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

Frank Snyder and Brian Flay
Department of Public Health, Oregon State University, Corvallis, Oregon, USA

Samuel Vuchinich, Alan Acock, and Isaac Washburn
Human Development and Family Studies, Oregon State University, Corvallis, Oregon, USA

Michael Beets
Department of Exercise Science, University of South Carolina, Columbia, South Carolina, USA

Kin-Kit Li
Department of Applied Social Studies, City University of Hong Kong, Hong Kong

Abstract: This article reports the effects of a comprehensive elementary school-based social-emotional and character education program on school-level achievement, absenteeism, and disciplinary outcomes utilizing a matched-pair, cluster-randomized, controlled design. The Positive Action Hawai‘i trial included 20 racially/ethnically diverse schools (M enrollment = 544) and was conducted from the 2002–03 through the 2005–06 academic years. Using school-level archival data, analyses comparing change from baseline (2002) to 1-year posttrial (2007) revealed that intervention schools scored 9.8% better on the TerraNova (2nd ed.) test for reading and 8.8% on math, that 20.7% better in Hawai‘i Content and Performance Standards scores for reading and 51.4% better in math, and that intervention schools reported 15.2% lower absenteeism and fewer suspensions (72.6%) and retentions (72.7%). Overall, effect sizes were moderate to large (range = 0.5–1.1) for all of the examined outcomes. Sensitivity analyses using permutation models and random-intercept growth curve models substantiated results. The results provide evidence that a comprehensive school-based program, specifically...
developed to target student behavior and character, can positively influence school-level achievement, attendance, and disciplinary outcomes concurrently.

**Keywords:** Randomized experiment, matched-pair, academic, achievement, discipline, social and character development

**INTRODUCTION**

Education has an urgent need to learn more about the role of behavior, social skills, and character in improving academic achievement (Eccles, 2004; Meece, Anderman, & Anderman, 2006). Since the No Child Left Behind Act passed, education has been focused on teaching to core content standards to improve academic achievement scores, particularly in reading and mathematics, for which schools are being held accountable (Hamilton et al., 2007). Teaching to, and support for, the behavioral, social, and character domains have been relegated to no or limited dedicated instructional time (Greenberg et al., 2003). Nevertheless, schools are expected to prevent violence, substance use, and other disruptive behaviors that are clearly linked to academic achievement (Fleming et al., 2005; Malecki & Elliott, 2002; Wentzel, 1993). The prevalence of discipline problems, for example, correlates positively with the prevalence of violent crimes within a school (Heaviside, Rowland, Williams, & Farris, 1999) which, in turn, affects attendance and academic achievement (Eaton, Brener, & Kann, 2008; Walberg, Yeh, & Mooney-Paton, 1974). Further, mental health concerns become more prevalent as students move into adolescence and can contribute to behavioral problems that detract from academic achievement (Costello, Mustillo, Erkanli, Keeler, & Angold, 2003). Disciplinary problems (Dinks, Cataldi, & Lin-Kelly, 2007; Eaton, Kann, et al., 2008; Eisenbraun, 2007) and underachievement abound (Coalition for Evidence-Based Policy, 2002; Perie, Moran, & Lurkus, 2005; Snyder, Dillow, & Hoffman, 2008).

To address these needs, numerous school-based programs have been developed to target problems of academic achievement (Slavin & Fashola, 1998; What Works Clearinghouse, n.d.). In addition, many other types of programs have offered the promise of improving academic performance indirectly through a focus on specific problem behaviors, such as substance use and violence (Battistich, Schaps, Watson, Solomon, & Lewis, 2000; Biglan et al., 2004; DuPaul & Stoner, 2004; Elias, Gara, Schuyler, Branden-Muller, & Sayette, 1991; Flay, 1985, 2009a, 2009b; Horowitz & Garber, 2006; Peters & McMahon, 1996; Sussman, Dent, Burton, Stacy, & Flay, 1995; Tolan & Guerra, 1994). Although some of these programs are promising, most are problem specific and tend to address only the microlevel or proximal predictors (e.g., attitudes toward a behavior) of a single problem (e.g., violent behavior; Catalano, Hawkins, Berglund, Pollard, & Arthur, 2002), not the multifaceted ultimate (e.g., safety of neighborhood) and distal (e.g., bonding to parents) factors
that influence many other important outcomes (Flay, 2002; Flay, Snyder, & Petrakis, in press; Petrakis, Flay, & Miller, 1995; Romer, 2003) Consequently, programs have had limited success (Catalano et al., 2002; Flay, 2002).

As practitioners, policymakers, and researchers have implemented programs and sought to raise academic achievement and address negative behaviors among youth, an increasing amount of evidence indicates a relationship among multiple behaviors (Botvin, Griffin, & Nichols, 2006; Botvin, Schinke, & Orlandi, 1995; Catalano, Berglund, Ryan, Lonczak, & Hawkins, 2004; Flay, 2002). Several mechanisms involving multiple behaviors have been identified in improving student behavior and performance (Greenberg et al., 2003; Zins, Weissberg, Wang, & Walberg, 2004). This suggests that key behaviors do not exist in isolation from each other. Moreover, prevention research offers ample empirical support showing that many youth outcomes, negative and positive, are influenced by similar risk and protective factors (Catalano et al., 2004; Catalano et al., 2002; Flay, 2002). That is, most, if not all, behaviors are linked (Flay, 2002). For example, the early initiation of alcohol and cigarette use and/or abuse is associated with lower academic test scores (Fleming et al., 2005). Further, early initiation of substance use and sexual activity can place youth at a greater risk of mental health disorders and aggressive behaviors (Gustavson et al., 2007; Hallfors, Waller, Bauer, Ford, & Halpern, 2005) and continuation of substance use through adolescence and into adulthood (Merline, O’Malley, Schulenber, Bachman, & Johnston, 2004).

Subsequently, there has been a movement toward more integrative and comprehensive programs that address multiple co-occurring behaviors and that involve families and communities. Such programs generally appear to be more effective (Battistich et al., 2000; Catalano et al., 2004; Derzon, Wilson, & Cunningham, 1999; Elias et al., 1991; Flay, 2000; Flay, Graumlich, Segawa, Burns, & Holliday, 2004; Hawkins, Catalano, Kosterman, Abbott, & Hill, 1999; Hawkins, Catalano, & Miller, 1992; Kellam & Anthony, 1998; Lerner, 1995). One of these programs currently being used nationally is the Positive Action (PA) program. PA is a comprehensive schoolwide social-emotional and character development (SACD) program (Flay & Allred, 2003; Flay, Allred, & Ordway, 2001) developed to specifically target the positive development of student behavior and character.

Based on prior studies, PA has been recognized in the character-education report by the U.S. Department of Education’s What Works Clearinghouse (2007) as the only “character education” program in the nation to meet the evidentiary requirements for improving both academics and behavior. Preliminary findings indicate that PA can positively influence school attendance, behavior, and achievement. Two previous quasi-experimental studies utilizing archival school-level data (Flay & Allred, 2003; Flay et al., 2001) reported beneficial effects on student achievement (e.g., math, reading, and science) and serious problem behaviors (e.g., suspensions and violence rates).
The first study (Flay et al., 2001) used School Report Card (SRC) data from two school districts that had used PA within a number of elementary schools for several years in the 1990s. Schools were rank ordered on poverty and mobility, and each PA school was matched with the best matched non-PA school(s) having similar ethnic distribution. Results indicated that PA schools scored significantly better than the non-PA schools in their percentile ranking of fourth-grade achievement scores and reported significantly fewer incidences of violence and lower rates of absenteeism. The second study (Flay & Allred, 2003) used a similar methodological approach but expanded the variables on which PA and non-PA schools were matched to include dependent variables (e.g., reading and math achievement) assessed before the introduction of PA. Results confirmed previous findings and demonstrated that involvement in PA during elementary school improved academic and disciplinary outcomes at both the elementary and secondary levels.

In sum, the prior quasi-experimental studies provide preliminary evidence regarding the effects of PA on academic achievement and disciplinary outcomes. However, these findings are in need of confirmation utilizing a randomized design (Flay, 1986; Flay et al., 2005), a standard considered vital before an intervention is ready for broad dissemination (Flay et al., 2005). Designs that use matching without random assignment leave open the possibility that variables other than those measured were responsible for observed posttest differences, rather than the intervention itself. In addition, the previous quasi-experimental studies lacked data on program implementation, a measurement that is desirable to ensure that implementation occurred and, if so, how well it occurred (Domitrovich & Greenberg, 2000; Durlak & DuPre, 2008; Flay et al., 2005).

Utilizing student self-report data from the current randomized trial, Beets and colleagues (2009) examined the preventive benefits of PA on rates of student self-report and teacher reports of student substance use, violence, and voluntary sexual activity. Results indicated lower rates of substance use, violence and sexual activity among students attending PA schools. Overall, this randomized trial (a) replicated findings from quasi-experimental studies regarding violence and substance use and (b) found that PA can also alter other behaviors, such as sexual activity, that the program does not address directly. Hence, even though PA did not teach sexual responsibility, for example, the SACD content produced effects on sexual activity. Previous results suggest a mechanism that leads PA to positively affect multiple outcomes, such as sexual responsibility and academic achievement, even though the program does not include explicit discussion of these outcomes.

The purpose of the present study was to apply a matched-pair, cluster-randomized, controlled design to evaluate the effects of PA on school-level indicators of academic achievement, absenteeism, and disciplinary outcomes. School-level data are useful for estimating causal effects but are underutilized (Stuart, 2007). The present study builds on extant research and is the first to report the effects of PA on school-level outcomes from a randomized, controlled
design; thus, it provides the most rigorous test yet conducted for whether PA can improve school-level performance and greatly reduces the possibility that factors other than the PA intervention are responsible for observed posttest group differences. PA was hypothesized to result in decreased absenteeism, disciplinary referrals and grade retentions and improved academic achievement.

**METHODS**

**Design and Sample**

The PA Hawai‘i trial was a matched-pair, cluster-randomized, controlled trial, conducted during the 2002–03 through 2005–06 school years, with a 1-year follow-up in 2007, in Hawai‘i elementary schools. The state is one large school district with diverse ethnic groups and a recognized need for improvement (i.e., low standardized test scores and a high percentage of students receiving free or reduced-price lunch). The trial took place in 20 public elementary (K-5 or K-6) schools (10 matched-pairs) on three Hawai‘ian islands. Eligible schools for the study were those elementary schools that (a) were located on O‘ahu, Maui, or Moloka‘i; (b) were K-5 or K-6 community schools (were not academy, charter, or special education); (c) had at least 25% of students receiving free or reduced-price lunch; (d) were in the state’s lower three quartiles of standardized test scores; and (e) had annual student mobility rates under 20%, thereby ensuring that at least 40% of a selected cohort was still in the same school by the end of the trial. To ensure comparability of the intervention and control schools with respect to baseline measures, 2000 SRC data on 111 eligible schools were used to stratify schools into strata ranked on an index based on (a) demographic variables of percentage free or reduced-price lunch, school size, percentage stability, and ethnic distribution; (b) characteristics of the student populations such as percentage special education, and limited English proficiency; and (c) indicators of student behavior and performance outcomes such as standardized test scores, absenteeism, and suspensions (Dent, Sussman, & Flay, 1993; Flay et al., 2004; Graham, Flay, Johnson, Hansen, & Collins, 1984). Schools were matched based on their index score, resulting in 19 utilizable strata. Matched pairs were randomly selected from within strata, with one school of each pair randomly assigned to either the intervention or control condition before recruitment.

Starting with schools only on O‘ahu (to limit travel costs), intervention schools were asked to implement PA, whereas the control schools were asked to continue “business as usual” without making any substantial SACD reforms. Once it was evident that no additional schools could be recruited on O‘ahu, recruitment began using strata from Maui and Moloka‘i. The final sample of schools was representative of Hawai‘ian schools, though with higher stability
Impact of the Positive Action Program 31

(as intended) and at higher risk (as intended) as indicated by percent free or reduced-price lunch and standardized test scores, respectively.

Intervention schools were offered the complete PA program free of charge and control schools were offered a monetary incentive during the randomized trial and the PA program upon completion of the trial. Three of the 10 control schools chose to receive the PA program after the formal trial; they were treated as controls at the follow-up to the present study, as anecdotal evidence suggests that they did not fully implement the program, and it is likely that schools need several years to fully implement a comprehensive program to see substantial benefits (Beets et al., 2009; Li et al., 2009).

Program Overview

The Positive Action program (http://www.positiveaction.net) is a comprehensive, schoolwide SACD program designed to improve academics, student behaviors, and character (Flay & Allred, in press). The program, developed in 1977 by Carol Gerber Allred, Ph.D., and revised since then as a result of process and outcome evaluations, is grounded in a broad theory of self-concept (Purkey, 1970; Purkey & Novak, 1970); is consistent with integrative, ecological theories of health behavior such as the Theory of Triadic Influence (Flay & Petraitis, 1994; Flay et al., in press), and is described in detail elsewhere (Flay & Allred, 2003; Flay et al., 2001). The full PA program consists of K-12 classroom curricula, of which only the elementary curriculum was used in the present randomized trial; a schoolwide climate development component, including teacher/staff training by the developer, a PA coordinator’s (principal’s) manual, school counselor’s program, and PA coordinator/committee guide; and family- and community-involvement programs.

The sequenced elementary curriculum consists of 140 lessons per grade, per academic year, offered in 15 to 20 min by classroom teachers. When fully implemented, the total time students are exposed to the program during a 35-week academic year is approximately 35 hr. Lessons cover six major units on topics related to self-concept (i.e., the relationship of thoughts, feelings, and actions) physical and intellectual actions (e.g., hygiene, nutrition, physical activity, avoidance of harmful substances, decision-making skills, creative thinking), social/emotional actions for managing oneself responsibly (e.g., self-control, time management), getting along with others (e.g., empathy, altruism, respect, conflict resolution), being honest with yourself and others (e.g., self-honesty, integrity, self-appraisal), and continuous self-improvement (e.g., goal setting, problem solving, courage to try new things, persistence). The classroom curricula utilize an interactive approach, whereby interaction between teacher and student is encouraged through the use of structured discussions and activities, and interaction between students is encouraged through structured or semi-structured small-group activities, including games, role plays,
and practice of skills. For example, students are asked how they like to be treated. Regardless of age, socioeconomic status, gender, or culture, students and adults suggest the same top values of respect, fairness, kindness, honesty, understanding/empathy, and love, consistent with others’ findings (Nucci, 1997). These values are then adopted as the code of conduct for the classroom and school (Flay & Allred, in press).

The school-climate kit consists of materials to encourage and reinforce the six units of PA, coordinating schoolwide implementation. Included in the kit, the PA coordinator’s (principal’s) manual directs the use of materials such as posters, music, tokens, and certificates. It also includes information on planning and conducting assemblies, creating a PA newsletter, and establishing a PA committee to create a schoolwide PA culture. In addition, a counselor’s program, implemented by school counselors, specializes in developing positive actions with students at higher risk and their classrooms, families, and the school as a whole. The family-involvement program is available in various levels of involvement and promotes the core elements of the classroom curriculum and reinforces schoolwide positive actions. The parent manual is designed for parents to use at home and includes materials that parallel the classroom curriculum. The present study did not include the more intensive family kit. The community-development component of PA was not used in this trial.

Prior to the beginning of each academic year, teachers, administrators, and support staff (e.g., counselors) attended PA training sessions conducted by the program developer. The training sessions lasted approximately 3 to 4 hr for the initial year, and 1 to 2 hr for each successive year. Booster sessions, conducted by the Hawai‘i-based project coordinator and lasting approximately 30 to 50 min, were provided an average of once per academic year for each school. In addition, mini-conferences were held in February of each year to bring together five or six leaders and staff (e.g., principals, counselors, teachers) from each of the 10 participating schools to share ideas and experiences as well as to get answers to any concerns regarding implementing the program.

### Data and Measures

**Archival School-Level Indicators.** Archival school-level data were obtained from the Hawai‘i Department of Education (HDE) as part of the state’s SRC data accountability system (HDE, n.d.-b), with different indicators available at different time points as shown in Table 1. The SRC data were included in schools’ School Status and Improvement Report, designed to provide information on schools’ performance and progress. Absenteeism, suspensions, retention in grade, and four academic achievement indicators served as the dependent variables for the present study; these were chosen because they were the publicly available indicators of school performance. Corresponding classroom- and student-level data were not available due to privacy considerations.
Table 1. Timeline of the Positive Action Hawai’i trial and compilation of school-level indicators

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Positive Action trial</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
</tbody>
</table>

School-level indicator

- Standardized Test Math\(^a\) (% ≥ average) | x | x | x | x | x | x | x | x | x | x | x |
- Math HCPS II (% proficient) | x | x | x | x | x | x | x | x | x | x | x |
- Standardized Test Reading\(^a\) (% ≥ average) | x | x | x | x | x | x | x | x | x | x | x |
- Reading HCPS II (% proficient) | x | x | x | x | x | x | x | x | x | x | x |
- Absenteeism (average days absent) | x | x | x | x | x | x | x | x | x | x | x |
- Suspensions (% of students) | x | x | x | x | x | x | x | x | x | x | x |
- Retentions (% of students) | x | x | x | x | x | x | x | x | x | x | x |

Note. HCPS II = Hawai’i Content and Performance Standards.

\(^\dagger\)Positive Action schools continued to implement the program after completion of the randomized trial; three control schools chose to receive Positive Action after the trial. \(^a\)Standardized test scores included SAT (Stanford 9) for 2002–06 and TerraNova (2nd ed.) for 2007.
School-level performance is an appropriate measure of program effectiveness because the PA Hawai‘i trial tested a schoolwide implementation of the program and whole schools were randomized to condition (Stuart, 2007).

The four school-level academic achievement variables included the Grade 5 math and reading standardized test (percentage scoring average or above; the HDE switched from the Stanford Achievement Test [SAT] to the TerraNova [2nd ed.] test at 1-year follow-up during the current study), and the Grade 4 math and reading Hawai‘i Content and Performance Standards (HCPS II; percentage proficient). The math and reading SAT and TerraNova (2nd ed.) are national norm-referenced tests that are utilized by school districts in the United States to assess achievement of students from kindergarten through high school. The math and reading HCPS II were developed by the HDE through a collaborative process involving teachers and HDE curriculum specialists and represent the HDE performance standards to meet No Child Left Behind mandates (Hawai‘i Department of Education, n.d.-a). The archival school-level academic achievement data were available continuously, from 2002 to 1-year post-trial, as intervention schools continued to implement the PA program. Achievement scores were not reported for one of the 10 pairs of schools because they had too few students at each grade level, so these schools were not included in the primary analysis. There were no missing data for the other dependent variables.

The other three school-level indicators used in this study included (a) absenteeism (average number of days absent per year), (b) suspensions (percentage suspended), and (c) retentions (percentage retained in grade, i.e., kept back a grade). Student suspensions may have occurred due to, for example, disorderly conduct, burglary, truancy, and contraband (e.g., possession of tobacco). Suspension data represent all grade levels at each school, and the retention variable included students who were retained in all grades except kindergarten. The archival school-level absenteeism data were available annually from 1997 to 2007, the suspension data from 1999 to 2007, and the retention data from 2002 to 2007.

Thus, the archival data utilized in the present analysis were collected from schools with a different student body each academic year, and intervention schools, over time, had increasing exposure to PA. For example, archival school-level data collected for PA schools during the 2005–2006 academic year represented schools with students who were exposed to the intervention for up to 4 years compared to the 2002–2003 academic year.

Implementation. As part of the PA Hawai‘i trial, sufficient data from year-end process evaluation surveys were collected from teachers at the end of the second (2004), third (2005), and final year (2006) of program implementation and are described in detail elsewhere (Beets et al., 2008). We used three school-level implementation indicators related to program exposure and adherence: (a) exposure, measured by seven items (i.e., six items referred to the six units in the PA curriculum and asked about how often the teachers taught the concept
throughout the school day, and an additional item assessed the amount of PA workbooks and activity sheets used during a typical day); (b) classroom material usage, measured by three items (i.e., how often teachers used PA materials/activities); and (c) schoolwide material usage, measured by three items (i.e., how often PA materials/activities were used throughout the school). All item responses ranged from 1 (never) to 5 (always). Alpha reliabilities were adequate (Beets et al., 2008).

The three school-level implementation indicators and an overall school-level implementation indicator were calculated at the 2nd (2004), 3rd (2005), and final year (2006) of program implementation using several steps. First, based on teachers’ responses to the items that comprised each of the different implementation indicators, we calculated mean teacher-level indicator scores. Second, using the teacher-level indicator scores, a mean school-level implementation indicator was calculated for every school each year. Last, an overall school-level implementation indicator was calculated by computing the mean across all schools for each year of program implementation.

During the spring of the final year of the 4-year randomized trial, data were collect from one school leader (i.e., principal, vice principal, counselor) from each treatment and control school regarding the SACD programs and/or activities that were conducted in their school during the prior 3 academic years. Respondents were asked to list up to 16 SACD programs. For each program, respondents indicated the number of weeks the program was offered, the amount of time (minutes) devoted to the program per week, and whether teachers attended/received training to deliver the program (yes/no).

Analyses

For our primary analysis, we used matched-paired $t$ tests, Hedges’ adjusted $g$ as a measure of effect size (Grissom & Kim, 2005; Hedges & Olkin, 1985), and percentage relative improvement (RI). To assess the robustness of results, permutation tests and random-intercept growth curve models were used for sensitivity analyses. The random-effects growth curve models provide some statistical control beyond randomization for potentially confounding unmeasured variables in case randomization was not totally successful with 10 schools per condition. This battery of statistical approaches was used separately for each of the outcomes and was applied to end-of-study (2006) and 1-year post trial (2007) outcomes.

Primary Analysis. First, matched-paired $t$ tests of difference scores were used to examine change in school-level outcomes by condition. For each outcome, two difference scores [posttest (2006) − baseline (2002) and 1-year post trial (2007) − baseline (2002)] were calculated for each pair of intervention and control schools and a paired $t$ test was performed. In a randomized design,
the difference in means provides an unbiased estimate of the true average intervention effect (Stuart, 2007).

Second, effect sizes for absenteeism, suspensions, retentions, and each of the four achievement outcomes were calculated by subtracting the mean difference of control schools from the mean difference of PA schools and dividing by the pooled posttest standard deviation. Hedges’ g (as well as other measures of effect size such as Cohen’s d and Glass’ d) has some positive bias; therefore, Hedges’ approximately unbiased adjusted g was calculated. Moreover, the adjusted g is an appropriate effect size calculation when the sample size is small (Grissom & Kim, 2005). Effect sizes were examined at posttest and at one-year post trial and were interpreted as small (0.2), moderate (0.5) or large (0.8; Cohen, 1977).

In addition, we calculated RI as an indicator of effect size that may be more understandable to practitioners. RI is the posttest difference between groups minus the baseline difference between groups, divided by the control group posttest level, that is, \( \frac{(PA_{mean} - C_{mean})_{posttest} - (PA_{mean} - C_{mean})_{baseline}}{C_{mean\ posttest}} \), expressed as a percentage.

Sensitivity Analysis. Subsequently, to avoid reliance on t-test assumptions alone and as a sensitivity analysis, permutation tests were conducted with Stata v10 *permute*, which estimates p values based on Monte Carlo simulations (Stata Corp., College Station, TX). Both paired t tests of differences and permutation models have demonstrated good performance in randomized trials when the number of pairs is small (Brookmeyer & Chen, 1998).

Last, random-intercept growth curve models (see the appendix) were conducted with Stata v10 *xtmixed* (Rabe-Hesketh & Skrondal, 2008) to account for all observations and to model school differences. That is, this allows a more complete analysis of the multiple waves of available data (five waves of data at posttest; six waves of data at 1-year post trial) and takes into account the pattern of change over time. The random-intercept model allows the intercept to vary between schools, which indicates that some schools tend to have, on average, better outcomes and other schools have worse outcomes. The random coefficient is fixed, which reflects that intervention effects are similar for all schools. To estimate effects with missing values present, full information maximum likelihood estimation was used which utilizes all available data to provide maximum likelihood estimation (Acock, 2005). For the present analyses, each growth curve involved approximately 100 observations (5 waves × 20 schools at posttest; 6 waves × 20 schools at 1-year post trial). Although this sample size is at the lower end of some suggested guidelines for this estimator, it is adequate as a supplementary sensitivity analysis, as different views exist regarding appropriate sample size (Singer & Willett, 2003).

For each outcome, from baseline through both posttest and 1-year post trial, we tested whether a quadratic term for time was significant using the likelihood-ratio (LR) test (Rabe-Hesketh & Skrondal, 2008). Through posttest,
results indicated that a quadratic model provided a significantly better fit for the data on reading HCPS II, LR $\chi^2(1) = 14.92, p < .001$, and absenteeism, LR $\chi^2(1) = 6.25, p < .05$. Through 1-year post trial, results showed that a quadratic model fit significantly better for math TerraNova, LR $\chi^2(1) = 4.04, p < .05$; reading TerraNova, LR $\chi^2(1) = 4.56, p < .05$; math HCPS II, LR $\chi^2(1) = 17.04, p < .001$; and absenteeism, LR $\chi^2(1) = 19.39, p < .001$.

For the remaining outcomes (school suspensions and retentions), from baseline through both posttest and 1-year post trial, we conducted random-intercept Poisson models with Stata v10 \texttt{xtpoisson} (Rabe-Hesketh & Skrondal, 2008). As is common with elementary school-level data, frequency distributions for school suspensions and retentions were skewed at both posttest and 1-year post trial. Hence, a random-intercept Poisson model was used to account for this skewed distribution. The mean and variance of the suspension and retention variables were similar through posttest (suspensions [$M = 0.95$; variance $= 1.09$], retentions [$M = 0.99$; variance $= 0.92$]) and 1-year post trial (suspensions [$M = 1.07$; variance $= 1.72$], and retentions [$M = 0.94$; variance $= 0.88$]), an assumption of the Poisson model (Snijders & Bosker, 1999); therefore, we did not adjust for overdispersion. Similarly, as previously discussed, an LR test was used to compare random-intercept Poisson models with the inclusion of a quadratic term. Only the result for suspensions, LR $\chi^2(1) = 4.85, p < .05$, at 1-year post trial demonstrated a quadratic model provided a better fit for the data.

In addition, to test whether the pattern of curvilinear change was different in PA and control schools, a Year Squared $\times$ Condition interaction term was included in the quadratic models, and an LR test was performed. Results indicated that the inclusion of an interaction term did not significantly improve any of the quadratic models and, hence, was not included in the final models.

**RESULTS**

**Baseline Equivalency**

At the 2002 baseline no significant differences ($p \geq .05$) existed between intervention and control schools on any of the SRC variables (Table 2; Table 4 displays outcome variables). Thus, the methods of developing strata and random selection and assignment were effective for these variables. Schools were racially/ethnically diverse with a mean enrollment of 544 ($SD = 276.41$).

**Implementation**

There was some variability in school-level implementation between schools, with small improvements across years (Table 3). Regarding the three
<table>
<thead>
<tr>
<th>Characteristics</th>
<th>2002 (Baseline)</th>
<th>2006 (Posttest)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Control</td>
<td>Positive Action</td>
</tr>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>Enrollment</td>
<td>478.80</td>
<td>207.06</td>
</tr>
<tr>
<td>Racial/Ethnic Distribution (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>African American</td>
<td>1.79</td>
<td>3.20</td>
</tr>
<tr>
<td>Chinese</td>
<td>2.05</td>
<td>3.66</td>
</tr>
<tr>
<td>Filipino</td>
<td>11.61</td>
<td>14.20</td>
</tr>
<tr>
<td>Hawai‘ian</td>
<td>5.61</td>
<td>5.98</td>
</tr>
<tr>
<td>Hispanic</td>
<td>2.45</td>
<td>2.35</td>
</tr>
<tr>
<td>Indochinese</td>
<td>2.02</td>
<td>5.62</td>
</tr>
<tr>
<td>Japanese</td>
<td>4.26</td>
<td>3.57</td>
</tr>
<tr>
<td>Korean</td>
<td>1.19</td>
<td>2.12</td>
</tr>
<tr>
<td>Native American</td>
<td>0.44</td>
<td>0.37</td>
</tr>
<tr>
<td>Part Hawai‘ian</td>
<td>31.86</td>
<td>28.37</td>
</tr>
<tr>
<td>Portuguese</td>
<td>1.41</td>
<td>1.94</td>
</tr>
<tr>
<td>Samoan</td>
<td>3.11</td>
<td>4.83</td>
</tr>
<tr>
<td>White</td>
<td>17.52</td>
<td>18.05</td>
</tr>
<tr>
<td>Other</td>
<td>14.69</td>
<td>14.01</td>
</tr>
<tr>
<td>Stability (%)</td>
<td>90.82</td>
<td>2.36</td>
</tr>
<tr>
<td>Free/reduced lunch (%)</td>
<td>54.32</td>
<td>26.40</td>
</tr>
<tr>
<td>Limited English proficiency (%)</td>
<td>11.83</td>
<td>15.30</td>
</tr>
<tr>
<td>Special education (%)</td>
<td>10.56</td>
<td>5.41</td>
</tr>
</tbody>
</table>

^aTwo-tailed t test; 18 df.
Impact of the Positive Action Program

Table 3. School-level implementation

<table>
<thead>
<tr>
<th>Indicator</th>
<th>2004</th>
<th>2005</th>
<th>2006</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$M^a$</td>
<td>$SD$</td>
<td>$M^b$</td>
</tr>
<tr>
<td>Exposure/amount</td>
<td>3.08</td>
<td>0.65</td>
<td>3.29</td>
</tr>
<tr>
<td>Classroom material usage</td>
<td>2.58</td>
<td>0.71</td>
<td>2.65</td>
</tr>
<tr>
<td>Schoolwide material usage</td>
<td>3.41</td>
<td>1.02</td>
<td>3.62</td>
</tr>
<tr>
<td>Overall</td>
<td>3.02</td>
<td>0.74</td>
<td>3.18</td>
</tr>
</tbody>
</table>

Note. Means correspond to item scale: 1 (never) to 5 (always).

$a^N = 8. b^N = 10.$

School-level indicators examined, schoolwide material usage demonstrated the highest school-level implementation. Implementation was adequate for each indicator; however, results indicated that schools could have implemented PA with greater fidelity.

We found that control schools reported implementing an average of 10.2 SACD programs compared with 4.2—in addition to PA—in the intervention schools. Teachers in control schools spent an average of 108 min per week on SACD-related activities. PA-school teachers spent the expected amount of time on PA (55.1 min/week), yet overall they still spent only 35 min/week more on SACD-related activities than teachers in control schools. Control schools reported that teachers were involved in SACD-related activities for an average of 24 weeks per school year. In contrast, teachers in intervention schools reported delivering PA almost every week of the school year as well as being involved in other SACD-related activities for 25 weeks/year. Both PA and control school teachers reported receiving training to implement approximately half of the SACD-related programs (52.3% and 53.3%, respectively) that they reported implementing other than PA (100% trained).

School-Level Raw Means

Raw means for school-level academic achievement, absenteeism, suspensions, and retentions are presented in Figures 1 and 2, respectively. Overall, for the academic achievement outcomes, raw means for PA and control schools were statistically similar at baseline and demonstrated a clearly discernable divergence over time. State averages for academic achievement are shown for comparison. Although the PA schools were well below state averages at baseline (as planned), they nearly met or exceeded the state averages for academic achievement at posttest and 1-year post trial.

Likewise, for the other school-level outcomes, PA and control schools diverged between baseline and posttest. For absenteeism and suspensions,
Figure 1. School-level means for math and reading achievement. Note. Hawai‘i Randomized Trial occurred 2002–03 to 2005–06. Standardized test scores included SAT (Stanford 9) for 2002–06 and TerraNova (2nd ed.) for 2007. PA = Positive Action; HCPS II = Hawai‘i Content and Performance Standards.
Figure 2. School-level means for absenteeism, suspensions, and retention. Note. Hawai‘i Randomized Trial occurred 2002–03 to 2005–06. PA = Positive Action.
prebaseline years of archival school-level data were available and provide an interrupted time series presentation. As expected, these outcomes were stable for several preprogram years with divergence occurring after the intervention.

**Matched Paired $t$ Tests and Effect Sizes**

The results of the matched paired $t$ tests of difference scores and effect size calculations at posttest and 1-year post trial are presented in Table 4. At posttest, results indicated that PA schools had significantly higher math ($p < .05$) and reading ($p < .05$) HCPS II scores; and significantly lower absenteeism ($p < .001$), with marginally fewer suspensions ($p = .056$). After completion of the randomized trial, at 1-year post trial as PA schools continued to implement the PA program, reading TerraNova ($p < .05$) and math ($p < .01$) and reading ($p < .05$) HCPS II were significantly higher among PA schools, and absenteeism ($p < .001$) and suspensions ($p < .05$) were significantly lower for PA schools. Overall, results indicated higher achievement and lower absenteeism and suspension outcomes for the PA schools. The permutation models provided similar statistically significant results as the matched paired $t$ tests at both posttests. That is, permutation tests at posttest indicated statistically significant results for math (marginal $p = .054$) and reading ($p < .01$) HCPS II and absenteeism ($p < .01$), and at 1-year post trial reading ($p < .05$) TerraNova, math ($p < .001$), and reading ($p < .05$) HCPS II, absenteeism ($p < .001$), and suspensions ($p < .05$) were significantly different for PA schools as compared to control schools.

To provide a basis for comparing the magnitude of the intervention effects we found with effects found in other trials, effect sizes were calculated. As shown in Table 4, all of the effect sizes were moderate to large, regardless of the level of significance. Corresponding effect size calculations demonstrated moderate to large treatment effects for the academic achievement, absenteeism, and disciplinary outcomes at posttest, with larger effects at 1-year post trial. Similarly, RIs were larger at 1-year post trial.

**Random-Intercept Growth Curve Models**

The estimates for the intervention effect on academic achievement scores (random-intercept models) from baseline through posttest and 1-year post trial are presented in Table 5. At posttest, the intraclass correlation coefficient (ICC; expressed as the proportion of the total outcome variation that is attributable to differences among schools) for the unconditional means models (Singer & Willett, 2003) were .72, .67, .87, and .72 for math SAT and HCPS II and reading SAT and HCPS II, respectively. At 1-year post trial, the ICC for the unconditional means models were .68, .46, .87, and .66 for math TerraNova and HCPS II and reading TerraNova and HCPS II, respectively,
Table 4. Baseline measures, school-level matched paired \(t\) tests of difference scores, and effect sizes for math achievement, reading achievement, absenteeism, suspensions, and retentions

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Condition</th>
<th>2002 (Baseline)</th>
<th>2006 (Posttest)</th>
<th>2007 (1-Year Post Trial)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>(M)</td>
<td>(SD)</td>
<td>(p^b)</td>
</tr>
<tr>
<td>Stand. Test Math^f</td>
<td>Control</td>
<td>76.56</td>
<td>13.73</td>
<td>0.957</td>
</tr>
<tr>
<td>(% ≥ average)</td>
<td>Control</td>
<td>76.22</td>
<td>11.72</td>
<td></td>
</tr>
<tr>
<td>Math HCPS II</td>
<td>Control</td>
<td>15.56</td>
<td>10.01</td>
<td>1.000</td>
</tr>
<tr>
<td>(% proficient)</td>
<td>PA</td>
<td>15.56</td>
<td>7.81</td>
<td></td>
</tr>
<tr>
<td>Stand. Test Reading^f</td>
<td>Control</td>
<td>71.74</td>
<td>13.36</td>
<td>0.962</td>
</tr>
<tr>
<td>(% ≥ average)</td>
<td>PA</td>
<td>71.44</td>
<td>13.12</td>
<td></td>
</tr>
<tr>
<td>Reading HCPS II</td>
<td>Control</td>
<td>35.67</td>
<td>16.67</td>
<td>0.904</td>
</tr>
<tr>
<td>(% proficient)</td>
<td>PA</td>
<td>34.89</td>
<td>9.37</td>
<td></td>
</tr>
<tr>
<td>Absenteeism (Average days</td>
<td>Control</td>
<td>11.00</td>
<td>2.27</td>
<td>0.872</td>
</tr>
<tr>
<td>absent)</td>
<td>PA</td>
<td>11.18</td>
<td>2.66</td>
<td></td>
</tr>
<tr>
<td>Suspensions</td>
<td>Control</td>
<td>0.98</td>
<td>1.11</td>
<td>0.777</td>
</tr>
<tr>
<td>(% of students)</td>
<td>PA</td>
<td>1.12</td>
<td>1.10</td>
<td></td>
</tr>
<tr>
<td>Retentions (% of students)</td>
<td>Control</td>
<td>1.50</td>
<td>0.972</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>PA</td>
<td>1.50</td>
<td>1.08</td>
<td></td>
</tr>
</tbody>
</table>

Note. ES = effect size; RI = Relative Improvement; \(\text{PA} = \text{Positive Action}\); HCPS II = Hawai‘i Content and Performance Standards. 
\(^a\)Two-tailed \(t\) test; 18 \(df\) for all variables except achievement-related variables (16 \(df\)). \(^b\)Mean difference = posttest - baseline; 1-year post trial - baseline. \(^c\)Two-tailed paired \(t\)-test difference score; 8 \(df\) for achievement-related variables and 9 \(df\) for other variables. \(^d\)Hedges’ \(g\) effect size (unbiased adjusted \(g\)) of mean difference. \(^e\)RI = \(\frac{(\text{PA}_{\text{mean}} - \text{C}_{\text{mean}})_{\text{posttest}} - (\text{PA}_{\text{mean}} - \text{C}_{\text{mean}})_{\text{baseline}}}{\text{C}_{\text{mean}} \text{ posttest}}\). 
\(^f\)Standardized test scores included Stanford Achievement Test (Stanford 9) for 2002–06 and TerraNova (2nd ed.) for 2007.
Table 5. School-level random-intercept growth model estimates for math and reading achievement

<table>
<thead>
<tr>
<th></th>
<th>Stand. Test Matha (% Average or Better)</th>
<th>Math HCPS II (% Proficient)</th>
<th>Stand. Test Readinga (% Average or Better)</th>
<th>Reading HCPS II (% Proficient)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Bb</td>
<td>SE</td>
<td>Bb</td>
<td>SE</td>
</tr>
<tr>
<td>2006 (Posttest)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>75.79***</td>
<td>3.26</td>
<td>14.68***</td>
<td>2.97</td>
</tr>
<tr>
<td>Year</td>
<td>0.70</td>
<td>0.52</td>
<td>0.77</td>
<td>0.47</td>
</tr>
<tr>
<td>Year²</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Condition (0 = C; 1 = PA)</td>
<td>-2.39</td>
<td>4.60</td>
<td>-2.50</td>
<td>4.21</td>
</tr>
<tr>
<td>Year × Condition</td>
<td>1.19†</td>
<td>0.73</td>
<td>2.10**</td>
<td>0.67</td>
</tr>
<tr>
<td>Random effects</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>School-level variance</td>
<td>69.13</td>
<td>24.65</td>
<td>57.44</td>
<td>20.50</td>
</tr>
<tr>
<td>Residual variance</td>
<td>23.89</td>
<td>33.98</td>
<td>20.17</td>
<td>3.36</td>
</tr>
<tr>
<td>2007 (1-year post trial)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>70.91***</td>
<td>3.76</td>
<td>20.21***</td>
<td>3.66</td>
</tr>
<tr>
<td>Year</td>
<td>5.49***</td>
<td>1.48</td>
<td>-4.28**</td>
<td>1.50</td>
</tr>
<tr>
<td>Year²</td>
<td>-0.96***</td>
<td>0.20</td>
<td>0.89***</td>
<td>0.21</td>
</tr>
<tr>
<td>Condition (0 = C; 1 = PA)</td>
<td>-2.73</td>
<td>4.41</td>
<td>-3.87</td>
<td>4.41</td>
</tr>
<tr>
<td>Year × Condition</td>
<td>1.34*</td>
<td>0.59</td>
<td>2.69***</td>
<td>0.60</td>
</tr>
<tr>
<td>Random effects</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>School-level variance</td>
<td>71.63</td>
<td>25.41</td>
<td>62.67</td>
<td>22.49</td>
</tr>
<tr>
<td>Residual variance</td>
<td>27.48</td>
<td>4.10</td>
<td>28.55</td>
<td>4.26</td>
</tr>
</tbody>
</table>

Note. C = control; PA = Positive Action.

*aStandardized test scores included SAT (Stanford 9) for 2002–06 and TerraNova (2nd ed.) for 2007. bB estimate based on random-intercept model.

†p < .10. *p < .05. **p < .01. ***p < .001; all two-tailed.
indicating that most of the variation in academic achievement lies between schools, rather than within schools over time. Overall, through both posttest and 1-year post trial, the random-intercept models’ Year × Condition interactions substantiated results of the matched-paired t tests and permutation models, indicating higher achievement increases in PA schools. For change from baseline through 1-year post trial, the time by condition interactions for math TerraNova (B = 1.34, p < .05) and HCPSII (B = 2.69, p < .001) and reading TerraNova (B = 1.35, p < .01) and HCPS II (B = 2.10, p < .05) were all statistically significant. These effects indicate about a 2 percentage point advantage per year for the PA group compared to the control group due to the intervention, or about a 12 percentage point advantage across the 6-year period.

The estimates for the intervention effect on the absenteeism, suspension, and retention outcomes (random-intercept and random-intercept Poisson models) from baseline through both posttest and 1-year post trial are presented in Table 6. Parameter estimates and incidence rate ratios (IRR) are each presented for the random-intercept Poisson models, as an intercept parameter is not calculated for IRR estimates and, additionally, a residual variance estimate is not part of such models (Rabe-Hesketh & Skrondal, 2008). At posttest, the ICCs for the unconditional means models were .88, .52, and .47 for absenteeism, suspensions, and retentions, respectively. The ICC values for the Poisson models are approximations and were calculated utilizing a similar approach as used for the random-intercept models (Goldstein, Browne, & Rasbash, 2002). At 1-year post trial, the ICCs for the unconditional means models were .88, .52, and .41 for absenteeism, suspensions, and retentions, respectively. Thus, much of the variation in absenteeism, nearly half of the variation in suspensions, and less than half the total variation in retentions can be attributable to differences between schools.

Regarding absenteeism, from baseline through both posttest (Year × Condition B = −0.45, p < .001) and 1-year post trial (Year × Condition B = −0.36, p < .001), the random-intercept growth models substantiated results of the matched-paired t tests, demonstrating a significant reduction in absenteeism among PA schools relative to control schools. However, as compared to the matched-paired t tests, inconsistent results emerged for the suspension and retention outcomes. The random-intercept growth curves indicated a marginally significant (B = −0.20, p = .06; IRR [95% CI] = 0.82 [0.67, 1.01]) Year × Condition interaction for the suspension outcome from baseline to 1-year post trial, where the t tests did not. Further, inconsistent with the nonsignificant matched-paired t test, the Retention Year × Condition interactions through posttest (B = −0.30, p < .05; IRR = 0.74 [0.54–1.00]) and 1-year post trial (B = −0.30, p < .05; IRR = 0.74 [0.58–0.95]) were statistically significant. Therefore, overall, the random-intercept and random-intercept Poisson models demonstrate decreased absenteeism, disciplinary, and retention outcomes among PA schools relative to control schools.
Table 6. School-level random-intercept growth model estimates for absenteeism, suspensions, retentions

<table>
<thead>
<tr>
<th></th>
<th>Absenteeism (Average Days Absent/Year)</th>
<th>Suspensions (% Suspended)</th>
<th>Retentions (% Retained)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$B^a$</td>
<td>SE</td>
<td>$B^b$</td>
</tr>
<tr>
<td><strong>2006 (Posttest)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Fixed effects</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>11.56***</td>
<td>0.80</td>
<td>-0.15</td>
</tr>
<tr>
<td>Year</td>
<td>-0.54*</td>
<td>0.27</td>
<td>0.12</td>
</tr>
<tr>
<td>Year$^2$</td>
<td>0.11**</td>
<td>0.04</td>
<td>0.15</td>
</tr>
<tr>
<td>Condition ($0 = C; 1 = PA$)</td>
<td>0.47</td>
<td>1.04</td>
<td>0.15</td>
</tr>
<tr>
<td>Year $\times$ Condition</td>
<td>-0.45***</td>
<td>0.10</td>
<td>-0.28$^\dagger$</td>
</tr>
<tr>
<td><strong>Random effects</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>School-level variance</td>
<td>4.85</td>
<td>1.57</td>
<td>0.44</td>
</tr>
<tr>
<td>Residual variance</td>
<td>0.54</td>
<td>0.09</td>
<td></td>
</tr>
<tr>
<td><strong>2007 (1-year post trial)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Fixed effects</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>11.52***</td>
<td>0.78</td>
<td>0.49</td>
</tr>
<tr>
<td>Year</td>
<td>-0.45*</td>
<td>0.20</td>
<td>-0.39</td>
</tr>
<tr>
<td>Year$^2$</td>
<td>0.09***</td>
<td>0.03</td>
<td>0.08$^*$</td>
</tr>
<tr>
<td>Condition ($0 = C; 1 = PA$)</td>
<td>0.28</td>
<td>1.05</td>
<td>-0.05</td>
</tr>
<tr>
<td>Year $\times$ Condition</td>
<td>-0.36***</td>
<td>0.08</td>
<td>-0.20$^\dagger$</td>
</tr>
<tr>
<td><strong>Random effects</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>School-level variance</td>
<td>5.01</td>
<td>1.61</td>
<td>0.54</td>
</tr>
<tr>
<td>Residual variance</td>
<td>0.54</td>
<td>0.08</td>
<td></td>
</tr>
</tbody>
</table>

*Note. C = control; PA = Positive Action; IRR = incidence rate ratio; CI = 95% confidence interval.

$^a$B estimate based on random-intercept model. $^b$B estimate based on random-intercept Poisson model. $^c$IRR estimate based on random-intercept Poisson model.

$^\dagger$p < .10. $^*p < .05. $^{**}p < .01. $^{***}p < .001; all two-tailed.
DISCUSSION

The present study extends previous research on the capabilities of school-based interventions targeting social-emotional and character development to improve academic performance and attendance and reduce disciplinary problems and grade retention in schools. This study also confirms earlier preliminary findings of beneficial results of the PA program from quasi-experimental studies (Flay & Allred, 2003; Flay et al., 2001) using a matched-pair, cluster-randomized, controlled trial. Specifically, as indicated by matched-paired $t$ tests and permutation models, PA schools scored significantly better than control schools in reading TerraNova and math and reading HCPS II and significantly lower absenteeism and suspensions at 1-year post trial. Moreover, random-intercept growth models demonstrated that PA schools showed significantly greater growth in math and reading TerraNova, math and reading HCPS II, and significantly lower absenteeism and retentions through 1-year post trial, with suspensions showing marginal significance. Indeed, school-level means for math and reading achievement demonstrated that PA schools, which were below state averages at baseline, nearly met or exceeded state averages by posttest and 1-year post trial. These findings were especially noteworthy because many of the schools were in low income areas and had a high level of racial/ethnic diversity.

The present results demonstrated moderate to large effect sizes on all of the observed outcomes and were likely the result of several notable attributes of the PA program. First, PA addresses distal influences on behavior in a multifaceted way; PA is a comprehensive approach that involves providing the curriculum to all grades in the school at once, involving all teachers and staff in the school, and involving parents and the community. The PA program assists students and adults to gain not only the knowledge, attitudes, norms, and skills that they might gain from other programs but also improved values, self-concept, family bonding, peer selection, communication, and appreciation of school, with the expected result of improvement in academic performance and a broad range of behaviors. These improved outcomes may occur because positive behaviors tend to correlate negatively with negative behaviors (Flay, 2002). More specifically, with regards to academic achievement, for example, PA increases positive behaviors and decreases disruptive behaviors, which in turn lead to more time on task for teaching and, in turn, more opportunity for student learning (Flay & Allred, in press). Also, improvements in students’ positive behaviors, such as attention and inhibitory control, can lead to increased academic achievement throughout formal schooling (McClelland, Acock, & Morrison, 2006).

Second, PA is “interactive” in delivery, using methods that integrate teacher/student contact and communication opportunities for the exchange of ideas, and utilize feedback and constructive criticism in a nonthreatening atmosphere (Tobler et al., 2000). Third, the results observed may also have been a consequence of the intensive nature of the program, with students receiving approximately 1 hr of exposure during a typical week over multiple school years.
Last, in the present study, we believe that the beneficial effects of the $PA$ program could have been even greater if the fidelity of implementation was excellent.

This analysis has some limitations. First, data regarding academic achievement, absenteeism, suspensions, and retention outcomes were not available at the student or classroom level. Because of this, variation in scores within students across years, or variation between students within schools could not be examined. As a result, individual student or classroom characteristics could not be included as predictors in the models to reduce unexplained variation. However, with random assignment, student and classroom characteristics should be about the same in the intervention and control groups. In addition, random-intercept models provide some statistical control for unmeasured differences between schools. Because every student’s score contributes to a school’s mean score, the design and analysis in this study provides a good test for intervention effects (Stuart, 2007). Future work that utilizes multilevel analysis of student-level indicators of academic achievement, absenteeism, and disciplinary outcomes would be beneficial.

Second, although school-level data are useful for estimating causal effects (Stuart, 2007), there may be inconsistencies among schools regarding how data, such as disciplinary-related referrals, are reported. Furthermore, it is possible that an intervention could influence how these data are reported. For example, a negative behavior that resulted in a disciplinary referral before an intervention may be handled in a different way after an intervention like $PA$.

A third limitation of our analyses is that only 20 schools participated in the study, with five waves of data resulting in 100 observations per random-effects growth curve model. Under conditions of small effect size and high ICC, this could result in relatively low statistical power to detect differences between treatment and control schools. This study found moderate to large effect sizes, but also large ICCs, so power was a concern. However, a successful matched-pair design can improve statistical power (Raudenbush, Martinez, & Spybrook, 2007), and our findings demonstrate a successful matched-pair design as well as its ability to detect statistical significance.

Fourth, there were a limited number of observations available for the random-effects growth curve models. With full information maximum likelihood estimation used in those models, a large sample is desirable (Hayes, 2006) to guarantee the accuracy of the estimates, although there are various viewpoints on what constitutes a large sample size (Singer & Willett, 2003). Our sample was large enough to use these models to compare the sensitivity of the matched-paired $t$ tests and permutation tests to an alternative statistical model, with different assumptions. The random-intercept models substantiated our findings from the more basic tests.

Fifth, although we demonstrated adequate implementation of $PA$ and realize the importance of implementation fidelity (Flay et al., 2005), we had insufficient data (i.e., insufficient variation given a sample of only 10 $PA$ schools) to examine implementation as a covariate. Also, we did not have data to observe the change in SACD-related activities in control schools. As indicated...
by the data procured during the last year of the 4-year trail, the widespread self-initiation of SACD-related activities, especially in control schools, can reduce the possible effect size that can be detected when evaluating school-based interventions (Hulleman & Cordray, 2009). In addition, because implementation data were not collected after completion of the randomized trial, we could not examine implementation at 1-year post trial. Future studies with larger samples of schools would be valuable to examine the effects of implementation fidelity on school-level outcomes.

Last, as with all other similar studies, results can be generalized only to schools that are willing to conduct such a program. Though our sample was adequate for this study, a larger representative sample of schools, or randomized trials at different locations, would allow generalization of results to a broader population.

These limitations notwithstanding, this study is the first to examine the effects of PA on school-level achievement, absenteeism, and disciplinary outcomes using a matched-pair, cluster-randomized, controlled design. The study extends research on the ways that changing a child’s developmental status in nonacademic areas can significantly enhance academic achievement (Catalano et al., 2004; Catalano et al., 2002; Flay, 2002) and actually may be essential for it. Future research should examine the specific mechanisms, moderators and mediators of social and character development intervention effects. Such knowledge would allow adjustments to PA that might increase the beneficial effect.

Unfortunately, elementary schools, with many demands for accountability, may concentrate solely on math, reading, and science achievement, and due to resource and time constraints, instruction regarding social and character development may be abandoned. The findings of this study provide evidence that the PA program, which has demonstrated effects on improving student behavior and character (Beets et al., 2009; Li et al., 2009), can also reduce school-level absenteeism and disciplinary outcomes and, concurrently, positively influence school-level achievement. Indeed, this study makes clear that a comprehensive school-based program that addresses multiple co-occurring behaviors can positively affect both behavior and academics.

ACKNOWLEDGMENTS

This project was funded by the National Institute on Drug Abuse (R01-DA13474). Additionally, The National Institute on Drug Abuse (DA018760) provided financial support for the completion of the work on this manuscript. The authors would like to thankfully recognize the support and involvement of the Hawaii school district and the principals, administrators, teachers, staff, students and their families at the participating schools. We also thank Howard Humphreys and Jonathan Wang for help with data collection and management.
Notice of potential conflict of interest: The research described herein was done using the program and the training and technical support of Positive Action, Inc. Dr. Flay’s spouse holds a significant financial interest in Positive Action, Inc.

REFERENCES


Impact of the Positive Action Program


**APPENDIX**

1. Random intercept mixed linear models:
   a. Random-intercept model

   \[ Y_{ij} = \beta_{oj} + \beta_1(\text{condition}_j) + \beta_2(\text{year}_{ij}) + \beta_3(\text{year}_{ij} \times \text{condition}_j) + \zeta_j + \epsilon_{ij} \]

   b. Random-intercept quadratic model

   \[ Y_{ij} = \beta_{oj} + \beta_1(\text{condition}_j) + \beta_2(\text{year}_{ij}) + \beta_3(\text{year}_{ij}^2) + \beta_4(\text{year}_{ij} \times \text{condition}_j) + \zeta_j + \epsilon_{ij} \]

   \[ Y_{ij} = \text{estimated outcome} \]

   \[ \beta_{oj} = \text{mean intercept} \]

   \[ \zeta_j = \text{random intercept} \]

   \[ \epsilon_{ij} = \text{level-1 residual} \]

2. Random-intercept Poisson models:
   The estimated outcome, \( Y_{ij} \), is assumed to have a Poisson distribution with expectation \( \mu_{ij} \).
   a. Random-intercept Poisson model

   \[ \mu_{ij} = \exp\{\beta_{oj} + \beta_1(\text{condition}_j) + \beta_2(\text{year}_{ij}) + \beta_3(\text{year}_{ij} \times \text{condition}_j) + \zeta_j \} \]

   b. Random-intercept Poisson quadratic model

   \[ \mu_{ij} = \exp\{\beta_{oj} + \beta_1(\text{condition}_j) + \beta_2(\text{year}_{ij}) + \beta_3(\text{year}_{ij}^2) + \beta_4(\text{year}_{ij} \times \text{condition}_j) + \zeta_j \} \]

   \[ \mu_{ij} = \text{mean rate at which outcome occurs.} \]