

## Effects of the *Positive Action* programme on problem behaviours in elementary school students: A matched-pair randomised control trial in Chicago

Kin-Kit Li<sup>a\*</sup>, Isaac Washburn<sup>b</sup>, David L. DuBois<sup>c</sup>, Samuel Vuchinich<sup>b</sup>, Peter Ji<sup>c</sup>, Vanessa Brechling<sup>c</sup>, Joseph Day<sup>c</sup>, Michael W. Beets<sup>d</sup>, Alan C. Acock<sup>b</sup>, Michael Berbaum<sup>c</sup>, Frank Snyder<sup>c</sup> and Brian R. Flay<sup>e\*\*</sup>

<sup>a</sup>Department of Applied Social Studies, City University of Hong Kong, Hong Kong; <sup>b</sup>Department of Human Development and Family Sciences, Oregon State University, Corvallis, OR 97331, USA; <sup>c</sup>Institute for Health Research and Policy, University of Illinois at Chicago, Chicago, IL 60608, USA; <sup>d</sup>Department of Exercise Science, Arnold School of Public Health, University of South Carolina, Columbia, SC 29208 USA; <sup>e</sup>Department of Public Health, College of Health and Human Sciences, Oregon State University, Corvallis, OR 97331, USA

(Received 25 March 2009; final version received 7 July 2010)

This study examined the effects of the *Positive Action* (PA) programme in Chicago Public Schools on problem behaviours among a cohort of elementary school students from grade three through grade five. Using a matched pair, randomised control design with 14 elementary schools, approximately 510 fifth graders self reported lifetime substance use, serious violence related behaviour, and current bullying and disruptive behaviours. Three level (i.e. students nested within schools within school pairs) overdispersed Poisson models were used to examine programme effects on the number of items endorsed for each of the four outcomes. Findings indicated that students in the intervention endorsed 31% fewer substance use behaviours (incidence rate ratio [IRR] 0.69), 37% fewer violence related behaviours (IRR 0.63) and 41% fewer bullying behaviours (IRR 0.59), respectively, compared to students in the control schools. Reduction in reported disruptive behaviours was of a similar magnitude (27%, IRR 0.73), but was not statistically significant. These results replicate findings of an earlier randomised trial of the PA programme and extend evidence of its effectiveness to youth attending large urban school systems.

**Keywords:** adolescent; matched pair; substance use; violence; bullying; disruptive behaviour

### Introduction

Violence and substance use are leading public health issues (Johnston, O'Malley, Bachman, & Schulenberg, 2008; Krug, Dahlberg, Mercy, Zwi, & Lozano, 2002). The initiation of these behaviours can date back as early as middle childhood.

---

\*Corresponding author. Email: ben.li@cityu.edu.hk

\*\*Corresponding author. Email: brian.flay@oregonstate.edu

A recent survey of 3530 third–fifth-graders, for example, found that 22% of the students reported involvement in bullying or being victimised at school (Glew, Fan, Katon, Rivara, & Kernic, 2005). A review of national and state-wide surveys similarly showed that the prevalence of having ever had more than a sip of alcohol was 10%, 16% and 29% for fourth, fifth and sixth graders, respectively (Donovan, 2007). Involvement in violence or substance use during childhood is clearly a source for concern. Evidence from longitudinal studies (e.g. Schulenberg & Maggs, 2008), furthermore, indicates that early manifestations of problem behaviour predict multiple long-term negative behaviour patterns and consequences.

Schools are regarded as an important context for implementing interventions to reduce problem behaviours. The Centers for Disease Control and Prevention (CDC), for example, has identified school-based approaches as best practices for aggression prevention (Thornton, Craft, Dahlberg, Lynch, & Baer, 2000). Timing of intervention is crucial as well given that prevention programmes need to be implemented before the onset of a problem (Nation et al., 2003). Andrews, Hampson, Barckley, Gerrard, and Gibbons (2008) suggested that alcohol and cigarette prevention programmes can start as early as second and fourth grade, respectively. Elementary schools clearly afford a promising context for interventions oriented towards preventing the onset of these and related problem behaviours.

Many problem behaviours co-occur and share some common causes (Brady, Tschann, Pasch, Flores, & Ozer, 2008; Flay, Snyder, & Petraitis, 2009). Accumulating research also supports the usefulness of comprehensive programmes that focus on multiple facets of character and social development for reducing multiple health risk behaviours (Flay & Allred, 2003; Hawkins, Kosterman, Catalano, Hill, & Abbott, 2005). School-based intervention programmes that are universal (i.e. programmes administered to all students in a classroom regardless of individual risks) demonstrate similar potential to reduce rates of substance use and violence (Hahn et al., 2007) and to have long-term benefits (Hawkins et al., 2005). The theory underlying these types of programmes often includes changes in the social environment, consistent with the *syndemic* perspective (Singer, 1994) and other comprehensive theories of health behaviour (Flay et al., 2009).

### ***Positive Action programme***

The *Positive Action* (PA) programme targets many influences on behaviour, both distal (e.g. school climate, teacher classroom management) and proximal (e.g. thoughts and feelings, self-efficacy) that are included as constructs in multiple health behaviour theories (Ajzen, 1991; Bandura, 1986; Hawkins & Weis, 1985). Many of these theories and perspectives are integrated into the Theory of Triadic Influence (TTI; Flay et al., 2009).

School-level analyses (relying on school report card data) using quasi-experimental, matched-control designs suggested the effectiveness of the PA programme for enhancing academic achievement and school involvement and reducing disciplinary referrals and violence in multiple locations (Flay & Allred, 2003; Flay, Allred, & Ordway, 2001). A recent evaluation of the programme in a 4-year trial in Hawaii that extended this research to a multilevel framework (i.e. individual data nested within schools) and the use of a matched-pair, randomised-control design (i.e. schools with similar characteristics were matched

into pairs and then schools within a pair were allocated randomly to either programme or control conditions) revealed significant reductions in fifth-grade student-reported substance use, violence and sexual behaviours in PA schools compared to control schools (Beets et al., 2009). This latter type of more rigorous research design provides the type of internal validity evidence that is critical to assessing the effectiveness of preventive interventions (Flay et al., 2005). It is important, however, for initial findings to be replicated, ideally with differing populations so as to better establish the external validity regarding an intervention's effectiveness. Evidence of external validity is particularly important for translation of research into practice and for decision makers and practitioners (Glasgow & Emmons, 2007).

### ***Purpose of the study***

This study had two major aims. The first was to replicate the existing evidence of effectiveness of the PA programme for reducing problem behaviours using the same multilevel framework and matched-pair, randomised-control design as in the Hawaii trial (Beets et al., 2008, 2009). The second aim was to examine the effectiveness of PA for problem behaviour reduction in a different population, elementary school students in a large urban school system – the Chicago Public Schools (CPS).

Evidence suggests that students in urban schools may have higher levels of aggression, delinquency and drug use when compared to those in rural schools (e.g. Farrell, Sullivan, Esposito, Meyer, & Valois, 2005). A number of societal processes may lead to such differences (Morales & Bonilla, 1993), including differential environmental exposure to stressors and the social consequences of poverty and inequality (Noguera, 2008). For instance, compared to students in rural and suburban schools, students in urban schools report higher exposure to violence (Campbell & Schwarz, 1996). In this study, as noted, one of our aims is to test whether the beneficial PA intervention effects generalise to urban populations.

## **Method**

### ***Design***

The Chicago trial of PA was under the auspices of the Social and Character Development (SACD) cooperative agreement funded by the Institute for Education Sciences of the US Department of Education. Chicago was one of seven sites nationwide evaluating the effects of different school-based interventions designed to promote students' SACD. Mathematica Policy Research (MPR) provided and administered a core set of survey questions to all the sites (Social and Character Development Research Consortium, in preparation). Site-specific questions were also included. The Chicago trial of PA took place in 14 CPS elementary schools during the 2004–2005 through 2006–2007 academic years, seven in the control condition and seven in the intervention condition. The study followed a single cohort of students who were third graders when treatment schools began implementing the programme. New students joining the grade level of this cohort in subsequent years were also included and followed. Five assessments were conducted during the study period: Fall 2004 (baseline assessment), Spring 2005, Fall 2005, Spring 2006 and 2007 (end of grade 5). Students for whom we obtained active parental consent and

personal assent for study participation and who provided data on substance use, violence, bullying and disruptive behaviours at the final assessment were included in the analyses. Although this study focused on negative behaviours, various psychological measures (e.g. self-efficacy, beliefs about aggression) and measures of prosocial behaviours were also included in the assessment. The results of the impact of the programme on these measures are available upon request. A high rate of return for parental consent forms by students (98%) was achieved by a class incentive and class visits and the parent affirmative rate (79%) was satisfactory (Ji et al., 2006). Rates of consent and assent and of data being obtained at baseline and at the final assessment were not significantly different across treatment and control conditions (Ji et al., 2006). Assessments and procedures were approved by the Institutional Review Boards of MPR, the University of Illinois at Chicago (UIC) and Oregon State University (OSU).

### *School and community context*

In the 2006–2007 school year, Hispanic and Black students accounted for 39% and 49% of the CPS student population, respectively, and 75% of CPS students were eligible for free or reduced-price lunch (Tang & Sable, 2009). Among students in grades 3–8, only 62% met or exceeded standards on the Illinois Standards Achievement Test (ISAT) in 2007 (Chicago Public Schools [CPS], 2009). About 16% of CPS students are English language learners, ranging from those who do not speak English to those with limited English abilities (CPS, 2010).

### *Participating schools*

Participating schools were drawn from a pool of 483 K-6 and K-8 CPS elementary schools. Several exclusion criteria were applied both to eliminate schools with extreme characteristics and to include schools with higher risks of problem behaviours (Figure 1). Using these criteria, 68 schools were eligible. These schools were invited to attend information sessions about the PA programme and the research study. Representatives from 36 schools attended an information session. Eighteen of these schools agreed to participate in the study with the understanding that they would be matched with another suitable school in the pool of 18 and randomly assigned to conditions (see Figure 1 for the matching criteria). Using this process, seven well-matched pairs of schools were created. All 14 schools in these matched pairs were successfully recruited into the study. Baseline equivalence between the programme and control conditions was evident at the school level on the full range of available school and neighbourhood characteristics (Ji, DuBois, Flay, & Brechling, 2008). Schools in the control condition received the PA programme and materials after the end of the trial period.

### *Student participants*

As noted, schools, not students, were the unit of randomisation. All schools were retained throughout the study, but student participation was highly variable across years. The mobility of students (i.e. moving in or out of schools) poses challenges for

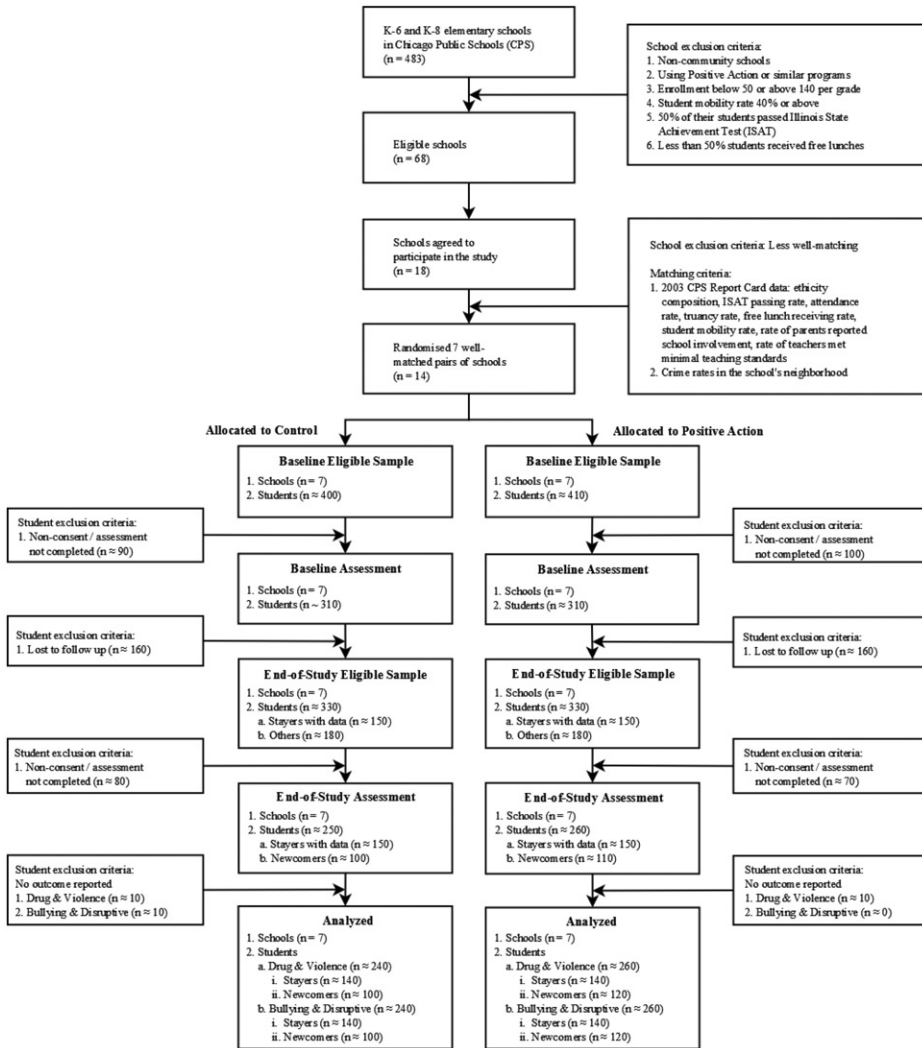


Figure 1. Diagram for participating schools and students.

school-based longitudinal studies. Because the outcome measures at the end of the study were of utmost importance for evaluating programme effectiveness, all students who completed these measures were included in the analyses.

Because some of the data reported herein were restricted-use data (Note 1 online), all *N*s are rounded to the nearest 10. At the end of the 3-year study period, approximately 510 fifth graders completed the questionnaires with the outcome measures. Slightly more than half of these students ( $n \approx 290$ ; 57%) were also part of the original sample of approximately 590 students at baseline. The final sample included about 250 control (49% girls) and 260 intervention (51% girls) students. The highest percentage self-identified their ethnicity as African-American (46%). The remaining ethnic groups included Hispanic (27%), White non-Hispanic (7%), Asian (3%) and other or mixed (17%).

### ***Intervention***

In this trial, programme schools received the K-8 portion of the PA classroom curriculum and school/staff training from the programme developer, plus kits for school preparation, school-wide climate development, counsellors and family classes. The K-8 PA classroom curriculum is scoped and sequenced, consisting of over 140 15-min, age-appropriate lessons per grade that are designed to be taught 4 days per week. More elaborated descriptions of the programme and unit contents can be found in Flay and Allred (2010) and at the programme website ([www.positiveaction.net](http://www.positiveaction.net)). Measures and materials are also available from the first author.

During each school year, the programme developer provided teacher and staff training to each school in the programme condition – approximately 4 h in the first year and 2 h in subsequent years. In addition, each year, we conducted multiple workshops for key players in the programme schools to provide further training on the school-wide components of the programme and to share experiences. We provided support for teachers and staff during the entire study period through individual consultation with a PA Implementation Coordinator. Visits by the Implementation Coordinator assessed the extent to which the programme components were in place, identified potential obstacles, such as staff turnover, which could potentially compromise implementation quality, and provided technical support to help overcome obstacles and enhance programme fidelity. We also attempted to provide regular consultation with principals and other school administrators to ensure that the PA curriculum and other programme components were implemented in adherence to the guidelines established by the programme developer. Finally, to enhance fidelity of implementation with our study cohort students in particular, we held a workshop with teachers of these students during both the second and third years of the study. All except these final workshops are standard supports that schools could purchase from Positive Action, Inc. Incentives were given to the participating schools for administrative support and to the teachers and parents for completing the surveys (Note 2 online).

### ***Measures***

#### *Programme implementation*

All teachers in the intervention schools were asked to complete a Unit Implementation Report at the end of each of the six units of the programme, that is, approximately every 6 weeks. The report was completed on the web and took teachers 5–10 min each time.

#### *Lifetime prevalence of substance use and serious violence-related behaviour*

Researcher-developed survey questions regarding students' lifetime substance use and serious violence-related behaviours (local site items) were not developmentally appropriate for younger students and, therefore, were administered only at the end of the study when students had reached fifth grade. Similar items have been used in previous studies (Beets et al., 2009; US Department of Health and Human Services, 2008). Each item started with a question stem, 'Have you ever', and continued with a specific behaviour descriptor. These included (1) smoked a cigarette, (2) drank

alcohol, (3) gotten drunk on alcohol, (4) used marijuana, (5) used other more serious drugs, (6) carried a knife, (7) threatened to cut or stab someone, (8) cut or stabbed someone on purpose, (9) been asked to join a gang, (10) hung out with gang members and (11) been a member of a gang. Response options were 1 (No), 2 (Yes, once), 3 (Yes, 2–5 times) and 4 (Yes, more than 5 times). Due to the low prevalence rates of the latter three responses, each item was dichotomised (0/1) to represent ‘No, never’ *versus* ‘ever’. We then created a count variable for both substance use (0–5) and serious violence-related behaviours (0–6), indicating the number of behaviours in which the student reported ever having engaged.

### *Bullying and disruptive behaviours*

Bullying and disruptive behaviours were assessed at each time point of the study using 12 items. Bullying behaviours were measured using six items selected from the Aggression Scale (Orpinas & Frankowski, 2001) in which children responded as to how often in the past 2 weeks they had engaged in verbal or physical aggression at school (e.g. teased others, shoved others). Internal consistency reliability of the scale was high for the present sample ( $\alpha=0.81$  and  $0.88$  at baseline and follow-up, respectively). The Frequency of Delinquent Behavior Scale (Dunford & Elliott, 1984) was used to measure disruptive behaviours, with items modified to refer to the school context. Children were asked to respond to six items about how often in the past couple of weeks they had engaged in different problem behaviours at school (e.g. taking something at school that belonged to others, skipping class). Reliability of this scale also was satisfactory ( $\alpha=0.76$  and  $0.81$  at baseline and follow-up, respectively). Response options were 0 (Never), 1 (Once or twice), 2 (A few times) and 3 (Many times) for all items. To facilitate a focus on relatively recurrent behaviours as well as comparison with the measures of substance use and serious violence-related behaviour, the former three response options were combined to create a dichotomous scale (0 = None to a few times; 1 = Many times). Items were then summed within domain to create separate count variables for bullying and disruptive behaviours (0–6).

### *Covariates*

To reduce potential confounding effects and increase statistical precision, self-reported age, gender, ethnicity and measures of baseline problem behaviours were included in analyses as control variables. For baseline problem behaviour, a mean score was computed by averaging the 12 bullying and disruptive behaviour items using the original response options (0–3). We conducted sensitivity analyses using different ways of controlling for baseline problem behaviour (e.g. one *vs.* two variables, continuous *vs.* count variables); results in all instances were consistent with those with report below.

### *Analysis*

Statistical analyses consisted of five steps. In step 1, multiple imputations were created to handle the missing values using switching regression (Royston, 2004) implemented in Stata (StataCorp, 2007). Multiple imputation produces less biased estimates than traditional approaches to handling missing data such as list-wise

deletion or mean substitution when data are not missing completely at random (Acock, 2005). Rather, multiple imputation assumes data are missing at random and appropriate mechanisms are included as covariates in the imputation model. An improved strategy for analysing multiply imputed data called ‘multiple imputation, then deletion (MID)’ (von Hippel, 2007) was adopted in which all cases were used for imputations but cases with imputed outcome variables were deleted from the analyses. In situations where missing values on the outcome variables are substantial and the imputation model for the outcome variables is difficult to specify, as in this study, MID is preferable. The extra information produced from cases with missing outcomes is useful for imputing missing covariates in other cases, but cases with imputed outcomes have nothing but random variation to contribute to the subsequent analysis (von Hippel, 2007).

For this study, the imputation model included students’ age, sex, ethnicity, school affiliation, baseline problem behaviours, problem behaviours at waves 2, 3 and 4 (computed the same as baseline problem behaviour) and an outcome composite score computed by averaging all 23 dichotomised items that comprised the four outcome behaviours at the final wave. A single composite outcome score was used rather than multiple scale-level or item-level scores to reduce missing data that needed to be handled by the algorithm. Results of sensitivity analyses showed that various ways of representing the outcomes in the imputation models did not alter the magnitude or significance of estimated programme effects. As suggested by Rubin (1987), five imputation datasets were generated. Using the MID procedures (von Hippel, 2007), all of the cases with imputed outcomes at grade 5 were deleted after the multiple imputation. Students who joined the schools after the beginning of the study (i.e. newcomers) accounted for the highest proportion of the imputed values on baseline problem behaviours, which was equivalent to 42% of the cases in the analysed data ( $n \approx 510$ ). All of the subsequent analyses were conducted using these imputed datasets. Parameter estimates were combined across the imputations using Rubin’s rules (Rubin, 1987).

In step 2, we tested the equivalence of our student cohort between PA schools and control schools. We have previously reported equivalence at the school level (Ji et al., 2008); however, because equivalence at the school level does not guarantee equivalence at the individual level, multilevel analyses were conducted to examine whether cohort (grade 3) students in the PA schools were different from students in the control schools on demographics and baseline problem behaviours.

In step 3, potential differences between stayers (i.e. students present at baseline and at the end of the study), dropouts (i.e. students present at baseline but not at the end of the study) and newcomers (i.e. students not in the study at baseline but present at the end) were examined on baseline problem behaviour and end-of-study problem behaviours. These analyses served to explore the possibility of alternative explanations due to the mobility of students (Hansen, Collins, Malotte, Johnson, & Fielding, 1985). To test whether stayers were the same as dropouts, multilevel regressions of baseline problem behaviour on stayer versus dropout group membership were conducted. Stayers and newcomers were compared on the four outcome measures at the final wave using multilevel overdispersed Poisson models. For the analyses of stayers and newcomers, students in the PA schools were not included, as stayers in the PA schools were expected to be different than the newcomers due to the intervention.



In step 4 of the analysis, a series of multilevel Poisson models, each of which included an overdispersion estimate (explained below), was conducted to examine the effects of the PA programme on problem behaviours; analyses were conducted. A successful matched-pair design (i.e. where variation between pairs far exceeds the variation within pairs) can enhance statistical power by matching units that are likely to be similar on the outcomes into pairs (Raudenbush, Martinez, & Spybrook, 2007). In this study, problem behaviours were tested using three-level models with students (Level 1) nested within schools (Level 2) and nested within pairs (Level 3). Introducing a third-level random effect partitioned the between-pair variation from the within-pair variation; hence, intervention effects could be tested with greater precision given that the pairs were well matched (Feng, Diehr, Peterson, & McLerran, 2001). Due to the relatively low frequency of problem behaviours, the variances of the outcome scales were larger than the mean; therefore, analyses testing for overdispersion were conducted (Long & Freese, 2006). A random effect at the student-level was introduced into the Poisson models accounting for overdispersion (Rabe-Hesketh & Skrondal, 2005).

The significance of the programme effects in the multilevel models computed in the statistical package, Stata, was tested against a standard normal distribution, which assumes a sufficiently large number of schools. Because only 14 schools were included, a set of sensitivity analyses using an adjusted df (12) was conducted to provide more conservative tests of the programme effects for each outcome behaviour (Snijders & Bosker, 1999). Because the tests of significance using an adjusted df could not be combined using Rubin's rules (Rubin, 1987), the ranges of the  $p$ -values for each outcome are reported.

In step 5, exploratory analyses were conducted comparing the programme effects for stayers and newcomers. Because stayers received more years of intervention compared to newcomers, results of the analyses were relevant to the question of whether programme effects depended on degree of programme exposure (i.e. a dose-response relationship). Like the analyses for the programme effects, multilevel overdispersed Poisson models were used. A binary variable representing the membership in the stayer (0) or newcomer group (1) and an interaction between this variable and intervention condition was included in the models. Sensitivity analyses using an adjusted df (12) were conducted.

## Results

### *Baseline equivalence*

Sample characteristics at baseline are presented online (Table S1). Gender composition and age at baseline were not significantly different between the programme and the control conditions, whereas differences on ethnicity composition were significant.  $\chi^2(4) = 10.24, p < 0.05$ . There were more African-American students and fewer students in the other/mixed ethnicity group for the control condition compared to the PA condition. Controlling for demographic variables and clustering of students, PA students were not significantly different from control students in the reported rates of problem behaviours,  $B = 0.13$ , ns (including stayers and newcomers;  $n \approx 510$ ).

### *Stayers/movers equivalence*

Across the multiple imputations, no significant difference on baseline problem behaviours was found between stayers and dropouts,  $B = -0.02$ , ns. Comparisons between stayers and newcomers in the control schools on the behavioural outcomes also showed non-significant differences (incidence rate ratios [IRRs] = 0.73–0.88). These results are in the direction of stayers endorsing fewer items on problem behaviours than dropouts and newcomers, although as noted the differences were not statistically significant.

### *Descriptive statistics at the final wave*

Consistent with the sample characteristics at baseline, differences on ethnicity were found between the programme and the control groups,  $\chi^2(4) = 11.64$ ,  $p < 0.05$ , at grade 5. The count frequency of the problem behaviours showed the clear patterns of long-tailed distributions. At fifth grade, about 35% of the students reporting having used at least one substance, and a similar proportion of students endorsed at least one violent behaviour. About 30% and 18% of students reported currently engaging in at least one bullying behaviour and disruptive behaviour, respectively.

### *Programme implementation*

There was variability between schools in all of the implementation indices (i.e. programme benchmarks), especially in first year, with improvements over time (Table S2 online). By the end of third year, one school was still implementing at a low level (meeting programme benchmarks at a level below 50%, on average, across all benchmarks), four at a moderate level (meeting programme benchmarks at levels between 50% and 60%) and two at moderate to high levels (meeting programme benchmarks at a level between 60% and 70%).

### *Outcome analyses*

Overdispersed Poisson models significantly improved the model fit compared with Poisson models without overdispersion,  $p < 0.001$ , for all likelihood ratio tests. Results of primary outcome analyses, as given in Table 1, showed students in the PA schools endorsed significantly fewer items for substance use (IRR = 0.69 [0.50–0.97]), serious violence (IRR = 0.63 [0.45–0.88]) and bullying behaviours (IRR = 0.59 [0.37–0.92]). PA students also reported engaging in fewer disruptive behaviours, IRR = 0.73 (0.43–1.24),  $p = 0.24$ , although this was not statistically significant. When an adjusted df (12) was used for testing programme effects, the ranges of the  $p$ -values for the programme effects on the count outcomes across imputation datasets were 0.05–0.06 for substance use, 0.02 for violence, 0.03–0.05 for bullying, and 0.24–0.31 for disruptive behaviours. The positive programme effects can be translated into 31% reduction in substance use behaviour, 36% reduction in violence behaviour, 41% reduction bullying behaviours and 27% (not significant) reduction in disruptive behaviours.

Table 1. Estimates of three-level overdispersed Poisson models ( $n \approx 500$ , for substance use and serious violence-related behaviour;  $n \approx 500$ , for bullying behaviour and disruptive behaviour).

	Substance use		Serious violence-related behaviour		Bullying behaviour		Disruptive behaviour	
	IRR <sup>a</sup>	(95% CI)	IRR	(95% CI)	IRR	(95% CI)	IRR	(95% CI)
<b>Fixed effects</b>								
Problem behaviour at baseline	1.46	(1.01–2.09)	1.54	(1.00–2.39)	1.65	(0.90–3.03)	2.18	(1.02–4.65)
Gender (0 = girls; 1 = boys)	1.75	(1.31–2.34)	2.02	(1.47–2.79)	2.14	(1.44–3.16)	2.35	(1.40–3.93)
Age at baseline (years)	1.46	(1.14–1.88)	1.75	(1.33–2.31)	1.56	(1.12–2.18)	1.99	(1.30–3.05)
Ethnic groups (reference: Black)	–	–	–	–	–	–	–	–
White	0.79	(0.36–1.71)	0.44	(0.18–1.10)	0.25	(0.08–0.75)	0.25	(0.06–1.08)
Hispanic	1.20	(0.78–1.86)	0.77	(0.46–1.31)	0.36	(0.19–0.67)	0.70	(0.36–1.36)
Asian	0.79	(0.28–2.22)	0.60	(0.19–1.86)	0.32	(0.08–1.29)	0.20	(0.02–2.11)
Other	1.39	(0.91–2.14)	1.16	(0.72–1.87)	1.05	(0.61–1.83)	1.15	(0.59–2.24)
Positive Action vs. control	<b>0.69</b>	(0.50–0.97)	<b>0.63</b>	(0.45–0.88)	<b>0.59</b>	(0.37–0.92)	0.73	(0.43–1.24)
Intercept	0.01	(0.00–0.10)	0.00	(0.00–0.03)	0.01	(0.00–0.11)	0.00	(0.00–0.01)
<b>Random effects</b>	Est.	SE	Est.	SE	Est.	SE	Est.	SE
Pair-level ( $n_{\text{pair}} = 7$ )	0.27	0.14	0.35	0.14	0.21	0.23	0.04	0.63
School-level ( $n_{\text{school}} = 14$ )	0.15	0.17	0.03	0.53	0.21	0.24	0.11	0.64
Individual-level <sup>a</sup>	0.70	0.11	1.00	0.10	1.32	0.12	1.42	0.17

Notes: Est., estimate. Results were derived from five multiply imputed datasets and combined using Rubin's rules (Rubin, 1987). The statistically significant programme effects ( $p < 0.05$ , two-tailed) are given in bold face.

<sup>a</sup>Fixed effects are presented in IRR for better interpretation. Random effects were averaged across the imputations and are presented in the unit of standard deviation.

### *Differential programme effects associated with programme exposure*

For the analyses examining differential programme effects associated with variations in programme exposure, the interaction between group membership (stayers vs. newcomers) and programme conditions was not significant for three of the four outcome measures at follow-up: substance use (IRR = 0.81 [0.45–1.45]), serious violence (IRR = 0.94 [0.50–1.78]) and disruptive behaviours (IRR = 0.43 [0.16–1.15],  $p = 0.09$ ). The interaction was significant for bullying behaviours (IRR = 0.44 [0.21–0.96]), with programme effects stronger for newcomers (IRR = 0.37) compared to stayers (IRR = 0.83). When the  $df = 12$  was used, there were no interaction effects for any of the outcome measures.

### **Discussion**

The Chicago trial was the second study to evaluate the PA programme using a multilevel framework and a matched-pair, randomised-control design. Findings of this study extend the experimental evidence of effectiveness of the PA programme reported for the Hawaii trial (Beets et al., 2009). Specifically, fifth-grade students reported 31%, 36%, 41% and 27% less lifetime substance use, lifetime violence, bullying behaviours and disruptive behaviours after 3 years of PA.

The observed reductions in disruptive behaviours, although in the expected direction, were not statistically significant. Although problem behaviours across diverse areas co-occur and likely share common causes (e.g. Brady et al., 2008), there may be substantial non-overlapping variation for each behaviour (Guilamo-Ramos, Litardo, & Jaccard, 2005). Substance use, violence and bullying are all health compromising behaviours, either to oneself or others, whereas disruptive behaviours are more oriented towards rule violations. Disruptive children over time may become rejected by peers (Ladd, Price, & Hart, 1990), which may lead to distrust of peers and eventual alienation from family and school (Slough & McMahan, 2008). Future development of the PA programme might focus on components to reduce disruptive behaviours more effectively. For instance, the content of the unit on social/emotional PAs for getting along with others includes development of a code of conduct and community service. These ideas could be extended to various classroom and school activities to further encourage good citizenship.

Meta-analyses have shown a mean effect size of 0.24 (SD = 0.30) for school-based substance-use prevention programmes with interactive components (Tobler et al., 2000), effect sizes of 0.20–0.35 for the impact of universal programmes for youth on aggressive/disruptive behaviours (Wilson & Lipsey, 2007) and effect sizes of 0.15–0.31 for universal school-based social and emotional learning programmes on conduct problems (Payton et al., 2008). The effect sizes observed in this study, which range from 0.27 to 0.41, are comparable to the upper end of the effects observed in these meta-analyses. In this and the Hawaii trial (Beets et al., 2009), PA lessons were delivered for 15 min each school day for 3 or 4 years. Students started the programme at a young age and most of them lived in high-poverty neighbourhoods. These factors may partially explain the relatively large effect sizes observed as time-intensive programmes as well as those that focus on younger age and lower socioeconomic status students have been associated with larger effects on problem behaviours in school-based programmes (Tobler et al., 2000; Wilson & Lipsey, 2007). Further discussion comparing this trial and the Hawaii trial can be found online (Note 3).

Results generally indicated that programme effects did not differ significantly by the degree of programme exposure among students in the programme schools. The findings suggest that programme effects extended to those who received less of the intervention. This lack of evidence of a dose-response relationship has at least three possible explanations. First, newcomers, as students entering a new school, might have been more open to and accepting of the programme curriculum. Second, the PA curriculum may have been implemented with greater fidelity over time and, hence, programme effects could have been facilitated for newcomers for this reason. It is possible, too, that effects were hindered to some degree for stayers who might have been habituated to earlier, less effective forms of the programme's implementation. Third, and most likely in our opinion, the programme led to changes in the school-wide climate, including the behaviour of the students and reinforcement of positive behaviours by teachers, that new students adjusted to relatively quickly when they entered the schools where it was being implemented.

A substantial proportion of teachers in this research reported that they did not believe that the PA programme was very or extremely likely to improve academics or student character (Table S2 online). In the Hawaii trial, Beets et al. (2008) found that attitudes towards PA were associated with school climate and teachers' beliefs regarding their responsibility to teach SACD concepts. With direct relevance to these findings, there appeared to be less than ideal levels of higher level administrative support for implementation of programmes such as PA within the school district in this trial. For example, many teachers informally communicated that they felt pressured to comply with district and building directives oriented towards meeting No Child Left Behind (NCLB) standards and as a result did not feel that they had time to implement the programme at expected levels, thus depriving them of the opportunity to potentially develop more favourable impressions of its potential effectiveness. Consequently, academic performance might have been prioritised over SACD as the primary responsibilities of the teachers. In addition, some teachers reported that they already were implementing SACD activities similar to the PA programme. This may have fostered an impression that the programme was limited in what it could offer beyond what they already had been doing.

From this and the Hawaii study, we have come to believe that it takes a much longer period of time for many low-performing schools to fully adopt and implement a comprehensive programme than it did previously and, indeed, that a high level of programme fidelity may be a largely unobtainable goal for many schools. Along with other comprehensive school reform programme developers and researchers (e.g. Fullan, 2001; Gorman et al., 2007), we believe that under recent conditions, many under-performing schools need 3–7 years to fully adopt and implement a comprehensive programme and see substantial benefits from it. In the current trial, most of the teachers (91% by the end of the intervention) could identify the state's and district's curriculum standards as being embedded in the PA lessons. Instead of taking away the instructional time, PA lessons may serve as readily available resources to facilitate the achievement of established educational standards when more fully adopted. Gaining more support for the programme from the higher level administration and developing a school-wide climate that holds teachers more accountable for the SACD of the students may contribute to a higher level of implementation (Beets et al., 2008).

During the course of the intervention, re-training sessions were offered every year to account for teacher turnover as a potential factor detracting from implementation

quality during the trial. Broader questions as to the sustainability of a relatively complex (i.e. multi-component) intervention such as PA once this type of external implementation support has ended are critically important. Clearly, sustained implementation has key implications for the potential of the PA programme to be delivered with acceptable fidelity on a larger, system-wide scale and thus yield substantial benefits for a school system's overall student population on a long-term basis. Investigation of these issues should be a priority in future research.

### ***Limitations***

There are several limitations of this research. First, the study did not have baseline data for about half of the sample. Rather than ignore the baseline data that were present, our strategy of using well-established multiple imputation techniques to estimate baseline values provided an important statistical control that provided a more accurate evaluation of intervention effects. Second, a self-report of negative behaviours was used as a basis for outcome measures. Because most of the negative behaviours would be considered undesirable (indeed some are illegal) for some segments of the sample, and probably socially desirable among the antisocial segments of the sample, it is likely that there was some measurement error. Reports from others (e.g. teachers, parents), from school records for disruptive and violence-related behaviours, and from laboratory tests for substance use would provide more precise measures. Third, as the study included only a relatively small number of schools, it was not possible to incorporate consideration of other school-level predictors, besides the intervention condition, that might influence the negative behaviours, potentially in interaction with the PA programme. Fourth, the mobility of students in this study was relatively high, thus potentially limiting generalisability of our findings to school populations with less mobility. It seems most likely that high levels of mobility serve to attenuate programme effects due to reduced exposure to the intervention among newcomers. Such circumstances, in any event, reflect real-world situations that make longitudinal evaluations of school-based interventions very challenging. Finally, the measures of lifetime prevalence of substance use and serious violence-related behaviour used in this research were context-free. Future studies can explore with whom and where these behaviours take place to further understanding of their situational determinants and correlates.

### **Conclusion**

For better integration of research and practice in public health, Glasgow and Emmons (2007) suggested 'do not expect a program to work perfectly initially, but plan for adaptation and refinement to fit local conditions and emerging issues' (p. 427). Understanding those differences contributes to developing a completely generalisable intervention. More well-designed evaluation of interventions with attention to the degree to which effects generalise across diverse contexts will be very useful in informing the selection and adaptation of effective programmes by decision makers and practitioners and hence closing the gap between research and practice. Findings of this study may be informative for identifying emerging issues for future adaptation and refinement of the programme and others like it.

Despite much previous research, we still do not have prevention or SACD programmes that produce the kind of effects, we would like or that do all that theory suggests is possible. The PA programme is one that comes close, in our estimation, to incorporating most of the factors that current theory and empirical data suggest for improved prevention and SACD. However, despite these characteristics and the positive results reported previously, this trial is only the second randomised trial of the PA programme. Furthermore, the programme's theoretical basis has been developed, at least in part, after the fact, and there are no studies yet that assess the effectiveness of the programme in changing the presumed mediator variables or that systematically investigate gender, risk status and other factors as potential moderators of programme impact. In this study, there was evidence of ample room for improvement in levels of programme implementation. As such, the potential exists for programme effects to continue to improve in schools using PA for several years due not only to greater levels of student exposure, but also gains in the level and integrity of implementation.

Replication studies such as this can help identify potential moderating environmental or contextual factors in successful programmes; however, empirically testing these moderating factors requires large-scale studies with multiple sites of intervention. We agree with Glasgow, Lichtenstein, and Marcus (2003) that increasing funding on the external validity evidence of intervention programmes could play an important role in narrowing the research–practice gap.

### **Acknowledgements**

The findings reported here are based on research funded by the Institute of Education Sciences (IES), US Department of Education, under cooperative agreement R215S020218 as part of the SACD Research Programme. The SACD Research Programme is a collaboration among IES, the CDC's Division of Violence Prevention, Mathematica Policy Research Inc. (MPR) and awardees of SACD cooperative agreements (Children's Institute, New York University, Oregon State University, University at Buffalo SUNY, University of Maryland, University of North Carolina Chapel Hill and Vanderbilt University).

The SACD research programme includes a multi programme evaluation data collected by MPR and complementary research study data collected by each grantee. The findings reported here are based partly on the Chicago portion of the multi programme data and the complementary research data collected by the University of Illinois and OSU (Brian Flay, Principal Investigator) under the SACD programme. The inclusion of the (Chicago portion of) the multi site data, which are restricted use data licensed from the US Department of Education, requires that all *N*s be rounded to the nearest 10. These findings may differ from the results reported for the SACD multi programme evaluation. The findings presented in this article may be based on different sample sizes of children, classrooms and teachers, outcome measures and/or analytic strategies seeking to answer different research questions.

Brian Flay and David DuBois conceived the study and obtained funding, David DuBois and UIC staff oversaw programme implementation, the programme developer (Carol G. Allred) provided teacher/staff training, UIC and MPR staff collected all data, Brian Flay and OSU co investigators and staff conducted data analysis, Kin Kit Li wrote the first draft of the article and all investigators and staff participated in article revisions.

The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the IES, CDC, MPR or every Consortium member, nor does mention of trade names, commercial products or organisations imply endorsement by the US Government.

We are extremely grateful to the participating CPS schools, their principals, teachers, students and parents. We also thank the CPS Research Review Board and Office of

Specialised Services, especially Dr Renee Grant Mitchell and Dr Inez Drummond, for their invaluable support of this research.

**Declaration of interest:** The research described herein was done using the programme, the training and technical support of Positive Action, Inc. Investigators and all research activities were supported financially by the IES research grant. Dr Flay's spouse holds a significant financial interest in Positive Action, Inc. Issues regarding conflict of interest were reported to the relevant institutions and appropriately managed following the institutional guidelines.

## References

- Acock, A.C. (2005). Working with missing values. *Journal of Marriage and Family*, *67*, 1012–1028.
- Ajzen, I. (1991). The theory of planned behavior. *Organization Behavior and Human Decision Processes*, *50*, 179–211.
- Andrews, J.A., Hampson, S.E., Barckley, M., Gerrard, M., & Gibbons, F.X. (2008). The effect of early cognitions on cigarette and alcohol use during adolescence. *Psychology of Addictive Behaviors*, *22*, 96–106.
- Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs, NJ: Prentice Hall.
- Beets, M.W., Flay, B.R., Vuchinich, S., Acock, A.C., Li, K. K., & Allred, C. (2008). School climate and teachers' beliefs and attitudes associated with implementation of the *Positive Action* program: A diffusion of innovations model. *Prevention Science*, *9*, 264–275.
- 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.
- Brady, S.S., Tschann, J.M., Pasch, L.A., Flores, E., & Ozer, E.J. (2008). Violence involvement, substance use, and sexual activity among Mexican American and European American adolescents. *Journal of Adolescent Health*, *43*, 285–295.
- Campbell, C., & Schwarz, D.F. (1996). Prevalence and impact of exposure to interpersonal violence among suburban and urban middle. *Pediatrics*, *98*, 396.
- Chicago Public Schools. (2009). *Illinois standards achievement test over time report*. Retrieved from <http://research.cps.k12.il.us/cps/accountweb/Reports/citywide.html>
- Chicago Public Schools. (2010). *English language learners*. Retrieved from <http://www.cps.edu/Schools/Pages/Englishlanguagelearners.aspx>
- Donovan, J.E. (2007). Really underage drinkers: The epidemiology of children's alcohol use in the United States. *Prevention Science*, *8*, 192–205.
- Dunford, F.W., & Elliott, D.S. (1984). Identifying career offenders using self reported data. *Journal of Research in Crime and Delinquency*, *21*, 57–86.
- Farrell, A.D., Sullivan, T.N., Esposito, L.E., Meyer, A.L., & Valois, R.F. (2005). A latent growth curve analysis of the structure of aggression, drug use, and delinquent behaviors and their interrelations over time in urban and rural adolescents. *Journal of Research on Adolescence*, *15*, 179–204.
- Feng, Z., Diehr, P., Peterson, A., & McLerran, D. (2001). Selected statistical issues in group randomized trials. *Annual Review of Public Health*, *22*, 167–187.
- Flay, B.R., & Allred, C.G. (2003). Long term effects of the Positive Action program. *American Journal of Health Behavior*, *27*, S6–S21.
- Flay, B.R., & Allred, C.G. (2010). The Positive Action Program: Improving academics, behavior and character by teaching comprehensive skills for successful learning and living. In T. Lovat & R. Toomey (Eds.), *International handbook on values education and student well being* (pp. 471–501). Dordrecht, Germany: Springer.



- 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.
- Flay, B.R., Biglan, A., Boruch, R.F., Castro, F.G., Gottfredson, D., Kellam, S., . . . , Ji, P. (2005). Standards of evidence: Criteria for efficacy, effectiveness and dissemination. *Prevention Science, 6*, 151-175.
- Flay, B.R., Snyder, F., & Petraitis, J. (2009). The theory of triadic influence. In R.J. DiClemente, M.C. Kegler, & R.A. Crosby (Eds.), *Emerging theories in health promotion practice and research* (2nd ed., pp. 451-510). San Francisco, CA: Jossey Bass.
- Fullan, M. (2001). *The new meaning of educational change* (3rd ed.). New York: Teachers College Press.
- Glasgow, R.E., & Emmons, K.M. (2007). How can we increase translation of research into practice? Types of evidence needed. *Annual Review of Public Health, 28*, 413-433.
- Glasgow, R.E., Lichtenstein, E., & Marcus, A.C. (2003). Why don't we see more translation of health promotion research to practice? Rethinking the efficacy to effectiveness transition. *American Journal of Public Health, 93*, 1261-1267.
- Glew, G.M., Fan, M. Y., Katon, W., Rivara, F.P., & Kernic, M.A. (2005). Bullying, psychosocial adjustment, and academic performance in elementary school. *Archives of Pediatrics and Adolescent Medicine, 159*, 1026-1031.
- Gorman, G.D., Slavin, R.E., Cheung, A.C.K., Chamberlain, A.M., Madden, N.A., & Chambers, B. (2007). Final reading outcomes of the national randomized field trial of Success for All. *American Educational Research Journal, 44*, 701-731.
- Guilamo Ramos, V., Litardo, H.A., & Jaccard, J. (2005). Prevention programs for reducing adolescent problem behaviors: Implications of the co occurrence of problem behaviors in adolescence. *Journal of Adolescent Health, 36*, 82-86.
- Hahn, R., Fuqua Whitley, D., Wethington, H., Lowy, J., Crosby, A., Fullilove, M., . . . , Dahlberg, L. (2007). Effectiveness of universal school based programs to prevent violent and aggressive behavior: A systematic review. *American Journal of Preventive Medicine, 33*, S114-S129.
- Hansen, W.B., Collins, L.M., Malotte, C.K., Johnson, C.A., & Fielding, J.E. (1985). Attrition in prevention research. *Journal of Behavioral Medicine, 8*, 261-275.
- Hawkins, J.D., Kosterman, R., Catalano, R.F., Hill, K.G., & Abbott, R.D. (2005). Promoting positive adult functioning through social development intervention in childhood: Long term effects from the Seattle Social Development Project. *Archives of Pediatrics and Adolescent Medicine, 159*, 25-31.
- Hawkins, J.D., & Weis, J.G. (1985). The social development model: An integrated approach to delinquency prevention. *Journal of Primary Prevention, 6*, 73-97.
- Ji, P., DuBois, D.L., Flay, B.R., & Brechling, V. (2008). "Congratulations, you have been randomized into the control group!": Issues to consider when recruiting schools for matched pair randomized control trials of prevention programs. *Journal of School Health, 78*, 131-139.
- Ji, P., Flay, B.R., DuBois, D.L., Brechling, V., Day, J., & Cantillon, D. (2006). Consent form return rates for third grade urban elementary students. *American Journal of Health Behavior, 30*, 467-474.
- Johnston, L.D., O'Malley, P.M., Bachman, J.G., & Schulenberg, J.E. (2008). *Monitoring the future national results on adolescent drug use: Overview of key findings, 2007*. Bethesda, MD: National Institute on Drug Abuse.
- Krug, E.G., Dahlberg, L.L., Mercy, J.A., Zwi, A.B., & Lozano, R. (2002). *World report on violence and health*. Geneva: World Health Organization.
- Ladd, G.W., Price, J.M., & Hart, C.H. (1990). Preschoolers' behavioral orientations and patterns of peer contact: Predictive of peer status? In S.R. Asher & J.D. Coie (Eds.), *Peer rejection in childhood* (pp. 90-115). Cambridge, UK: Cambridge University Press.

- Long, J.S., & Freese, J. (2006). *Regression models for categorical dependent variables using Stata*. College Station, TX: Stata Press.
- Morales, R., & Bonilla, F. (1993). *Latinos in a changing U.S. economy: Comparative perspectives on growing inequality*. Newbury Park, CA: Sage.
- Nation, M., Crusto, C., Wandersman, A., Kumpfer, K.L., Seybolt, D., Morrissey Kane, E., . . . , Davino, K. (2003). What works in prevention: Principles of effective prevention programs. *American Psychologist*, 58, 449-456.
- Noguera, P.A. (2008). *The trouble with Black boys*. San Francisco, CA: Jossey Bass.
- Orpinas, P., & Frankowski, R. (2001). The Aggression Scale: A self report measure of aggressive behavior for young adolescents. *Journal of Early Adolescence*, 20, 50-67.
- Payton, J., Weissberg, R.P., Durlak, J.A., Dymnicki, A.B., Taylor, R.D., Schellinger, K.B., & Pachan, M. (2008). *The positive impact of social and emotional learning for kindergarten to eighth grade students: Findings from three scientific reviews*. Chicago, IL: Collaborative for Academic, Social, and Emotional Learning.
- Rabe Hesketh, S., & Skrondal, A. (2005). *Multilevel and longitudinal modeling using Stata*. College Station, TX: StataCorp LP.
- Raudenbush, S.W., Martinez, A., & Spybrook, J. (2007). Strategies for improving precision in group randomized experiments. *Educational Evaluation and Policy Analysis*, 29, 5-29.
- Royston, P. (2004). Multiple imputation of missing values. *The Stata Journal*, 4, 227-241.
- Rubin, D.B. (1987). *Multiple imputation for nonresponse in surveys*. New York: Wiley.
- Schulenberg, J.E., & Maggs, J.L. (2008). Destiny matters: Distal developmental influences on adult alcohol use and abuse. *Addiction*, 103, 1-6.
- Singer, M. (1994). Aids and the health crisis of the U.S. urban poor: The perspective of critical medical anthropology. *Social Science and Medicine*, 39, 931-948.
- Slough, N.M., & McMahon, R.J. (2008). Preventing serious conduct problems in school age youth: The Fast Track program. *Cognitive and Behavioral Practice*, 15, 3-17.
- Snijders, T.A.B., & Bosker, R.J. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. Thousand Oaks, CA: Sage.
- Social and Character Development Research Consortium (2010). *Efficacy of schoolwide programs to promote social and character development and reduce problem behavior in elementary school children: Social and character development multi program evaluation report*. Washington, DC: Institute of Education Sciences, National Center for Education Research, US Department of Education.
- StataCorp. (2007). *Stata statistical software: Release 10*. College Station, TX: StataCorp LP.
- Tang, J., & Sable, J. (2009). Characteristics of the 100 largest public elementary and secondary school districts in the United States: 2006-07 (NCES 2009-342). Washington, DC: National Center for Education Statistics, Institute of Education Sciences, US Department of Education.
- Thornton, T.N., Craft, C.A., Dahlberg, L.L., Lynch, B.S., & Baer, K. (2000). *Best practices of youth violence prevention: A sourcebook for community action*. Atlanta, GA: Centers for Disease Control and Prevention.
- Tobler, N.S., Roona, M.R., Ochshorn, P., Marshall, D.G., Streke, A.V., & Stackpole, K.M. (2000). School based adolescent drug prevention programs: 1998 meta analysis. *The Journal of Primary Prevention*, 20, 275-336.
- US Department of Health and Human Services. (2008). *YRBSS: Youth Risk Behavior Surveillance System*. Retrieved from <http://www.cdc.gov/HealthyYouth/yrbs/>
- von Hippel, P.T. (2007). Regression with missing Ys: An improved strategy for analyzing multiply imputed data. *Sociological Methodology*, 37, 83-117.
- Wilson, S.J., & Lipsey, M.W. (2007). School based interventions for aggressive and disruptive behavior: Update of a meta analysis. *American Journal of Preventive Medicine*, 33, S130-S143.

### *Supplementary Online Materials*

Materials presented in this document are supplementary to the article titled “*Effects of the Positive Action Program on Problem Behaviors in Elementary School Students: A Matched-Pair Randomized Control Trial in Chicago.*” Materials are organized into the following notes regarding (a) the reporting of sample sizes, (b) incentives, and (c) the comparison between the impact of the program in the Chicago trial and the Hawaii trial, with the relevant references attached. Two supplementary tables describing the sample characteristics and the teachers’ report of meeting the implementation benchmarks are also presented.

#### *1. Reporting of Sample Sizes*

The Social and Character Development (SACD) research program includes a multi-program evaluation data collected by MPR and complementary research study data collected by each grantee. The findings reported here are based partly on the Chicago portion of the multi-program data and the complementary research data collected by the University of Illinois and Oregon State University (Brian Flay, Principal Investigator) under the SACD program. The inclusion of the (Chicago portion of) the multi-site data, which are restricted-use data licensed from the US Department of Education, requires that all *N*s be rounded to the nearest 10. These findings may differ from the results reported for the SACD multi-program evaluation. The findings presented in this manuscript may be based on different sample sizes of children, classrooms, and teachers, outcome measures, and/or analytic strategies seeking to answer different research questions.

#### *2. Incentives*

All of the participating schools received the PA program at no cost. Schools in the program condition received \$4,000 per year for three years to support a part-time PA coordinator and \$1,000 per year to support a PA committee, whereas controls school received \$1,000 per year in unrestricted funds. Incentives were also provided to teachers and parents for completing the surveys (about \$4,000 per school each year).

#### *3. Comparison between the Program Impact in the Chicago and the Hawaii Trials*

Although measures and data analytic procedures were not exactly comparable between the two trials, the effect sizes of the Hawaii trial (IRR of 0.41 for substance use and IRR of 0.42 for violent behaviors) were somewhat larger in magnitude than those reported here for the Chicago trial. One possible explanation for this difference is that PA had been implemented for four years in Hawaii compared with three years in Chicago. In addition, the social and physical environments in which the two trials took place may have exerted different challenges and opportunities for students. For instance, a study found that youth in Chicago attending schools with 20% or more Hispanic students were exposed to more alcohol advertising (Pasch, Komro, Perry, Hearst, & Farbaksh, 2008). Witnessing community violence was also found to be associated with childhood aggression through imitation and changes in associated cognitions in Chicago (Guerra, Huesmann, & Spindler, 2003). From a social-ecological perspective (McLeroy, Bibeau, Steckler, & Glanz, 1988), it takes multiple levels of influence to change behaviors (Flay, Snyder, & Petraitis, 2009). This underscores a need for programs to not only develop children’s characters and skills in dealing with various hazardous situations, but also foster community efforts to change the environment (e.g., enhancing constructive school climates, encouraging parent involvement). Although community-level change activities are available for the PA program, they were not implemented in the Chicago trial due to resource constraints.

*School-level evidence for the match-paired randomized control trial in Hawaii.* In addition to the current findings using a multilevel approach to examine changes in students’ problem behaviors within the larger school context, analyses of school-level archival data from the Hawaii trial has shown that schools offering the PA program improved their standardized reading and math scores over time compared with control schools (Snyder et al., 2010). School-level, archival indicators of absenteeism, suspensions, and retentions also differentially improved in PA schools with moderate to large effect sizes.

Table S1 *Sample characteristics at baseline and the final wave*

	Baseline		Final Wave	
	Control	PA	Control	PA
Number of schools	7	7	7	7
Number of students (total)	≈310	≈310	≈250	≈260
Stayers (n ≈ 300)	≈150	≈150	≈150	≈150
Dropouts (n ≈ 320)	≈160	≈160	--	--
Newcomers (n ≈ 210)	--	--	≈100	≈110
Gender (percent of boys)	44.95%	48.24%	42.97%	50.76%
Age at baseline / SD (years)	8.71 / .53	8.66 / .59	8.68 / .51	8.74 / .62
Race a				
White	7.54%	6.09%	6.85%	6.95%
African American	48.20%	40.71%	48.79%	44.02%
Hispanic American	23.61%	26.60%	29.84%	24.32%
Asian American	3.28%	1.28%	3.23%	2.32%
Other or mixed ethnicity	17.38%	25.32%	11.29%	22.39%
Baseline problem behaviors / SD b				
All baseline students (n ≈ 590)	.31 / .45	.38 / .51		
Stayers only (n ≈ 290)	.23 / .37	.38 / .52		
All final wave students (n ≈ 510)			.26 / .46	.38 / .49
Number of items endorsed c				
Substance use – 0	≈ 140	(58.33%)	≈ 170	(65.38%)
Substance use – 1	≈ 60	(25.00%)	≈ 60	(23.08%)
Substance use – 2	≈ 20	(8.33%)	≈ 20	(7.69%)
Substance use – 3	≈10	(4.17%)	≈ 10	(3.85%)
Substance use – 4	≈ 10	(4.17%)	≈ 0	(0.00%)
Substance use – 5	≈ 0	(0.00%)	≈ 0	(0.00%)
Serious violence – 0	≈ 150	(62.50%)	≈ 170	(68.00%)
Serious violence – 1	≈ 40	(16.67%)	≈ 40	(16.00%)
Serious violence – 2	≈ 20	(8.33%)	≈ 10	(4.00%)
Serious violence – 3	≈ 20	(8.33%)	≈ 10	(4.00%)
Serious violence – 4	≈ 10	(4.17%)	≈ 10	(4.00%)
Serious violence – 5	≈ 0	(0.00%)	≈ 10	(4.00%)
Serious violence – 6	≈ 0	(0.00%)	≈ 0	(0.00%)
Bullying behavior – 0	≈ 160	(66.67%)	≈ 190	(70.37%)
Bullying behavior – 1	≈ 30	(12.50%)	≈ 30	(11.11%)
Bullying behavior – 2	≈ 20	(8.33%)	≈ 20	(7.41%)
Bullying behavior – 3	≈ 20	(8.33%)	≈ 10	(3.70%)
Bullying behavior – 4	≈ 10	(4.17%)	≈ 10	(3.70%)
Bullying behavior – 5	≈ 0	(0.00%)	≈ 0	(0.00%)
Bullying behavior – 6	≈ 0	(0.00%)	≈ 10	(3.70%)
Disruptive behavior – 0	≈ 200	(80.00%)	≈ 210	(84.00%)
Disruptive behavior – 1	≈ 20	(8.00%)	≈ 30	(12.00%)
Disruptive behavior – 2	≈ 10	(4.00%)	≈ 10	(4.00%)
Disruptive behavior – 3	≈ 10	(4.00%)	≈ 0	(0.00%)
Disruptive behavior – 4	≈ 0	(0.00%)	≈ 0	(0.00%)
Disruptive behavior – 5	≈ 10	(4.00%)	≈ 0	(0.00%)
Disruptive behavior – 6	≈ 0	(0.00%)	≈ 0	(0.00%)

*Note:* PA=Positive Action. Stayer = Student who had data at baseline and the final wave. Dropout = Student who had data at baseline but not at the final wave. Newcomer = Student who had data at the final wave but not baseline. Because some of the data reported herein were restricted-use data from the US

Department of Education (see Note 1 above), all *N*s are rounded to the nearest 10.

<sup>a</sup> For demographic variables, only race compositions were different between the two groups at baseline ( $\chi^2(4)=10.24, p<.05$ ); this difference was similar at the final wave ( $\chi^2(4)=11.64, p<.05$ ).

<sup>b</sup> Possible range for baseline problem behaviors was 0-3. Three-level regression models (students nested within schools within pairs) were used for comparing baseline problem behaviors between students in PA schools and control schools on various samples. 1) For students who had data at baseline (including  $\approx 590$  students with valid responses on baseline problem behaviors and covariates), baseline problem behaviors was not significantly different between students in PA schools and students in control schools ( $B=.09, ns$ ). 2) For stayers (including  $\approx 290$  students with valid responses), the difference in baseline problem behaviors was close to significant across conditions ( $B=.15, p=.053$ ). 3) For students at the final wave, baseline problem behaviors was not significantly different across conditions ( $B=.13, ns$ ). Results in 3) were derived from 5 multiply imputed datasets because newcomers did not have data on baseline problem behaviors.

<sup>c</sup> Substance use, serious violence, bullying and disruptive behaviors are presented in the unit of number of items endorsed by the students (i.e., same as the outcome measures in the analyses). Possible ranges for numbers of behaviors endorsed were 0-5 for substance use and 0-6 for the other three count variables.

Table S2 *Percentage of teachers (all teachers, cohort or not) meeting implementation benchmarks on unit implementation reports by year of program and their beliefs about the program*

	2004-05	2005-06	2006-07
<b>Program Benchmarks</b>			
1. Teaching at least 4 lessons per week	61%	68%	66%
2. Distribute 5 or more Word of the Week Cards/week	22%	35%	35%
3. Distribute 5 or more PA Stickers/week	30%	39%	41%
4. Read 5 or more notes from ICU Box	42%	48%	49%
5. Play PA Music 2 or more days/week	19%	35%	33%
6. Spoke with 2 or more Parents about PA per week	25%	45%	43%
7. Identified District's or State's Core Curriculum	81%	93%	91%
<b>Standards in PA lessons</b>			
8. Whether PA assemblies were delivered in school	17%	47%	47%
<b>Teachers' Beliefs</b>			
9. Teacher believes s/he delivered program quite well or very well	64%	71%	71%
10. Teacher believes continued use of PA is very or extremely likely to improve student character	61%	63%	68%
11. Teacher believes continued use of PA is very or extremely likely to improve student academics	49%	53%	58%
<b>Teacher Response Rates</b>	<b>53%</b>	<b>75%</b>	<b>76%</b>

## References

- Flay, B. R., Snyder, F., & Petraitis, J. (2009). The Theory of Triadic Influence. In R. J. DiClemente, M. C. Kegler & R. A. Crosby (Eds.), *Emerging theories in health promotion practice and research* (2nd ed., pp. 451-510). San Francisco, CA: Jossey-Bass.
- Guerra, N. G., Huesmann, L. R., & Spindler, A. (2003). Community violence exposure, social cognition, and aggression among urban elementary school children. *Child Development, 74*, 1561-1576.
- McLeroy, K. R., Bibeau, D., Steckler, A., & Glanz, K. (1988). An ecological perspective on health promotion programs. *Health Education and Behavior, 15*, 351-377.
- Pasch, K. E., Komro, K. A., Perry, C. L., Hearst, M. O., & Farbakhsh, K. (2008). Does outdoor alcohol advertising around elementary schools vary by the ethnicity of students in the school? *Ethnicity and Health, in press*.
- Snyder, F., Flay, B., Vuchinich, S., Acock, A., Washburn, I., Beets, M., et al. (2010). Impact of a social-emotional and character development program on school-level indicators of academic achievement, absenteeism, and disciplinary outcomes: A matched-pair, cluster-randomized, controlled trial. *Journal of Research on Educational Effectiveness, 3*, 26-55.