/10.1016/j.avb.2011.10.001>.
- Matjasko, J. L., Massetti, G. M., & Bacon, S. (2016). Implementing and evaluating comprehensive evidence-based approaches to prevent youth violence: Partnering to create communities where youth are safe from violence. *Journal of Primary Prevention*, 37, 109–119. <https://doi.org/10.1007/s10935-016-0422-y>.
- Melton, G. B., Limber, S. P., Cunningham, P., Osgood, D. W., Chambers, J., Flerx, V., et al. (1998). *Violence among rural youth. Final report to the Office of Juvenile Justice and Delinquency Prevention*. Washington, DC: Office of Juvenile Justice and Delinquency Prevention.
- Olweus, D. (2005). A useful evaluation design, and effects of the Olweus Bullying Prevention Program. *Psychology, Crime, & Law*, 11, 389–402. <https://doi.org/10.1080/10683160500255471>.
- Olweus, D., & Alsaker, F. D. (1991). Assessing change in a cohort longitudinal study with hierarchical data. In D. Magnusson, L. R. Bergman, G. Rudinger, & B. Torestad (Eds.), *Problems and methods in longitudinal research* (pp. 107–132). New York: Cambridge University Press. <https://doi.org/10.1017/CBO9780511663260.008>.
- Olweus, D., & Limber, S. P. (2010). Bullying in school: Evaluation and dissemination of the Olweus Bullying Prevention Program. *American Journal of Orthopsychiatry*, 80, 124–134. <https://doi.org/10.1111/j.1939-0025.2010.01015.x>.
- O'Moore, A. M., & Minton, S. J. (2005). Evaluation of the effectiveness of an anti-bullying programme in primary schools. *Aggressive Behavior*, 31, 609–622. <https://doi.org/10.1002/ab.20098>.
- Pepler, D. J., Craig, W. M., Ziegler, S., & Charach, A. (1994). An evaluation of an anti-bullying intervention in Toronto schools. *Canadian Journal of Community*, 13, 95–110. <https://doi.org/10.7870/cjcmh-1994-0014>.
- Schroeder, B. A., Messina, A., Schroeder, D., Good, K., Barto, S., Saylor, J., & Masiello, M. (2012). The implementation of a statewide bullying prevention program: Preliminary findings from the field and the importance of coalitions. *Health Promotion Practice*, 13, 489–495. <https://doi.org/10.1177/1524839910386887>.
- Sullivan, T. N., Sutherland, K. S., Farrell, A. D., & Taylor, K. A. (2016). School as venues for prevention programming. In K. Bosworth (Ed.), *Prevention science in school settings: Complex relationships and processes* (pp. 201–226). Springer: New York.
- Zimmerman, M. A., Stewart, S. E., Morrel-Samuels, S., Franzen, S., & Reischl, T. M. (2011). Youth empowerment solutions for peaceful communities: Combining theory and practice in a community-level violence prevention curriculum. *Health Promotion Practice*, 12, 425–439. <https://doi.org/10.1177/1524839909357316>.