American Journal of Orthopsychiatry

Evaluating Dosage Effects for the Positive Action Program: How Implementation Impacts Internalizing Symptoms, Aggression, School Hassles, and Self-Esteem

Paul R. Smokowski, Shenyang Guo, Qi Wu, Caroline B. R. Evans, Katie L. Cotter, and Martica Bacallao

Online First Publication, March 7, 2016. http://dx.doi.org/10.1037/ort0000167

American Journal of Orthopsychiatry

http://dx.doi.org/10.1037/ort0000167

Evaluating Dosage Effects for the Positive Action Program: How Implementation Impacts Internalizing Symptoms, Aggression, School Hassles, and Self-Esteem

Paul R. Smokowski
University of North Carolina – Chapel Hill and University of Kansas

Shenyang Guo
Washington University

Qi Wu and Caroline B. R. Evans
University of North Carolina – Chapel Hill

Katie L. Cotter
Arizona State University

Martica Bacallao
University of Kansas

Positive Action (PA) is a school-based intervention for elementary-, middle-, and high-school students that aims to decrease problem behaviors (e.g., violence, substance use) and increase positive behaviors (e.g., academic achievement, school engagement). PA has a long history of documented success achieving these aims, making it an Evidence Based Practice (EBP). Intervention research on EBP's has established the importance of implementation fidelity, especially with regard to program dosage; failure to properly implement an EBP can have negative consequences on targeted outcomes, especially if participants are exposed to a low dosage of the program (e.g., fewer lessons than specified). Much of the current research on PA has neglected to examine how program dosage impacts PA's effect on targeted outcomes. Using propensity score models, multiple imputation, and a 2-level hierarchical linear model, the current study fills this gap and examines how different dosages of PA as measured by years participating in PA and number of PA lessons, impacts adolescent internalizing symptoms, aggression, perceptions of school hassles, and self-esteem over a 3-year period. The current sample included middle school students in grades 6, 7, and 8 ($N = 5{,}894$). The findings indicate that students who received 3 years of the PA intervention and a high number of PA lessons had a significantly higher self-esteem score than those who received 0 years of PA or zero lessons. Participants who received 1 year of PA also reported significantly lower school hassle scores than those who received 0 years. Dosage had no statistically significant effects on aggression or internalizing score. Implications are discussed.

Positive Action (PA) is an intervention implemented in elementary, middle, and high school that aims to improve academic achievement, school attendance, problem behaviors (e.g., substance use, violence, disruptive behaviors, dropping out of school, sexual behavior), parent–child bonding, family cohesion, and family conflict (National Registry of Evidence-Based Programs and Practices, 2014). The PA curriculum consists of a series of kits, with age-appropriate lesson plans and materials that assist teachers in guiding students through PA. The elementary-school PA curriculum consists of seven kits for use with kindergarten-aged

Paul R. Smokowski, University of North Carolina – Chapel Hill, North Carolina Academic Center for Excellence in Youth Violence Prevention, and School of Social Welfare, University of Kansas; Shenyang Guo, Department of Social Work, Washington University; Qi Wu and Caroline B. R. Evans, University of North Carolina – Chapel Hill, North Carolina Academic Center for Excellence in Youth Violence Prevention; Katie L. Cotter, Department of Social Work, Arizona State University; Martica Bacallao, School of Social Welfare, University of Kansas.

Caroline B. R. Evans is now at the School of Social Welfare, University of Kansas.

Funding for this research was provided through a cooperative agreement with the U.S. Centers for Disease Control and Prevention's National Center for Injury Prevention and Control (5 U01 CE001948-03).

Correspondence concerning this article should be addressed to Paul R. Smokowski, University of Kansas, School of Social Welfare, 1545 Lilac Lane, Lawrence KS 66045. E-mail: smokowski@ku.edu

youth to youth in grade 6; each kit consists of 140 brief lessons ranging from about 10 to 15 minutes in length. The middle school PA curriculum consists of two kits for use with youth in grades 7 and 8 that contain 82 lessons ranging from about 15 to 20 minutes in length. Every kit has materials that address the following six units: (a) Self-Concept, designed to enhance students' self-understanding and identity; (b) Positive Actions for Your Body and Mind, designed to teach proper hygiene, exercise, and creative thinking skills; (c) Managing Yourself Responsibly, designed to teach students skills to effectively manage time and resources; (d) Treating Others the Way You Like to be Treated, designed to teach students positive social skills; (e) Telling Yourself the Truth, designed to teach the importance of self-honesty; and (f) Improving Yourself Continually, designed to teach students how to apply Positive Action skills in all areas of life and convey the idea that self-improvement is an ongoing, continual process (Positive Action, n.d.).

Created in 1977, PA has a long history of documented success and it is currently listed on the Substance Abuse and Mental Health Services Administration's (SAMHSA) National Registry of Evidence-Based Programs and Practices (NREPP) website, establishing PA as an Evidence-Based Practice (EBP; NREPP, 2014). Further, Blueprints for Healthy Youth Development, a database of violence, delinquency, and drug prevention programs, has labeled PA as a *model program*, a designation reserved for programs that have undergone stringent evaluation (i.e., one high-quality randomized controlled trial, or one high-quality randomized controlled trial and one high-quality quasi-experimental evaluation) and demonstrate positive program effects at least 12 months following program completion (Blueprints for Healthy Development, 2012). The U.S. Department of Justice recognized PA as an effective program given the strong evidence indicating that PA achieves its intended outcomes (National Institute of Justice, n.d.). In addition, the U.S. Department of Education's (DoE) *What Works Clearing House* gave PA a positive rating in 2007 based on evidence that the intervention had positive effects on its targeted outcomes (DoE, 2007) and in 2009, the DoE's Safe, Disciplined, and Drug-Free Schools Expert Panel labeled PA as a promising program (DoE, 2009). Despite the evidence and accolades supporting the general utility of PA, little research has examined whether and how the implementation of PA (e.g., program dosage) affects the programs' impact on targeted outcomes. The current study addresses this gap in the literature and assesses how program dosage of PA as measured by number of years participating in the PA program and number of PA lessons, impacts internalizing symptoms, aggression, school hassles, and self-esteem in a large sample of rural, middle school aged youth.

Evidence-Based Practice, Implementation Science, and Program Dosage

Evidence-Based Practices (EBP's) are interventions supported by scientific evidence indicating that they improve participant outcomes (Drake et al., 2001). Four "intermediate outcomes" or steps are required to successfully utilize EBP's: (a) Identification of and access to EBP's relevant and appropriate for the desired population and purpose; (b) Acceptance of the scientific evidence supporting the EBP and the decision to adopt a particular EBP; (c) Implementation of the EBP; and (d) Evaluation of the efficacy of the EBP (Proctor, 2004; p. 228). Step 3, implementation, is a vital step that impacts how EBP's effect targeted outcomes and poor implementation fidelity often results in diminished program effects (Ennett et al., 2011). A thorough review of intervention research led researchers to identify five core aspects of implementation fidelity: adherence (the degree to which program components were delivered according to program manuals), exposure (the amount of a program participants received as measured by number of sessions, session length, or program frequency), quality of delivery (how well the program was delivered as measured by factors such as leader preparedness and enthusiasm, and effectiveness of sessions), participant responsiveness (a measure of how responsive participants were in each session), and program differentiation (a measure to ensure participants in each experimental condition received only the planned intervention; Dane & Schneider, 1998). Exposure is a critical component of implementation fidelity (Ennett et al., 2011), and participant attendance (i.e., the amount of dosage of a program youth receive) effects how an intervention impacts targeted outcomes (Odom, 2009). Despite the impact that exposure has on intervention results, researchers have found that for school-based interventions, teachers typically expose youth to two thirds to three-quarters of an intervention (see Ennett et al., 2011 for a review), indicating that youth rarely receive the full dosage of school-based interventions. Indeed, close scrutiny of PA program evaluations shows no evidence of implementation quality or dosage being linked to program outcomes. With a curriculum that contains more than 75 lessons at every level, the program designer reports that the average number of lessons implemented during past studies of PA is 25, one third of the recommended lesson count (C. Allred, personal communication, 2013). This partial implementation of school-based interventions likely decreases the intended positive effects these programs have on improving student behaviors and academic outcomes. It also leaves an open question as to whether past program effects were actually attributable to other unmeasured factors rather than lesson content. The current study examines whether and how exposure to the PA program impacts PA's effects on targeted outcomes.

Past Research on Positive Action and Dosage

Although there is a large body of research assessing the effects of the PA program, few studies have examined whether and how program dosage impacts outcomes. For example, a six-year, eight-wave study using a sample of 1,170 low-income urban youth in grades 3 through 8 found that relative to control schools that did not participate in PA, schools participating in PA reported reduced absenteeism, a decrease in the natural increase in students' disaffection with learning, higher teacher ratings of student academic motivation and ability, and increased math and reading scores, especially for African American boys and youth receiving free and reduced price lunch (Bavarian et al., 2013). However, this study neglected to assess the number of years students' were exposed to the program or the number of PA lessons in which students participated, rendering it impossible to determine whether exposure to PA impacted changes in the targeted outcomes. We assume that PA was responsible for the observed changes; however, program exposure was not documented, leaving scant evidence.

Other evaluations of PA have reported that program participation resulted in decreased violence (Beets et al., 2009; Flay, Allred,

& Ordway, 2001; Flay & Allred, 2003; Lewis, Schure, et al., 2013; Li et al., 2011; Snyder et al., 2013), absenteeism (Snyder et al., 2009), suspensions (Snyder et al., 2009), bullying (Li et al., 2011), substance use (Beets et al., 2009), sexual activity (Beets et al., 2009), depression and anxiety (mediated by social-emotional and character development; Lewis, DuBois, et al., 2013), school hassles (Guo et al., 2015), and substance use (Li et al., 2011; Snyder et al., 2013). In addition, research on PA has also shown program participation resulted in increased academic performance (Flay et al., 2001; Flay & Allred, 2003; Snyder et al., 2009, 2013), improved school involvement (Flay & Allred, 2003), and self-esteem (Guo et al., 2015). However, the majority of these studies neglected to assess whether and how the amount of program exposure impacted the targeted outcomes. Only two of the aforementioned studies attempted to assess the impact of program exposure. One evaluation of PA used a matched-pairs randomized control design with 14 elementary schools in Chicago to test a dose–response relationship between exposure to PA and changes in targeted outcomes. Researchers found that compared with newcomers to the PA program (youth who had not received any of the PA program), stayers (youth who had been engaged in the PA program for more time) did not show any differences in terms of substance use, serious violence, and disruptive behavior. However, newcomers showed stronger program effects for bullying compared to stayers (Li et al., 2011). Although this study examined PA dosage, it did so in a rudimentary way, using a binary variable to indicate whether each participant was a stayer or newcomer; thus, there was no measure of exact number years or lessons received.

A second study of youth in grade 5 attending rural and urban schools in Hawaii found that, compared with youth who received 1 to 2 years of PA, youth who received 3 to 4 years of PA reported significantly less substance use and violent behavior according to self- and teacher-reports and significantly less self-reported sexual activity. Although this study did not assess how the number of PA lessons impacted outcomes, the findings indicated that an accumulation of exposure to PA over time improved program effects (Beets et al., 2009).

Taken together, past research on PA highlights that program participation successfully decreases a number of problematic behaviors (e.g., violence, substance use) and increases positive behaviors (e.g., academic functioning, school engagement). However, little PA research examines how exposure to the program (as measured by years or number of lessons) impacts program effects on targeted outcomes. Preliminary findings suggest that engagement in PA over 3 to 4 years is most beneficial, however more research is needed to confirm these findings.

Hypothesis for Current Study

Using four waves of panel data, the current analysis investigates whether and how the amount of exposure to the PA program (i.e., number of years participating in PA and number of PA lessons) is associated with adolescent functioning. Specifically, it was hypothesized that compared with youth who received no years of PA or no PA lessons, youth who received increasing years of PA or more PA lessons would report decreased internalizing symptoms, aggression, and school hassles and increased self-esteem following program completion.

Method

Current Study

A collaborative partnership between the Centers for Disease Control and Prevention (CDC) and the North Carolina Youth Violence Prevention Center (NC-YVPC) funded the current study. The study sample came from the NC-YVPC Rural Adaptation Project (RAP), a 5-year longitudinal panel study of more than 7,000 middle- and high-school students from 27 public middle schools and 11 public high schools in two rural, low income counties in North Carolina; one county served as the intervention county and received three interventions (Positive Action, Teen Court, and Parenting Wisely), whereas the other county received no interventions and served as the comparison county. A very small number of youth who participated in PA also went through Teen Court or had a parent who participated in Parenting Wisely, thus it is unlikely that exposure to these two interventions impacted the effects of the PA program. The current study used four waves of the RAP panel data collected between 2011 and 2014 to assess how the dosage of Positive Action (PA) impacts adolescent functioning.

Sample

In Year 1 of the RAP study, all middle school students (grades 6 through 8) in the comparison county were included in the sample. Because the intervention county was larger both geographically and in student population, a random sample of 40% of youth in grades 6 through 8 were included from the intervention county. Students from both counties were followed longitudinally as they moved through middle school and into high school. At the start of each new academic year, the incoming cohort of sixth graders from the comparison county and a random sample of 500 sixth graders from the intervention county were added to the RAP sample. During the 5-year RAP study, PA was administered for three years (Years 2, 3, and 4 of the RAP study) to about 4,700 students in grades 6, 7, and 8 across the 13 middle schools in the intervention county.

Analytic Sample and Sample Size

The final sample consisted of 5,894 participants. The racial/ethnic diversity of the current sample mirrored the surrounding community and 28.0% of participants identified as White, 27.0% as American Indian, 25.0% as African American, 12.0% as mixed race/other, and 8.0% as Latino/Hispanic. About half of the sample (51.0%) was female, 89.0% of participants received free or reduced price lunch, and average age at baseline was 12.78 years. See Table 1 for baseline descriptive statistics and Tables 3 and 4 for analytic sample sizes by year and dosage analysis (i.e., years and lessons).

Implementation and Fidelity Procedures

Ideally, when PA is implemented in middle school classrooms (grades 6, 7, and 8), teachers should teach brief lessons (about 15 min) two or three days per week. In an attempt to achieve this

Table 1. *Baseline Descriptive Statistics*

Scale/variable	% or mean	*SE*
Gender (male)		
Female	.51	.008
Race (White)		
African American	.250	.007
Hispanic/Latino	.080	.004
Native American	.270	.004
Mixed race/other	.120	.005
Age at baseline	12.780	.016
Receipt of free/reduced lunch (no)		
Yes	.89	.005
School satisfaction	2.37	.008
School danger	1.79	.006
School hassles	1.49	.009
Discrimination experiences	1.43	.009
Religious orientation	2.30	.009
Parent support	2.68	.008
Teacher support	3.15	.009
Parent–child conflict	1.99	.039
Friend support	2.47	.009
Delinquent friends	1.38	.007
Peer pressure	1.32	.006
Peer rejection	1.30	.007
Aggression	1.33	.006
Internalizing symptoms	1.43	.008
Self-esteem	2.71	.007
Future optimism	3.460	.009

optimal implementation, NC-YVPC staff provided program materials, training, and supervision for all 13 middle schools in the intervention county. Each fall, NC-YVPC staff trained teachers and counselors in how to facilitate the program by modeling the implementation of PA lessons. The majority of middle schools implemented PA during social studies or health class and taught two or three lessons per week. To ensure consistency, NC-YVPC staff occasionally assisted teachers in implementing PA lessons.

By November in Year 2 of the RAP study (Year 1 of PA implementation), the 65 teachers from the 13 intervention middle schools, took over implementing the PA program. NC-YVPC staff monitored teachers' progress for implementation fidelity, observed teachers teaching PA, and completed rating forms to document that teachers demonstrated adequate implementation skills. Teachers used weekly Implementation Logs, provided by NC-YVPC staff, to document each completed lesson. NC-YVPC staff collected these logs and entered the information into an Excel spreadsheet to allow close monitoring of the progress within and across the 13 intervention schools. The Implementation Logs allowed for close monitoring of the dosage that students received, which was measured by the number of lessons taught and by the duration of the lessons. Despite barriers to implementation present in school environments (e.g., schedule changes, teacher absences), all 65 teachers reached, and most exceeded, their implementation goals for number of PA lessons, especially during Years 3 and 4 of the RAP study. Year 1 of PA implementation (Year 2 of the RAP study) was the most difficult because of complications recruiting appropriate staff to implement PA, building rapport and trust with principals, and finding incentives for teachers. School-level reports from NC-YVPC PA staff highlighted the barriers to implementation such as difficulties fitting PA into the existing school curriculum, unannounced changes to the school schedule that impacted PA implementation, teacher absences, teacher turnover, and teacher reassignment. However, by the third and final implementation year of PA (Year 4 of the RAP study), the 16 grade 6 teachers taught 1,193 PA lessons, ranging from 74 to 77 lessons per teacher over the year and exceeding their goal of 73 lessons for the year. The 24 grade 7 teachers taught 1,527 PA lessons, ranging from 63 to 65 lessons per teacher over the year and dramatically exceeding their goal of 51 lessons for the year. The grade 8 teachers taught 1,509 PA lessons, ranging from 60 to 64 lessons per teacher over the year, and exceeding their yearly goal of 45 lessons. According to the PA program designer, classroom teachers typically implement about 25 lessons per year (C. Allred, Personal communication, 2013). The implementation record obtained in the current study is almost triple that cited by the program designer, an impressive finding, especially given the chaotic and low income school district in the implementation county. In years 2 and 3 of PA implementation, teachers assigned to implement PA received a $50 incentive each month if lesson goals were met, which dramatically increased teacher motivation and fidelity to the PA program.

Additional PA Materials: Climate and Counselor Kits

In addition to the lesson kits for each grade level (described above), PA also has supplemental materials (climate and counselor kits) used to help schools establish an environment that reinforces the lessons of the PA program. The PA website describes the climate kits as containing:

> . . . items and activities to reinforce positive actions: words-of-the-week, newsletter templates, assemblies, stickers, tokens, positive notes, and so forth They reinforce the good feelings that students have when they do positive actions (that is, they help develop intrinsic, rather than extrinsic, motivation to engage in positive actions; Positive Action, n.d.).

NC-YVPC provided the 13 middle schools implementing PA with Climate Kits. The principal of each school chose teachers, staff, or community members volunteering in the schools to use the Climate Kit materials around the school (e.g., hallways, classrooms, offices) to highlight and reinforce PA program themes. In addition, counselors in the 13 schools received the Counselor Kits that contained "the text, *Positive Actions for Living,* used for additional counseling sessions and with individuals, small groups and families" (Positive Action, n.d.). Both the Climate and Counselor Kits were provided by the NC-YVPC study.

Data Collection Procedures

Following IRB approval from a major research university, almost identical data collection procedures were used in the intervention and comparison counties. In accordance with school district policies, the comparison county adopted the assessment as part of normal school proceedings, whereas the intervention county sent a letter home to all parents explaining the RAP study. Parents from the intervention county who did not want their child

to participate in the study sent a letter requesting nonparticipation, and their child was removed from the study roster. Students in both counties were advised that participation was voluntary and they were free to decline; students assented to participate by reading and electronically signing an assessment screen before beginning the online assessment. Students completed assessments in school computer labs, which were closely monitored by NC-YVPC research staff. To maintain confidentiality, each participant was assigned a unique identification number and no identifying information was collected.

Measures

A modified version of the School Success Profile (SSP; Bowen & Richman, 2008) was the primary measure used in the current study. The SSP is a 195-item online, youth self-report survey with 22 subscales measuring perceptions and attitudes about school, friends, family, neighborhood, self, health, and well-being. The SSP has been widely used since its creation in 1993 and has well-documented reliability and validity (Bowen, Rose, & Bowen, 2005). The RAP project used a modified version of the SSP, the School Success Profile Plus (SSP+), which included 17 of the original SSP subscale and 12 additional subscales; the SSP+ was created for the current project. The four dependent variables used in the current study were *internalizing symptoms*, *aggression*, *school hassles*, *and self-esteem.* Each model included the following measures as independent variables: demographics (i.e., gender, race, age, receipt of free/reduced price lunch), school experiences (i.e., school satisfaction, school danger, school hassles, discrimination experiences), social interactions with adults (i.e., parent support, teacher support, parent–child conflict), social interactions with peers (i.e., friend support, delinquent friends, peer pressure, friend rejection, religious orientation), and psychological well-being (i.e., future optimism, internalizing symptoms, aggression, self-esteem). Internalizing symptoms, aggression, school hassles, and self-esteem served as independent variables in the three models when they were not used as dependent variables. See Table 1 for baseline descriptive statistics and Table 2 for a description of measures.

Analytic Plan

The current study employed a quasi-experimental design and compared the outcome variables of internalizing symptoms, aggression, school hassles, and self-esteem among the group who received no PA intervention with the groups who received different doses of the PA intervention. Because the study did not and could not implement a randomized controlled trial, the evaluation of intervention dosage encounters a fundamental challenge: determining to which extent the changes on youth's outcomes over time can be attributed to the intervention per se and the amount of intervention. Under this context, selection into any level of dosage was not and could not be random, and hence, should be modeled first. Besides controlling for selection bias, the data analysis also needs to address several key methodological issues such as missing data imputation, the violation of a normality assumption embedded in the linear model, clustering effects inevitably existing in the study of change trajectories, and so on.

The need to correct for selectivity. Receipt of the PA intervention was not random, therefore youth from the nontreatment group and youth from the groups that received various doses of the PA treatment were imbalanced on covariates. Because of this imbalance, running a covariance control, such as a linear regression model or a hierarchical linear model, inevitably encounters a statistical problem known as endogeneity, resulting in biased and inefficient models (Guo & Fraser, 2015; Imbens, 2004; Sobel, 1996). To correct for endogeneity, we applied the Neyman-Rubin counterfactual framework as a conceptual model to guide the data analysis and each group with different treatment doses was compared with the zero-dose nonintervention group. According to the logic of the counterfactual model, individuals in the different treatment groups and the individuals in the nontreatment group have potential outcomes in both states, that is, the one in which they are observed and the one in which they are not observed. The Neyman-Rubin framework offers a practical way to evaluate counterfactuals, when random assignment is not used or when comparability is compromised by attrition or other factors. We employed propensity scores to balance different dosage groups. A propensity score is a conditional probability of a participant receiving treatment, given observed covariates (Rosenbaum & Rubin, 1983). The propensity score can be conceptualized as a balancing score representing a vector of covariates or "conditioning variables."

Dosage analysis. The propensity scores for each level of treatment were estimated by using a multinomial logit model and then we conducted an outcome analysis that employed the inverse of a specific propensity score as a sampling weight (Imbens, 2000). In this analysis, we conducted two separate dosage analysis using two different dosage variables: dosage of years and dosage of lessons. The dosage-of-years variable has four values: 0, 1, 2, and 3. Thus, four groups were classified in the dosage-of-years analysis. The majority of participants received 0 years of PA intervention (n = 3,936; 66.78%), 15% received one year of PA intervention (n = 884), 14.05% received two years of PA intervention (n = 828), and 4.17% received three years PA intervention (n = 246). Based on the dosage of lessons variable, five groups were classified: 0 lessons (n = 3,475; 58.96%), 0 to 31 lessons (n = 572, 9.7%), 31 to 63 lessons (n = 629, 10.67%), 63 to 103 lessons (n = 627, 10.64%), and 103 to 174.5 lessons (n = 591, 10.03%). In each dosage analysis, we ran three steps based on Imben's approach.

Step 1. We estimated generalized propensity scores (GPS) by using a mutinomial logit model. The conditional probability of receiving a particular dose of treatment was defined given the observed covariates as the GPS. We have four or five treatment doses in the two dosage analyses, and then each participant had four or five generalized propensity scores based on the number of groups.

Step 2. We conducted outcome analyses by following the process of propensity score weighting. Specifically, we calculated the inverse of a specific GPS and defined the inversed propensity score as a sampling weight to be used in outcome analysis. Denoting $e(x_{i,d}) = pr(D = d \mid X = x)$ as the generalized propensity score of receiving treatment dose d for participant i with observed covariates x, then the inverse of the GPS (i.e., $1/e(x_{i,d})$) is defined as a sampling weight for participant i. The outcome analysis then is a weighted model using $1/e(x_{i,d})$. Although each participant has

Table 2. *Description of Measures*

Measure	Type	Alpha (year 1; 2; 3; 4)	Response options	Example items
Self-esteem (5 items)	Level 1 Dependent	.87; .91; .92; .94	Not like me, A little like me, A lot like me	■ I am able to do things as well as most other people ■ I have confidence in myself
Aggression (12 items)	Level 1 Dependent	.86; .87; .86; .90	Not like me, A little like me, A lot like me	■ I get in many fights ■ I break rules at home, school, or elsewhere
Internalizing symptoms (7 items)	Level 1 Dependent	.89; .90; .91; .95	Not like me, A little like me, A lot like me	■ I often feel nervous or tense ■ I often feel fearful or anxious
School hassles (13 items)	Level 1 Dependent	.90; .92; .92; .93	Never, Once or twice, More than twice	■ Someone treated you in a disrespectful way ■ Someone at school pushed, shoved, or hit you
School satisfaction (7 items)	Level 2 Independent	.84; .85; .87; .88	Not like me, A little like me, A lot like me	■ I enjoy going to this school ■ I get along well with teachers at this school
School danger (11 items)	Level 2 Independent	.85; .87; .89; .91	Does not happen, Happens sometimes, Happens a lot	■ How often does each of the following happen at your school? Examples items included: Fights among students and Students carrying weapons.
Discrimination experiences (3 items)	Level 2 Independent	.71; .75; .76; .76	Never, Sometimes, Frequently, Always	■ How often do people dislike you because of your race or ethnicity? ■ How often have you seen friends treated unfairly because of their race or ethnicity?
Religious orientation (3 items)	Level 2 Independent	.88; .91; .92; .93	Not like me, A little like me, A lot like me	■ My religious faith gives me strength ■ My religious faith influences the decisions I make
Parent support (5 items)	Level 2 Independent	.89; .92; .93; .94	Never, Once or twice, More than twice	■ How often did the adults in your home let you know that you were loved? ■ How often did the adults in your home tell you that you did a good job?
Teacher support (8 items)	Level 2 Independent	.88; .90; .92; .92	Strongly disagree, Disagree, Agree, Strongly agree	■ My teachers care about me ■ My teachers give me a lot of encouragement
Parent–Child conflict (10 items)	Level 2 Independent	.82; .83; .84; .85	True, False	■ At least 3times a week, my parent(s) and I get angry at each other. ■ My parent(s) put me down
Friend support (5 items)	Level 2 Independent	.89; .91; .92; .94	Not like me, A little like me, A lot like me	■ I can count on my friends for support ■ I can trust my friends
Delinquent friends (9 items)	Level 2 Independent	.90; .91; .91; .92	Not like me, A little like me, A lot like me	■ I have friends who get in trouble with the police ■ I have friends who cut classes
Peer pressure (5 items)	Level 2 Independent	.73; .77; .79; .83	Not like me, A little like me, A lot like me	■ I let my friends talk me into doing things I really don't want to do ■ I tend to go along with the crowd
Friend rejection (3 items)	Level 2 Independent	70; .74; .75; .80	Not like me, A little like me, A lot like me	■ I am made fun of by my friends ■ I wish my friends would show me more respect
Future optimism (12 items)	Level 2 Independent	.88; .90; .92;, 97	Strongly disagree Disagree, Agree, Strongly agree	■ When I think about my future, I feel very positive ■ I see myself accomplishing great things in life

d (four or five in this analysis) propensity scores obtained from the multinomial logit model, we only used one such score in the following propensity score weighting analysis. This propensity score is the predicted probability for participant *i* to fall into the *d* dose category that is used, and the inverse of this score is used and defined in the propensity score weighting analysis.

Step 3. A similar weighted model using each covariate as a dependent variable was performed to check balance on the cova-

riate. For the dosage of years analysis, 26 variables in the intervention groups with different doses had significant differences compared with those in nonintervention group before the propensity score adjustment. The propensity score adjustment improved in terms of group balance. Only 10 variables were not balanced between different comparison groups after the adjustment. For the dosage of lessons analysis, 34 variables were not balanced before propensity score adjustment, and after adjustment, only 11 variables were not balanced between different comparison groups.

The need to control for clustering effects and the growth curve analysis. The current evaluation aims to analyze the impacts of different doses of the PA intervention on participants' change on four outcome variables over a 3-year study period. As such, only students who provided data for at least two waves were included in the analysis. Students who had data on only one wave were excluded from the analysis, because for these students the data do not qualify to the definition of outcome change.

The dataset in the current analysis has a typical nesting structure, because study times (i.e., four waves or occasions) are nested within students, and students are nested within schools. To correct for these clustering effects the violation of independent-observation assumption embedded in a linear regression model, we used hierarchical liner modeling (HLM). Because the data show trivial clustering of students within a school (i.e., the intraclass correlation coefficient on most outcomes was below .03), and as such, we applied a two-level HLM models instead of treating school as a third level. The two-level HLM is shown by the following combined equation:

$$\ln(Y_{ti}) = \gamma_{00} + \gamma_{10}(Time)_{ti} + \sum_{q=1}^{Q} \gamma_{0q}(X)_{qi} + r_{0i} + e_{ti}$$

where $\ln(Y_{ti})$ is the outcome variable of interest, $(Time)_{ti}$ is the time variable measured in months from baseline or Wave 1, $(X)_{qi}$ are Q student-level variables, r_{Oi} is a random effect for the ith student, and e_{ti} is a residual term incorporating temporal random effect for the ith student at time t.

Additional efforts made to address data issues are described below. First, all four outcome variables of interest had a skewed distribution and did not meet the normality assumptions about the outcome variable embedded in an HLM model. As such, we followed the convention in econometrics to take a natural-logarithm transformation of the dependent variable. Second, the analysis specifies a linear time variable only, which is typical for growth curve analysis using three- or four-point panel data where a quadratic or other type of curvilinear model adds unnecessary complexity (Raudenbush & Bryk, 2002). Third, we chose $Q = 21$ (for the outcome of internalizing symptoms, aggression, and school hassle) or 22 (for the outcome of self-esteem), or employed 21 or 22 predictor variables at level 2. Three of the outcome variables (internalizing symptoms, aggression, and school hassles) were treated as both outcome variables and covariates; when they were not the outcome variables, they were controlled in the analysis as covariates. However, self-esteem was only treated as an outcome variable in our analysis and was not used as a covariate. Thus, besides the 19 common covariates, aggression and school hassles were controlled to predict internalizing symptoms; internalizing symptoms and school hassles were controlled to predict aggression; and internalizing symptoms and aggression were controlled to predict school hassle. When self-esteem was the outcome variable, internalizing symptoms, aggression, and school hassles were used as covariates. Therefore, we had $Q = 21$ for three outcomes and $Q = 22$ for the last outcome (self-esteem). These 21 or 22 variables may be categorized into the following five types: (a) demographic variables (i.e., race, gender, age at baseline, receipt of free or reduced price lunch; (b) school experiences (i.e., school satisfaction, school danger, school hassles, discrimination experiences); (c) social interactions with adults (i.e., parent support, teacher support, parent–child conflict; (d) social interactions with peers (i.e., friend support, delinquent friends, peer pressure, friend rejection, religious orientation), and (e) psychological well-being (i.e., future optimism, internalizing symptoms, aggression, self-esteem). All these 21 or 22 variables were measured at the time when the student entered the study.

Finally, because the dependent variable in our final model takes a natural-logarithm transformation, we used the exponent of estimated coefficient [$exp(B)$] to ease the burden of interpretation of findings. The rationale for such presentation is that we control for all other variables included in the model at the level of zero. Doing so, all other coefficients are cancelled out, and only the estimated intercept and slope of interest remain in the equation. Using a ratio to compare two groups of a dichotomous variable, the estimated intercept is further dropped out. Suppose X is a dichotomous predictor variable, the ratio of model-predicted outcome values for the two groups of X under the condition of controlling for all other predictor variables at the zero level is as follows:

$$\frac{Y|X=1}{Y|X=0} = \frac{\exp(\hat{\beta}_0 + \hat{\beta}_1)}{\exp(\hat{\beta}_0)} = \exp(\hat{\beta}_0 + \hat{\beta}_1 - \hat{\beta}_0) = \exp(\hat{\beta}_1)$$

where $\hat{\beta}_0$ is model-estimated intercept and $\hat{\beta}_1$ is model-estimated slope for variable X. Thus, we can interpret the finding of the difference between $X = 1$ and $X = 0$ on the outcome Y as: the group of $X = 1$ on average has an outcome that is $[100 - 100^{*} \exp(\hat{\beta}_1)]\%$ lower than the outcome of the group of $X = 0$ when $\exp(\hat{\beta}_1) < 1$, and the group of $X = 1$ on average has an outcome that is $[100^{*}\exp(\hat{\beta}_1)]\% - 100]\%$ higher than the outcome of the group of $X = 0$ when $\exp(\hat{\beta}_1)]\% > 1$. A primarily X variable of interest is the Intervention Group with different doses: the percentage difference on this variable shows the direction of certain PA treatment dose impact, and test of its statistical significance generalizes the effect to the targeted population.

Because this study is the a program evaluation that examines whether the PA intervention has beneficial impacts for the participants, we performed directional hypothesis tests of treatment effects based on the theory of change of PA intervention and the coding schemes. For the four different outcome variables, we had a hypothesized sign for each coefficient for one-tailed tests. The hypothesized sign for "Aggression", "Internalizing" or "School Hassles" is "−," indicating hypotheses about a negative sign of the coefficient. The hypothesized sign for "Self-esteem" is "+," indicating hypotheses about a positive sign of the coefficient. For the demographic variables and other covariates, we performed nondirectional hypothesis tests with a given level of statistical significance.

Imputed data and the application of Rubin's rule. A typical problem frequently encountered in data collection studying change of youths' outcomes over time is the occurrence of missing data. We followed guidelines on handling missing data (Allison, 2002) and conducted multiple imputation. We created 15 imputed files and the final results were aggregated from these 15 multiply imputed files. There are not many existing studies showing the analytic procedure that combines the multiple imputation and growth curve modeling with propensity score dosage analysis into one procedure. Ultimately, the current investigation ought to incorporate three statistical models into one: the multiple imputation of missing data, the HLM-type growth curve analysis, and propensity score dosage analysis.

Based on a careful study of existing literature and numerous experimental runs, we developed the following analytic protocol: we first estimated the propensity scores for each of the 15 imputed files; we then conducted the weighted growth curve analysis using the estimated propensity scores as the weights of dosage for each of the 15 imputed files; and finally we aggregated the 15 sets of estimated results into one by using Rubin's (1987) rule. For the balance check, we ran several logistic regressions or OLS regressions with the weights of dosage in each imputed file. In the balance check, each covariate was treated as a dependent variable and the treatment doses were independent variables. The above protocol was applied to each of the four outcome variables.

Results

Results of the evaluation of the PA treatment dosage effects on internalizing symptoms, aggression, school hassles, and self-esteem are shown in Table 3 (dosage of years) and Table 4 (dosage of lessons). The results shown in the table are exp(B) or exponentiations of estimated coefficients, so a value of exp(B) that is greater than one indicates a positive sign of the coefficient, while a value of exp(B) that is less than one indicates a negative sign of the coefficient. The evaluation of treatment doses finds that after controlling for selection biases, a high dose of PA treatment generates statistically significant beneficial effects on the change of self-esteem score for both dosage of years and dosage of lessons. PA treatment also generates statistically significant beneficial effects on the change of school hassle score for dosage of years. The finding on the change of internalizing score shows an unexpected effect: the youth who received the intervention had higher internalizing scores than youth who did not receive the intervention. However, this unexpected effect was not statistically significant from the one-tailed test reported in this study, although it was statistically significant in a two-tailed test. Different doses of PA intervention had mixed impacts on aggression scores, but the impacts were not statistically significant.

Specifically, the results showed that, other things being equal and controlling for selection biases, students who received 3-years of PA intervention had a self-esteem score that was 5.3% higher than those who received zero years of the PA intervention ($p < .001$); students who received 103 to 174.5 PA lessons had a self-esteem score that was 6.2% higher than those who received zero PA lessons ($p < .001$). However, contrary to our hypothesis, students who received one year and two years of PA dosage had a self-esteem score that was 1.2% and 1.1% lower than those who received zero years of PA, but this was not statistically significant. And students who received lower dosages of PA lessons (0–31, 31–63, 63–103) had a self-esteem score that was 1.6%, 0.3%, and 3% lower than students who received zero PA lessons, but this difference was not statistically significant.

Different doses of the PA intervention also showed beneficial impacts on students' school hassles scores. For the analysis using dosage of years, results showed that students who received one year of the PA intervention had a school hassle score that was 1.6% lower than those who received zero years of the PA intervention ($p < .05$). Students who received two years of the PA intervention had a school hassles score that was 0.5% lower than those who received zero years of the PA intervention, although it was not statistically significant. However, counter to our hypothesis, students who received three years of the PA intervention had a school hassles score that was 2.6% higher than those who received zero years of the intervention, although this was not statistically significant. In addition, most of the results by using the dosage of lessons variable showed beneficial impacts on students' school hassle scores, although these effects were not significant. For example, students who received 0 to 31 lessons had a school hassles score that was 1.1% lower than those who received zero PA lessons (not statistically significant); Students who received 31 to 63 PA lessons had a school hassles score that was 1.6% lower than those who received zero PA lessons (not statistically significant); students who received 103 to 174.5 number of lessons of the PA intervention had a school hassles score that was 8.4% lower than those who received zero PA lessons (not statistically significant). However, those with 63 to 103 PA lessons had a school hassles score that was 0.7% higher than those who received no PA lessons (not statistically significant).

With regard to the change of the internalizing score, although all the different doses of the PA intervention for both year and lessons were associated with an unexpectedly higher internalizing score, the relationship was not statistically significant in the one-tailed test. However, the following relationships were significant in the two-tailed test, which were not reported in our results table: Students who received three years of the PA intervention had an internalizing score that was 24.4% higher than those who received zero years of the PA intervention ($p < .001$, two-tailed); Students who received 0 to 31 lessons of the PA intervention had an internalizing score that was 2.6% higher than those who received zero lessons of the PA intervention ($p < .05$, two-tailed); Students who received 63 to 103 lessons of the PA intervention had an internalizing score that was 4% higher than those who received zero lessons of the PA intervention ($p < .05$, two-tailed). Because the results on baseline check showed that a statistically significant mean difference between intervention groups and nonintervention group at baseline (the treated group had higher internalizing score to start with, $p < .001$), students' higher internalizing score in the intervention group were the result of higher internalizing score at baseline, rather than the failure of the PA intervention.

With regard to the change of the aggression score, some doses of intervention showed beneficial impacts, although not statistically significant. For example, students who received one year of the PA intervention had an aggression score that was

Table 3. *Growth Curve Modeling Results for Dosage of Years*

Fixed and random effects	Estimation based on 15 imputed files exp(B)			
	Aggression score	Internalizing score	Self-esteem	School hassles
Fixed effect				
Level 1: Time				
Time (months since baseline)	.998**	.999	.998***	.995***
Level 2: Individual				
Dosage of years (0 year)				
1 year	.997	1.011	.988	.984*
2 years	1.021	1.028	.989	.995
3 years	1.130	1.244	1.053***	1.026
Gender (Male)				
Female	1.055***	1.126***	.958***	.997
Race (White)				
African American	1.014	.996	1.082***	.970
Hispanic	.956	1.024	1.058**	.974
Native American	1.003	.997	1.043***	.969*
Mixed race and other	1.012	1.025	1.067***	.979
Age at baseline	1.009	1.007	.992	.990
Receipt of Free/reduced lunch (No)				
Yes	1.001	1.042*	1.026*	.989
School hassles	1.064***	1.096***	.977	
Internalizing symptoms	1.055***		.965**	1.047**
Aggression		1.156***	1.029	1.076**
Parent–child conflict	1.012***	1.019***	.996	1.006
Friend rejection	.967*	.997	.990	1.046**
Religious orientation	.938***	.957***	1.091***	.994
School satisfaction	.948***	.982	1.034**	.98
Future optimism	.990	.967	1.036**	1.023*
Parent support	1.007	.988	1.017	.998
Teacher support	1.024	1.031*	.996	.998
Friend support	1.025*	1.021	1.005	1.005
Delinquent friends	1.076***	.997	1.008	.974
Peer pressure	.999	.997	.993	.992
Perceived discrimination	.993	1.025	.983	1.046**
School danger	1.022	.977	.999	1.101***
Intercept	1.018	.941	2.006***	1.116
Random Effect (Variance Component)				
Level 2 intercept	.035***	.056***	.032***	.04***
Model Wald χ^2 (*df*) shown by one imputed file	723.53 (25)	846.28 (25)	799.75 (26)	683.19 (25)
Number of students in one imputed file				
At wave 1 (Time = 0 month)	3735	3715	3785	4271
At wave 2 (Time = 12 months)	3999	3981	4031	5140
At wave 3 (Time = 24 months)	4872	4839	4910	5894
At wave 4 (Time = 36 months)	4175	4216	4292	5894

Note. Reference group for categorical variables is shown in parenthesis after variable name.
* $p < .05$. ** $p < .01$. *** $p < .001$. One-tailed for directional hypothesis of treatment effects or two-tailed for nondirectional hypothesis test for other covariates.

0.3% lower than those who received zero years of the PA intervention. However, students who received two and three years of the PA intervention had an aggression score that was 2% and 13% higher than those who received zero years of the PA intervention (not statistically significant). Students who received 63 to 103 PA lessons had an aggression score that was 0.2% lower than those who received zero lessons (not statistically significant). However, students who received 0 to 31, 31 to 63, and 103 to 174.5 lessons had an aggression score that was 0.3%, 1.6%, and 0.8% higher than students who received zero PA lessons (not statistically significant).

Discussion

The current study filled a gap in existing PA research and examined how dosage of the PA program as measured by years and number of lessons was associated with changes in internalizing symptoms, aggression, school hassles, and self-esteem over a three-year period. Although extensive research has been conducted establishing PA as an EBP, the majority of this research was conducted in urban areas and neglected to examine the impact of program dosage on targeted outcomes, especially in rural environments. The amount of exposure to an EBP influences the effect of

Table 4. *Growth Curve Modeling Results For Dosage of Lessons*

Fixed and random effects	Estimation based on 15 imputed files exp(B)			
	Aggression score	Internalizing score	Self-esteem	School hassles
Fixed effect				
Level 1: Time				
Time (months since baseline)	.999*	.999	.998***	.995***
Level 2: Individual				
Dosage of lessons (0 lesson)				
0–31	1.003	1.026	.984	.989
31–63	1.016	1.016	.997	.984
63–103	.998	1.040	.970	1.007
103–174.5	1.008	1.118	1.062***	.916
Gender (male)				
Female	1.059***	1.130***	.971*	.997
Race (White)				
African American	1.013	1.019	1.060**	.991
Hispanic	.955	1.036	1.039	.988
Native American	.992	1.005	1.034*	.969
Mixed race and other	1.001	1.014	1.056**	.981
Age at baseline	1.008	1.007	.986*	.986*
Receipt of free/reduced lunch (No)				
Yes	1.004	1.028	1.037*	.988
School hassles	1.058**	1.090***	.982	
Internalizing symptoms	1.069***		.972	1.052**
Aggression		1.162***	1.036	1.078**
Parent child conflict	1.008	1.017***	.994	1.001
Friend rejection	.963*	.980	.983	1.040
Religious orientation	.939***	.972	1.092***	.995
School satisfaction	.961*	.993	1.039*	.992
Future optimism	.984	.961*	1.038**	1.020
Parent support	1.002	.990	1.008	.972
Teacher support	1.020	1.031	.991	.996
Friend support	1.034*	1.015	1.002	.988
Delinquent friends	1.089***	.980	.002	.974
Peer pressure	.995	.999	1.015	.990
Perceived discrimination	.998	1.015	.968**	1.039*
School danger	1.040	.995	1.007	1.106***
Intercept	.975	.908	2.129***	1.326*
Random effect (variance component)				
Level 2 intercept	.035***	.058***	.033***	.040***
Model Wald χ^2 (*df*) shown by one imputed file	538.57 (26)	385.20 (26)	1047.75 (27)	427.23 (26)
Number of students in one imputed file				
At wave 1 (Time = 0 month)	3735	3735	3785	4271
At wave 2 (Time = 12 months)	3999	3999	4031	5140
At wave 3 (Time = 24 months)	4872	4872	4910	5894
At wave 4 (Time = 36 months)	4175	4175	4292	5894

Note. Reference group for categorical variables is shown in parenthesis after variable name.
* $p < .05$. ** $p < .01$. *** $p < .001$. One-tailed for directional hypothesis of treatment effects or two-tailed for nondirectional hypothesis test for other covariates.

the EBP on targeted outcomes (Ennett et al., 2011; Odom, 2009), highlighting the importance of the current study. Based on past research, it was hypothesized that increases in exposure to the PA program as measured by years and number of PA lessons would be associated with decreases in internalizing symptoms, aggression, and school hassles and increases in self-esteem compared to youth who had zero years of PA or zero PA lessons. These hypotheses were partially supported.

Although aggression did decrease as PA dosage increased, this decrease was not statistically significant and was inconsistent (aggression was lower for one year of the program, but higher for two and three years of the program and was lower for 63–103 lessons, but higher for the other amounts of lessons). Given the violent context of the community in which PA was implemented, it is likely that participants were frequently exposed to violence and conflict, which normalized these behaviors and perhaps countered the peaceful teachings of PA. Thus, as PA worked to teach students to behave respectfully and nonviolently, the interpersonal and community context worked against this message, resulting in aggression scores that remained unaffected by dosage of the PA

program. Current findings indicate that even a high program dosage of PA was not enough to combat the constant barrage of violence to which participants were exposed.

These nonsignificant findings for aggression, which includes bullying behaviors and violence, run contrary to past reports of PA effectiveness (Beets et al., 2009; Flay et al., 2001; Flay & Allred, 2003; Guo et al., 2015; Lewis, Schure, et al., 2013; Li et al., 2011; Snyder et al., 2013). This brings up two unavoidable points for future clarification. First, when program evaluations include sophisticated dosage measures, moving beyond simple treatment versus comparison group analyses, perhaps the PA program is less effective than originally thought. EBP's once given special designations such as "model program" quickly get disseminated, often without continuing rigorous program evaluations. Dissemination can become business rather than science. Considering that past PA reports were based on a small number of rigorous trials, we strongly recommend the continuation of program evaluations for this EBP. Further, people adopting the program should closely scrutinize implementation quality and dosage. Second, as the research on PA accumulates, researchers should consider if the program is more effective in certain environments. Past studies may have shown effectiveness in lower risk environments (i.e., Hawaii; Beets et al., 2009). The current study took place in a particularly high-risk context and clearly shows limited effectiveness for lowering aggression, even when the program was implemented over three years and with an exceptional level of lesson dosage.

Also in line with our hypothesis, students who received one year (statistically significant) and two years (not statistically significant) of PA had school hassles scores that were lower than students who received zero-years of PA, however, three years of the program caused a nonsignificant increase in school hassles. The results after one year suggest that by attempting to increase students' positive behaviors and actions, PA has the potential to improve the overall school culture and climate, resulting in decreased school hassles (e.g., experiences of verbal, physical, and relational aggression). However, by year three of the program, school hassles had increased nonsignificantly. A similar pattern existed for lesson dosage, although all results were nonsignificant. When the PA program was new (0–31 lessons and 31–63 lessons) school hassles were lower, but after students had participated for a while (63–103 lessons and 103–174.5 lessons) school hassles increased. Perhaps when students are first introduced to PA, the lessons are novel and exciting and students enthusiastically embrace the notion of behaving positively, which results in increased prosocial behavior and decreased school hassles. However, as the novelty wears off over the years or over an increased number of lessons, students may become less likely to follow the tenants of the program causing a decrease in pro-social behavior and an increase in school hassles.

Finally, in line with our hypothesis, the current study found that increased exposure to PA as measured by an increased number of years and lessons was associated with a significant increase in self-esteem relative to zero years or zero lessons of PA. Students who had three years of PA or 103 to 174.5 lessons reported significant increases in self-esteem, however fewer years and a lower number of lessons resulted in nonsignificant decreases in self-esteem. Many of the six PA units focus on topics that likely influence and improve youths' self-esteem. For example, the self-concept unit helps students' shape their identity and understanding of themselves; increased self-understanding likely leads youth to value and appreciate themselves as individuals, thus bolstering their self-esteem. PA also encourages youth to take care of themselves both physically and mentally by reinforcing proper hygiene, exercise, and creative thinking skills (i.e., Positive Actions for Your Body and Mind unit). In this regard, PA teaches youth to value their physical appearance, physical health, and mental health and encourages them to engage in activities that promote physical and emotional wellbeing; looking and feeling healthy helps increase self-esteem. Many PA lessons focused on social interactions (i.e., Treating Others the Way You Like to be Treated unit), which encouraged youth to treat others as they would like to be treated. This likely helped improve youths' social interactions, perhaps leading to an increased social circle and more perceived social support, and ultimately making youth feel well liked and good about themselves (i.e., higher self-esteem). Given the direct conceptual link between many of the PA lesson units and improved self-esteem, it follows that the PA program would likely impact self-esteem. However, current findings indicate that in order to achieve this improvement in self-esteem, prolonged exposure to the PA program is needed. This finding is in line with past research that found improvements in violence, substance use, and sexual activity were not apparent until three or four years of PA had been administered (Beets et al., 2009), suggesting that PA teachings accumulate over time and result in positive improvements in youth outcomes after a few years of program participation. Youth with less years of PA or fewer lessons reported nonsignificant decreases in self-esteem. Perhaps minimal exposure to the program made youth aware of their deficits in self-confidence, resulting in decreased self-esteem. As youth were exposed to increased amounts of the PA program, they acquired the skills to increase their self-confidence and self-esteem.

This need for prolonged program implementation underscores the importance of studying dosage and implementation quality. Our evaluation did not show that teaching more lessons or more years is better, except in enhancing self-esteem and decreasing school hassles. The typical student in this study received 74 "core" lessons in grade 6, 64 in grade 7, and 60 grade 8, representing an effort that takes extraordinary commitment, hundreds of hours of instruction time, and approximately $100,000 in program materials purchased to serve 4,700 students annually in 13 schools. The next generation of studies of PA needs to discern whether a more efficient, parsimonious subset of lessons can be identified to decrease consumer burden. Further, researchers must begin to conduct cost-benefit studies to answer both the question, "Does the program do what it says it will?" and the question, "Is it worth the cost in monetary, intellectual, and human capital that is spent on implementation?" Currently, EBPs are often disseminated before the second cost-benefit question is answered, leaving the risk with consumers who lack this information up front.

Limitations

The current findings must be considered in light of certain limitations. First, randomly assigning schools to the intervention or control group would have been ideal, however this was not feasible; thus, propensity score analysis was used to correct for this limitation. Second, it would have been ideal for youth to fill out the

online assessments in private rooms, as their answers might have been impacted by the presence of their peers. Given limitations of time and space, this was not possible and trained research staff closely monitored the data collection process to ensure privacy and confidentiality. Third, the intervention and control counties had a few differences between them and it would have been ideal if two identical counties could have been used; future research on PA should attempt to use a matched-control group that is more similar to the intervention group. Finally, given that multiple teachers taught PA, it is likely that the lessons were taught in slightly different ways, which could have impacted the results. However, this is a limitation of all large scale intervention studies.

Conclusion

The current study investigated the effects of PA dosage, as measured by years and number of lessons, on the outcomes of internalizing symptoms, aggression, school hassles, and self-esteem. The findings indicated that increased exposure to the PA program in terms of both years and lessons was related to statistically significant increases in self-esteem and school hassle. Increased exposure to the PA program over years was associated with statistically significant increases in internalizing symptoms from the two-tailed tests, but the impacts were not significant from one-tailed tests. Contrary to past studies, dosage of the PA intervention was not statistically significantly associated with changes in aggression. Finally, effects of the PA program on targeted outcomes might be impacted by characteristics of the surrounding school and community context, suggesting that modifications to the program based on the surrounding milieu and level of risk might be warranted. Future program evaluations need to include dosage and implementation quality measures and establish cost-benefit ratios for evidenced based programs that are widely disseminated.

Keywords: positive action program; dosage; adolescent; implementation; rural

References

Allison, P. D. (2002). *Missing data*. Thousand Oaks, CA: Sage.

Bavarian, N., Lewis, K. M., Dubois, D. L., Acock, A., Vuchinich, S., Silverthorn, N., . . . Flay, B. R. (2013). Using social-emotional and character development to improve academic outcomes: A matched-pair, cluster-randomized controlled trial in low-income, urban schools. *The Journal of School Health, 83,* 771–779. http://dx.doi.org/10.1111/josh.12093

Beets, M. W., Flay, B. R., Vuchinich, S., Snyder, F. J., Acock, A., Li, K. K., . . . Durlak, J. (2009). Use of a social and character development program to prevent substance use, violent behaviors, and sexual activity among elementary-school students in Hawaii. *American Journal of Public Health, 99,* 1438–1445. http://dx.doi.org/10.2105/AJPH.2008.142919

Blueprints for Healthy Youth Development. (2012). *Program criteria*. Retrieved from http://www.blueprintsprograms.com/programCriteria.php

Bowen, G. L., & Richman, J. M. (2008). *The school success profile*. Chapel Hill, NC: University of North Carolina.

Bowen, G. L., Rose, R. A., & Bowen, N. K. (2005). *The reliability and validity of the school success profile*. Philadelphia, PA: Xlibris.

Dane, A. V., & Schneider, B. H. (1998). Program integrity in primary and early secondary prevention: Are implementation effects out of control? *Clinical Psychology Review, 18,* 23–45. http://dx.doi.org/10.1016/S0272-7358(97)00043-3

Drake, R. E., Goldman, H. H., Leff, H. S., Lehman, A. F., Dixon, L., Mueser, K. T., & Torrey, W. C. (2001). Implementing evidence-based practices in routine mental health service settings. *Psychiatric Services, 52,* 179–182. http://dx.doi.org/10.1176/appi.ps.52.2.179

Ennett, S. T., Haws, S., Ringwalt, C. L., Vincus, A. A., Hanley, S., Bowling, J. M., & Rohrbach, L. A. (2011). Evidence-based practice in school substance use prevention: Fidelity of implementation under real-world conditions. *Health Education Research, 26,* 361–371. http://dx.doi.org/10.1093/her/cyr013

Flay, B. R., & Allred, C. G. (2003). Long-term effects of the Positive Action program. *American Journal of Health Behavior, 27,* S6–S21. http://dx.doi.org/10.5993/AJHB.27.1.s1.2

Flay, B. R., Allred, C. G., & Ordway, N. (2001). Effects of the Positive Action program on achievement and discipline: Two matched-control comparisons. *Prevention Science, 2,* 71–89. http://dx.doi.org/10.1023/A:1011591613728

Guo, S., & Fraser, W. M. (2015). *Propensity score analysis: Statistical methods and applications* (2nd ed.). Thousand Oaks, CA: Sage.

Guo, S., Wu, Q., Smokowski, P. R., Bacallao, M., Evans, C. B. R., & Cotter, K. L. (2015). A longitudinal evaluation of the Positive Action program in a low-income, racially diverse, rural county: Effects on self-esteem, school hassles, aggression, and internalizing symptoms. Manuscript submitted for publication.

Imbens, G. W. (2000). The role of the propensity score in estimating does-response functions. *Biometrika, 87,* 706–710. http://dx.doi.org/10.1093/biomet/87.3.706

Imbens, G. W. (2004). Nonparametric estimation of average treatment effects under exogeneity: A review. *The Review of Economics and Statistics, 86,* 4–29. http://dx.doi.org/10.1162/003465304323023651

Lewis, K. M., DuBois, D. L., Bavarian, N., Acock, A., Silverthorn, N., Day, J., . . . Flay, B. R. (2013). Effects of Positive Action on the emotional health of urban youth: A cluster-randomized trial. *Journal of Adolescent Health, 53,* 706–711. http://dx.doi.org/10.1016/j.jadohealth.2013.06.012

Lewis, K. M., Schure, M. B., Bavarian, N., DuBois, D. L., Day, J., Ji, P., . . . Flay, B. R. (2013). Problem behavior and urban, low-income youth: A randomized controlled trial of positive action in Chicago. *American Journal of Preventive Medicine, 44,* 622–630. http://dx.doi.org/10.1016/j.amepre.2013.01.030

Li, K. K., Washburn, I., DuBois, D. L., Vuchinich, S., Ji, P., Brechling, V., . . . Flay, B. R. (2011). Effects of the Positive Action programme on problem behaviours in elementary school students: A matched-pair randomised control trial in Chicago. *Psychology & Health, 26,* 187–204. http://dx.doi.org/10.1080/08870446.2011.531574

National Institute of Justice. (n.d.). Program profile: Positive Action. Retrieved from http://www.crimesolutions.gov/ProgramDetails.aspx?ID=113

National Registry of Evidence-Based Programs and Practices. (2014). Positive Action. Retrieved from http://www.nrepp.samhsa.gov/ViewIntervention.aspx?id=400

Odom, S. L. (2009). The tie that binds: Evidence-based practice, implementation science, and outcomes for children. *Topics in Early Childhood Special Education, 29,* 53–61. http://dx.doi.org/10.1177/0271121408329171

Positive Action. (n.d.) *Positive Action overview*. Retrieved from https://www.positiveaction.net/

Proctor, E. K. (2004). Leverage points for the implementation of evidence-based practice. *Brief Treatment and Crisis Intervention, 4,* 227–242. http://dx.doi.org/10.1093/brief-treatment/mhh020

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika, 70,* 41–55. http://dx.doi.org/10.1093/biomet/70.1.41

Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. New York, NY: Wiley. http://dx.doi.org/10.1002/9780470316696

Snyder, F. J., Acock, A. C., Vuchinich, S., Beets, M. W., Washburn, I. J., & Flay, B. R. (2013). Preventing negative behaviors among elementary-school students through enhancing students' social-emotional and character development. *American Journal of Health Promotion, 28,* 50–58. http://dx.doi.org/10.4278/ajhp.120419-QUAN-207.2

Snyder, F., Flay, B., Vuchinich, S., Acock, A., Washburn, I., Beets, M., & Li, K. K. (2009). Impact of a social-emotional and character development program on school-level indicators of academic achievement, absenteeism, and disciplinary outcome: A matched-pair, cluster-randomized, controlled trial. *Journal of Research on Educational Effectiveness, 3,* 26–55. http://dx.doi.org/10.1080/19345740903353436

Sobel, M. E. (1996). An introduction to causal inference. *Sociological Methods & Research, 24,* 353–379. http://dx.doi.org/10.1177/0049124196024003004

U.S. Department of Education. (2007). *What works clearing house*. Retrieved from http://ies.ed.gov/ncee/wwc/interventionreport.aspx?sid=380

U.S. Department of Education. (2009). *Save, disciplined, and drug-free schools*. Retrieved from http://www2.ed.gov/admins/lead/safety/exemplary01/report_pg9.html