CITATION PAGE

Bavarian, N., Lewis, K. M., DuBois, D. L., Acock, A., Vuchinich, S., Silverthorn, N., Snyder, F. J., Day, J., Ji, P., & Flay, B. R. (2013). Using social-emotional and character development to improve academic outcomes: A matched-pair, cluster-randomized trial in low-income, urban schools. *Journal of School Health*, *83*, 771-779.

Peer-Reviewed, Pre-Publication Version.

ABSTRACT

BACKGROUND: School-based social-emotional and character development (SECD) programs can influence not only SECD, but also academic-related outcomes. The purpose of this study was to evaluate the impact of one SECD program, *Positive Action* (*PA*), on educational outcomes among low-income, urban youth.

METHODS: The longitudinal study used a matched-pair, cluster-randomized controlled design. Student-reported disaffection with learning and academic grades, and teacher ratings of academic ability and motivation were assessed for a cohort followed from grades 3 to 8. Aggregate school records were used to assess standardized test performance (for entire school, cohort, and demographic subgroups) and absenteeism (entire school). Multilevel growth-curve analyses tested program effects.

RESULTS: *PA* significantly improved growth in academic motivation and mitigated disaffection with learning. There was a positive impact of *PA* on absenteeism and marginally significant impact on math performance of all students. There were favorable program effects on reading for African American males and cohort students transitioning between grades 7 and 8, and on math for females and low-income students.

CONCLUSIONS: A school-based SECD program was found to influence academic outcomes among students living in low-income, urban communities. Future research should examine mechanisms by which changes in SECD influence changes in academic outcomes.

Keywords: Child & Adolescent Health; Emotional Health; Public Health

Using social-emotional and character development to improve academic outcomes: A matched-pair, cluster-randomized trial in low-income, urban schools

A growing body of research indicates that school-based social-emotional and character development (SECD) and SECD-like programs (e.g., social-emotional learning, positive youth development) can influence health behaviors and academic achievement among low-income minority youth, a population disproportionately affected by disparities in health[1] and education.[2] In their meta-analysis examining the impact of school-based mental health and behavioral programs set in low-income, urban schools, Farahmand and colleagues[3] reported a mean effect size (generally Hedges *g*) on academic outcomes of 0.24. Durlak and colleagues[4] reported a mean effect size (generally Hedges *g*) on academic outcomes of 0.27 in their meta-analysis on school-based social-emotional learning (SEL) programs. With respect to health-related outcomes, the Durlak meta-analysis[4] also showed SEL programs decreased conduct problems (effect size = 0.22) and emotional distress (effect size = 0.24), and improved positive social behaviors (effect size = 0.24). While encouraging, there is a need to accumulate further evidence regarding the capacity of SECD programs to promote academic outcomes, especially when implemented in low-income, urban schools. Accordingly, the primary purpose of this study was to examine the impact of one comprehensive, school-wide SECD program, *Positive Action*, on academic outcomes using a longitudinal cluster-randomized controlled design in low-income, urban schools.

Positive Action[5] is grounded in theories of self-concept,[6-8] is consistent with social-ecological theories of health behaviors such as the Theory of Triadic Influence (TTI),[9, 10] and proposes positive feelings, thoughts, and actions result in fewer negative behaviors and enhanced motivation to learn. The core curriculum is taught through 6 units: self-concept, positive actions for mind and body, positive social-emotional actions focusing on getting along with others, and managing, being honest with, and continually improving oneself. The sequenced classroom curriculum consists of over 140 15-20 minute, age-appropriate lessons per grade taught 4 days per week for grades K-6, and 70 20 minute lessons taught 2 days per week for grades 7 and 8. The *PA* program also includes teacher, counselor, family, and community training, and school-wide climate development; the school-climate kit, which was used by every school in the trial assigned to the *PA* condition, focuses on using curriculum lessons and school activities to promote further positive actions amongst students, the school, families, and the community. More information about *PA* is available at http://www.positiveaction.net.

Prior research has demonstrated that the *PA* program impacts a range of risk and resilience factors linked to academic outcomes, as well as academic outcomes themselves. [6] In an analysis of 3 longitudinal randomized controlled trials (RCT) of *PA* involving students aged 6 to 11 years, *PA* partially mitigated the decrease in number of positive behaviors endorsed by youth across time. [11] In a matched-pair RCT of *PA* involving 20 schools in Hawai'i, *PA* was shown to create whole-school contextual change and improve school quality. [12] Students in schools receiving *PA* were also less likely to engage in substance use, violent behaviors, or sexual activity, [13] and *PA* schools had significantly higher school-level academic achievement and less absenteeism. [14]

Limitations in prior *PA* research should be addressed. For example, the academic impact of *PA* during the middle-school years has not yet been examined. Doing so is critical, as the adolescent years represent a key developmental period with new academic and social demands. Also, the need exists to collect academic-related data from students and teachers so that precursors of academic achievement (e.g., engagement with learning) that cannot be measured by school-level archival records alone can be assessed. Lastly, the need exists for experimental designs of *PA* in low-income, urban settings. The present study addresses these limitations by 1) following a cohort of students during the elementary- and middle-school years, 2) including student self-reports and teacher ratings of students, and 3) being set in a low-income, urban setting. The purpose was to test the hypothesis that academic performance across time would be better among schools and students receiving *PA*, than those not receiving *PA*.

METHODS

Participants

Participating schools were drawn from 483 K-6 and K-8 Chicago Public Schools. Schools were excluded from participation if they: 1) were non-community schools (e.g., charter schools and magnet schools), 2) already had *PA* or a similar intervention, 3) had an enrollment rate below 50 or above 140 students per grade, 4) had annual student mobility rates over 40%, 5) had more than 50% of students who passed the Illinois State Achievement Test (ISAT), and 6) had fewer than 50% of students who received free lunch. The latter two criteria ensured the selection of high-risk schools. A total of 68 schools met eligibility criteria, of which 18 agreed to participate, and the 7 best-matched pairs (the N that funding would support) were selected for participation; the following variables were used in the matching process: ethnicity, percentage of students who met or exceeded criteria for passing the ISAT, attendance rate, truancy rate, percentage of students who received free lunch, percentage of students who enrolled in or left school during the academic year, number of students per grade, percentage of parents reported to demonstrate school involvement, percentage of teachers employed by the school who met minimal teaching standards, and crime rate for the neighborhood in which the school was located. [15-18] A series of *t*-tests revealed that the 7 pairs of schools did not significantly differ from the remainder of the 68 schools eligible for the study, and the *PA* and control schools were not significantly different from each other on any of the matching variables. [15, 17] Throughout the 6 years of the study, 100% of schools were retained.

The total number of students in the analytic sample was 1,170, of whom approximately 53% were female; approximately 48% were African American, 27% Hispanic and 19% other (e.g., White, Asian, Native American, or “Other”). A total of 247 teachers completed student assessments; 75% of teachers were female; 43% White, 36% African American, 13% Hispanic and 8% other (Asian and Native American).

Instruments

Student self-report measures

Disaffection with learning was assessed using four items from a measure of student engagement developed by Furrer and Skinner.[19] Principal components factor analysis on student responses showed this measure loaded strongly onto one factor at both Wave 1 (loadings greater than or equal to 0.66) and Wave 8 (loadings greater than or equal to 0.67). Items were rated on a four-point Likert scale (“Disagree A LOT” to “Agree A LOT”) and included “When I’m in class, I think about other things” and “When I’m in class, my mind wanders”. A mean of the items was used to create a composite score, whereby higher scores reflected having more disaffection. Cronbach’s alpha across the 8 waves of data ranged from 0.64 to 0.71.

To assess the impact on *academic grades*, students were asked, “What grades have you been getting this school year?” with response options ranging from 1 to 9 (e.g., 1 = Mostly F’s, 4 = mix of C’s and D’s, and 9 = Mostly A’s).

Teacher ratings of students

Teachers assessed students using pre-existing measures of academic ability and motivation.[20, 21] Each consented student was rated in the areas of reading, mathematics, academic performance, and intellectual functioning using a 5-point Likert scale (1 = Far below grade level to 5 = Far above grade level). Due to multicollinearity (i.e., correlations of 0.84 and higher) between these items, a composite score was created, with higher scores indicating higher teacher ratings of students' *academic ability*. Cronbach's alpha for the composite measure ranged from 0.97 to 0.98. *Academic motivation* was assessed with a single-item measure, with response options ranging from "Extremely low" to "Extremely high".

School-level archival data

Because state test data provide a policy-relevant measure of achievement,[22] archival reading and math scores of non-English Language Learners on a standardized, school-administered, statewide test (the *ISAT*) were gathered from the Chicago Public Schools website.[23] The website provided information on the percentages of students tested (all students, grade-specific, and demographic subgroups) whose scores fell into each category (i.e. Warning, Not Meeting Standards, Meeting Standards, or Exceeding Standards). A single weighted average of the percentages of students falling into each achievement level was created for each school (i.e., [[1 × % of students at Warning level] + [2 × % of students NOT meeting standards] + [3 × % of students meeting standards] + [4 × % of students exceeding standards]]) for both reading and math, overall and by demographic sub-groups.

A *value-added metric index of ISAT* performance was also reported by the school district.[24] These indices control for the prior year ISAT scores of students as well as other relevant factors (i.e., grade level, gender, race/ethnicity, low income status, English Language Learner status, Individualized Education Plan status, homelessness, and mobility) and are designed to reflect the extent to which scores for a group of students improved (or declined) more than would be predicted based on these factors. Data were available for our student cohort transitioning from grades 7 to 8 (2009-2010).

The school district reported average daily attendance rates for each school on a scale from 0 to 100%; these statistics were converted to a measure of average daily *absenteeism* by subtracting 100 from each school's respective year-end attendance.

Procedure

The Chicago trial of *PA* was longitudinal (i.e. 6 years and 8 waves) at the school level and used a place-focused, intent-to-treat design with a dynamic cohort at the student level. [25] Surveys were administered to students beginning in grade 3 (fall 2004), and at 7 additional time points (waves) over 6 years: spring 2005, fall 2005, spring 2006, spring 2007, fall 2008, spring 2009, and spring 2010 (end of grade 8).

Parental consent was obtained before students, parents, or teachers completed surveys when students were in grade 3, with students joining the study at later waves consented at the time of entry into the study. All students were re-consented for the second phase of funding at wave 6. At baseline, 79% of parents provided consent; consent rates ranged from 65% to 78% for waves 2 through 5, and from 58% to 64% for waves 6 through 8.

The total number of students in the analytic sample across all waves was 1,170. Of the original 624 students in grade 3 at the beginning of the trial, only 131 (i.e., 21%) remained at grade 8, reflecting the high mobility by low-income urban students. With respect to maintenance of the baseline sample size, 363 students were present at wave 8 (i.e., approximately 61% of the Wave 1 sample size); the decrease in N over time is consistent with the trend among Chicago

Public Schools to decrease in size during the study period, together with lower consent rates at wave 6 through 8.[15]

To substantiate student self-reports, teacher assessments of students and archival data were used. Student assessments were completed by teachers at all waves excepting wave 6 (the transition from one funding cycle to the next). Percentages of consented students for whom teachers completed ratings for at each wave (excepting wave 6) ranged from 72% to 93%. Archival ISAT and absenteeism data were collected for the 3 academic years prior to the baseline, as well as throughout the duration of the study.

Data Analyses

Analyses were conducted using Stata version 12.1. Preliminary analyses involved assessing distributions of each outcome and calculating intraclass correlations, Cronbach's alphas, and correlations between the student and teacher variables at Waves 1 and 8.

Primary analyses consisted of estimating multilevel growth-curve models to account for all observations and to model school differences. These were 3-level, time within students within schools, analyses for student-level measures, and 2-level, time within schools, analyses for the aggregated school-level data. We used Stata's "xtmixed" command for normally distributed outcomes, and "xttobit" for outcomes with a positively or negatively skewed distribution (i.e. censored below or above, respectively).[26]

A random-intercept model was fitted using the following equations for student- and school-level analysis, respectively:

$$\hat{Y}_{tij} = \beta_0 + \beta_1(\text{condition}_j) + \beta_2(\text{time}_{tij}) + \beta_3(\text{condition}_j \times \text{time}_{tij}) + \zeta_j + \zeta_{ij} + \varepsilon_{tij} \quad \text{[Student-level]}$$

$$\hat{Y}_{tj} = \beta_{0j} + \beta_1(\text{condition}_j) + \beta_2(\text{year}_{tj}) + \beta_3(\text{year}_{tj} \times \text{condition}_j) + \zeta_j + \varepsilon_{tj} \quad \text{[School-level]}$$

$\hat{Y}_{tij}$ and $\hat{Y}_{tj}$ represent the estimated score on a particular outcome at a particular time *t* (measured as study duration, in years, for student-level models, and as academic year in school-level models). Additionally, *i* represents a student, *j* represents a school, β_0 represents the mean intercept and ζ_j is deviation of a school's mean score from the mean score for all schools. ζ_{ij} is deviation of each student's score from their school's mean, and ε_{tij} and ε_{tj} are the residual. The original models included quadratic terms for time and the interaction of condition by time. Nonsignificant higher order terms were dropped from the model for parsimony, while outcomes with significant quadratic terms (e.g., condition $\times$ time^2) were graphed to facilitate interpretation of growth trajectories (not shown).

When applicable, analyses with student-level variables were run using both the fully reduced random-intercept and random-coefficients models, with the former model nested within the latter model. A likelihood-ratio test was performed to determine whether the random-coefficients model was a better fit for the data. [26]

Due to the power and sample size limitations, and because the *a priori* directional hypothesis was that the *PA* schools would have greater improvements across time, *one-tailed* p-values were used in tests of effects of the *PA* program on school-level outcomes.[27] In the analyses using ISAT weighted averages, 6 matched pairs were retained (for reasons discussed elsewhere); [15] all 7 matched pairs were retained for the endpoint value-added ISAT analysis and for the absenteeism growth-curve analysis. For all outcomes (student-level and school-level) analyzed using growth-curve analyses, effect sizes were calculated using the method described by Lipsey and Wilson. [28]

Sensitivity analyses assessed the robustness of results from the primary analyses. A first approach involved including a "pairs" variable as an additional level in each of the best-fitting models to determine whether adding a fourth level would affect findings. Second, to provide a more conservative test (from a statistical power perspective) of program effects for each outcome, the test statistic provided by Stata (which assumes a large sample size) in the primary analyses (N=14 schools) was compared to the critical value for a two-tailed *t*-distribution with 12 degrees of freedom at a 95% confidence level (2.18).[29]

For student-level data, the possible *moderating effects* of gender and student mobility were examined. The effect of student mobility groups was examined using results from a latent class analysis[15] in which a 5-class solution was found to be the most appropriate fit for the data: 1) stayers (average study duration of 5.72 years, N=158), 2) temporary participants (present for grade 4 and/or 5 only; average study duration of 1.30 years; N=196), 3) late joiners (average study duration of 1.38 years; N=308); 4) early leavers (average study duration of 0.94 years; N=263), and 5) late leavers (average study duration of 3.23 years; N=287); stayers served as the reference group.

RESULTS

The intraclass correlations (ICCs) for the student-level measures were generally low, with none of the ICCs for student-reported and only 1 of the 14 ICCs for teacher-reported outcomes above 0.10. Scale reliabilities (reported above) were generally high, with a clear increase in Cronbach's alphas as students aged. Table 1 shows the correlations between the student and teacher variables at Waves 1 (beginning of grade 3) and 8 (end of grade 8).

Program effects (significant condition × time and condition × time^2 interactions) were present for disaffection with learning (Table 2). Students in *PA* schools started off higher than those in control schools (i.e., more reported disaffection with learning). There was then an overall trend toward a net increase in disaffection with learning by the end of the study period in both *PA* and control schools; the pattern of change was linear in control schools and curvilinear within *PA* schools. As shown in Table 2, there was evidence of a program effect on teacher ratings of student academic motivation in the form of significant condition × time and condition × $time^2$ interactions. For students in *PA* schools, after an initial period of modest decline there was an accelerating increase, whereas for control school students there was a gradually decreasing trend. The net result was notably higher predicted levels of teacher-rated academic motivation for students in *PA* schools. Sensitivity analyses at the pair level supported this finding (results not shown).

With respect to teacher-rated academic ability, a significant condition × time interaction was found in the random-intercept model. In the random-coefficients model, which provided a better fit, the condition × time interaction was not significant ($B = 0.03$, $p < .05$ in random-intercept model; $B = 0.02$, $p > .05$ in random-coefficients model). For both teacher-rating measures, there was no evidence of moderation of program effects by mobility group; gender moderation was observed for academic ability, with *PA* boys being rated higher by teachers than control boys.

Growth-curve analyses for the weighted composite measure of ISAT scores for all students in *PA* and non-*PA* schools did not reveal evidence of a program effect for Reading. There was, however, evidence of marginal program effects for Math (see Table 3). When “pairs” was included in the random-intercept model, this finding remained marginal (results not shown). With respect to demographic subgroups, significant condition × time interactions were seen in Reading performance for African American male students ($B = 0.03$, *one-tailed* $p < .05$). The condition × time interaction remained significant in the pair-level analysis (results not shown). Marginal results (p-values less than or equal to .10) indicative of favorable growth in *PA* schools as compared to control schools, were observed for Reading performance for male students and

African American students, and for Math performance for females and students receiving free or reduced-price lunch.

Endpoint regression analyses for our study cohort, using the value-added metric of the same standardized test, showed significant results in Reading, but not Math. As compared to students in control schools making the grade 7 to 8 transition, students in *PA* schools performed significantly better in reading (B = 1.26, *one-tailed* p=0.013, effect size =0.83, results not shown).

As shown in Table 2, growth-curve analyses showed there was lower absenteeism at *PA* schools than control schools (B = -0.16, *one tailed* p = 0.015). Sensitivity analyses using the pair-level variable and the adjusted degrees of freedom supported these findings (results not shown).

Table 4 shows the estimated means of our outcomes at baseline and endpoint, as well as the effect sizes for each outcome. The largest effect sizes for school-level measures were for absenteeism (effect size = -0.78) and reading performance on the ISAT for African American males (effect size = 1.50). With respect to student-level measures, the largest effect size was observed for teacher ratings of academic motivation (effect size = 0.39).

DISCUSSION

In the Chicago trial of *PA*, the intervention had a positive impact on absenteeism, mitigated a natural increase in disaffection with learning, and *PA* teachers rated their students as experiencing greater growth in academic motivation and ability; these findings are encouraging, as these outcomes are predictors of long-term academic achievement and school completion. [30-32] Favorable growth was also observed with respect to ISAT Reading and Math performance, particularly for African American males and students receiving free or reduced-price lunch. Socioeconomic background (i.e. low-income), gender (i.e. being male) and ethnicity (i.e.

African-American, Hispanic, and Native American youth) are known predictors of school drop-out, and school drop-out is associated with a multitude of negative outcomes. [31] As prevention programs can only influence those factors amenable to change (e.g., motivation to learn), it is encouraging that this trial also demonstrated improvements in test scores for these high-risk groups.

The impact on academic-related outcomes observed in this study may be attributed to a number of factors. For example, the skills fostered by the *PA* program (e.g., problem solving, self-control, emotional regulation, and attention), and lesson plans focusing on improving motivation to learn and do well in school, may in part explain the observed results.[5] In addition, the promotion of positive behaviors may have resulted in less time being spent by teachers on classroom management and, subsequently, more time devoted to interactive strategies that create an intellectually stimulating environment.[5] Moreover, the impact on academics may have been mediated through improvements in attachment to school and teachers.

This study is the first to examine the academic impact of *PA* in a low-income, urban setting, and supplements Snyder and colleagues'[14] findings on the academic impact of *PA* in Hawai'i by including data from students and teachers of students in the elementary- and middle-school grades. The study also adds to the research of Madsen and colleagues',[33] who evaluated the impact of a physical-activity focused, school-based, Positive Youth Development program in low-income Bay Area California schools using a quasi-experimental time series design; namely, the researchers found that each additional year of exposure to the program resulted in significantly higher scores in meaningful participation in school and academic-related goals and aspirations of youth. In the current study, for those measures with significant program effects, the effect size for disaffection with learning (effect size = -0.19) was smaller than the effect sizes for academic outcomes reported by the research teams led by Farahmand[3] and Durlak.[4] On the other hand, other measures in this study (e.g., academic motivation, absenteeism, ISAT Math results) had larger effect sizes than those observed in the aforementioned studies.

Limitations

This study is not without its limitations. Student and teacher-reports on academic measures are subject to social desirability bias; this potential bias was addressed by supplementing student and teacher reports with archival measures representing the actual performance of students on standardized tests. Another possible limitation of the study is that students in the intervention group may have acted differently because they knew they were receiving the *PA* program, a phenomenon known as the Hawthorne effect. This limitation was addressed through the trial's use of a control group of students and teachers who were also aware they were being observed as part of a study. With respect to external validity, the findings are generalizable only to similar schools (i.e. low-income, urban schools) that would self-select to participate in a trial of this nature. The small number of pairs and schools (i.e., 7 and 14, respectively) could influence statistical power; however, that significant findings were found in primary and sensitivity analyses suggest that our findings are robust. Additionally, student mobility led to high turnover of students, which is problematic as it can become difficult to determine whether observed effects can be attributed to the intervention or differential attrition.[25] A recent approach to analyzing mobility patterns is latent class analysis (LCA),[34, 35] and the present study contributes to the LCA literature by examining students who enter a study, not just those who exit;[15] program effects were not found to differ by mobility class.

Limitations notwithstanding, the present study has several strengths. The longitudinal nature of this RCT allowed examination of school performance across 6 years, encompassing both elementary- and middle-school grades. The data from multiple sources as well as the sensitivity analyses provide confidence in study findings. In addition to standardized test performance, our study also reported on theoretically-expected mediators of academic success (e.g., disaffection with learning). Moreover, this study involved a sample of students in a high-risk setting. Thus, policymakers aiming to alleviate educational disparities should use scientific data from this and other evidence-based studies to advocate for comprehensive school-based SECD programming.

Conclusions

Findings from the present study reinforce prior findings that SECD-like programs can improve academic achievement as well as improve student behavior and health. Future studies should determine the mechanism by which SECD programs such as *Positive Action* improve academic outcomes (e.g., mediation through factors that SECD programs seek to foster, such as attachment with teacher and school, improved school climate, emotional regulation, attention, executive function, and increased self-control). Future research could also supplement student and teacher reports by gathering data from parents that may influence academic performance (e.g., parent's highest level of education).

IMPLICATIONS FOR SCHOOL HEALTH

In an era where increased pressures to "teach to the test" may lead school officials to feel as though they have neither the time nor money to invest in evidence-based prevention programming,[36] there is an increasing need to demonstrate the impact that multifaceted prevention programs can have on academic performance and student and community wellness.[37] When taken together with preliminary research showing the impact of this trial on health behaviors,[38] results from this study demonstrate the possibility of addressing the proverbial "two birds” (i.e. health and academics) with “one stone” (i.e. school-based social-emotional and character development programs).

Human Subjects Approval Statement

The research presented herein was approved by the institutional review boards of Oregon State University and the University of Illinois at Chicago, the Research Review Board at Chicago Public Schools and the Public/Private Ventures Institutional Review Board for Mathematica Policy Research Inc.

Acknowledgements

The findings reported here are based on research funded by grants from the Institute of Education Sciences (IES), U.S. Department of Education:R305L030072, R305L030004 and R305A080253 to the University of Illinois at Chicago (2003-2005) and Oregon State University (2005-2012). The SACD Research Program is a collaboration among IES, the Centers for Disease Control and Prevention’s (CDC) 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). Moreover, the preparation of this manuscript was supported, in part by the National Institute on Alcohol Abuse and Alcoholism(NIAAAT32 AA014125).

The SACD research program includes multi-program evaluation data collected by MPR and complementary research study data collected by each grantee. The findings reported here are based only on the Chicago portion of the multi-program data and the complementary research data collected by the University of Illinois at Chicago and Oregon State University (Brian Flay, Principal Investigator) under the SACD program.

The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the Institute of Education Sciences, CDC, MPR, or every Consortium member, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government. Correspondence concerning this article should be addressed to NBavarian@berkeley.edu or Brian.Flay@oregonstate.edu

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 Specialized Services, especially Drs. Renee Grant-Mitchell and Inez Drummond, for their invaluable support of this research.

Notice of potential conflict of interest: The research described herein was done using the program, the training, and technical support of *Positive Action, Inc*. in which Dr. Flay's spouse holds a significant financial interest. Issues regarding conflict of interest were reported to the relevant institutions and appropriately managed following the institutional guidelines.

REFERENCES

1. Braveman PA, Cubbin C, Egerter S, et al. Socioeconomic disparities in health in the United States: what the patterns tell us. *Am J Public Health*. 2010;100(S1):S186-196.

2. Aud S, Fox MA, KewalRamani A. *Status And Trends In The Education Of Racial And Ethnic Groups. NCES 2010-015*. National Center for Education Statistics: Washington, DC; 2010

3. Farahmand FK, Grant KE, Polo AJ, et al. School-based mental health and behavioral programs for low-income, urban youth: a systematic and meta-analytic review. *Clin Psychol*. 2011;18(4):372-390.

4. Durlak JA, Weissberg RP, Dymnicki AB, et al. The impact of enhancing students' social and emotional learning: a meta-analysis of school-based universal interventions. *Child Dev*. 2011;82(1):405-432.

5. Flay BR, Allred CG. The Positive Action program: improving academics, behavior, and character by teaching comprehensive skills for successful learning and living. In: Lovat T, Toomey R, Clement N, eds. *International Research Handbook On Values Education And Student Wellbeing*. Dordrecht: Springer Netherlands; 2010:471-501.

6. DuBois DL, Flay BR, Fagen MC. Self-esteem enhancement theory. In: DiClemente RJ, Crosby RA, Kegler MC, eds. *Emerging Theories In Health Promotion Practice And Research.* 2nd ed. San Francisco, CA: Jossey-Bass; 2009:97-130.

7. Purkey WW. *Self-concept and School Achievement*. Englewood Cliffs, NJ: Prentice-Hall; 1970.

8. Purkey WW, Novak J. *Inviting School Success: A Self-concept Approach To Teaching And Learning*. Belmont, CA: Wadsworth; 1970.

9. Flay BR, Petraitis, J. The theory of triadic influence: a new theory of health behavior with implications for preventive interventions. In: Albrecht G, ed. *Advances in Medical Sociology.* Volume 4. Greenwich, CT: JAI Press; 1994:19-44.

10. Flay BR, Snyder F, Petraitis J. The theory of triadic influence. In: DiClemente RJ, Crosby RA, Kegler MC, eds. *Emerging Theories In Health Promotion Practice And Research.* 2nd ed. San Francisco, CA: Jossey-Bass; 2009:451-510.

11. Washburn IJ, Acock A, Vuchinich S, et al. Effects of a social-emotional and character development program on the trajectory of behaviors associated with social-emotional and character development: findings from three randomized trials. *Prev Sci*. 2011;12(3):314-323.

12. Snyder FJ, Vuchinich S, Acock A, et al. Improving elementary school quality through the use of a social-emotional and character development program: a matched-pair, cluster-randomized, controlled trial in Hawai'i. *J Sch Health*. 2012;82(1):11-20.

13. Beets MW, Flay BR, Vuchinich S, et al. Use of a social and character development program to prevent substance use, violent behaviors, and sexual activity among elementary-school students in Hawaii. *Am J Public Health*. 2009;99(8):1438-1445.

14. Snyder F, Vuchinich S, Acock A, et al. Impact of the Positive Action program on school-level indicators of academic achievement, absenteeism, and disciplinary outcomes: a matched-pair, cluster randomized, controlled trial. *J Res Educ Eff*. 2010;3(1):26-55.

15. Lewis KM, DuBois DL, Ji P, et al. Meeting the challenges of longitudinal cluster-based trials in schools: Lessons from the Chicago trial of Positive Action. Oregon State University. In review.

16. Schochet P, Novak T. Computer program to construct school pairs. Unpublished manuscript. Mathematica Policy Research. 2013.

17. Ji P, DuBois DL, Flay BR, et al. “Congratulations, you have been randomized into the control group!(?)”: issues to consider when recruiting schools for matched-pair randomized control trials of prevention programs. *J Sch Health*. 2008;78(3):131-139.

18. Li K-K, Washburn I, DuBois DL, et al. Effects of the Positive Action programme on problem behaviours in elementary school students: a matched-pair randomised control trial in Chicago. *Psychol Health*. 2011;26(2):187-204.

19. Furrer C, Skinner E. Sense of relatedness as a factor in children’s academic engagement and performance. *J Educ Psychol*. 2003;95(1):148-162.

20. Gresham FM, Elliot SN. *Social skills rating system*. Circle Pines, MN: American Guidance Service; 1990.

21. Achenbach TM. *Manual for the Teacher’s Report Form and 1991 profile*. Burlington, VT: University of Vermont, Department of Psychiatry; 1991.

22. Somers M, Zhu P, Wong E. *Whether and How to Use State Tests to Measure Student Achievement in a Multi-State Randomized Experiment: An Empirical Assessment Based on Four Recent Evaluations* (NCES 2012-4015). U. S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office; 2011.

23. Chicago Public Schools. Chicago Public Schools. Available at: http://www.cps.edu/Pages/home.aspx. Accessed March 6, 2012.

24. Chicago Public Schools. *FAQ on the value-added metric*. Available at: http://research.cps.k12.il.us/export/sites/default/accountweb/Research/ValueAdded/valueadd_faq.pdf. Accessed March 6, 2012.

25. Vuchinich S, Flay BR, Aber L, et al. Person mobility in the design and analysis of cluster-randomized cohort prevention trials. *Prev Sci*. 2012;13(3):300-313.

26. Rabe-Hesketh S, Skrondal A. *Multilevel And Longitudinal Modeling Using Stata.* 2nd ed. College Station, TX: Stata Press; 2008.

27. Knottnerus JA, Bouter LM. The ethics of sample size: two-sided testing and one-sided thinking. *J Clin Epidemiol*. 2001;54(2):109-110.

28. Lipsey MW, Wilson DB. *Practical Meta-analysis*. Thousand Oaks, CA: Sage; 2001.

29. Raudenbush SW, Bryk AS. *Hierarchical Linear Models: Applications And Data Analysis Methods.* 2nd ed. Newbury Park, CA: Sage; 2002.

30. Suh S, Suh J, Houston I. Predictors of categorical at-risk high school dropouts. *J Couns Dev*. 2007;85(2):196-203.

31. Lehr CA, Johnson DR, Bremer CD, et al. *Essential Tools: Increasing Rates Of School Completion: Moving From Policy And Research To Practice*. Minneapolis.MN: ICI Publications Office; 2004.

32. Caraway K, Tucker CM, Reinke WM, et al. Self-efficacy, goal orientation, and fear of failure as predictors of school engagement in high school students. *Psychol Schools*. 2003;40(4):417-427.

33. Madsen KA, Hicks K, Thompson H. Physical activity and positive youth development: impact of a school-based program. *J Sch Health*. 2011;81(8):462-470

34. Beunckens C, Molenberghs G, Verbeke G, et al. Latent-class mixture model for incomplete longitudinal Gaussian data. *Biometrics.* 2008;64(1):96-105.

35. Roy J. Modeling longitudinal data with nonignorable dropouts using a latent dropout class model. *Biometrics.* 2003;59(4):829-836.

36. Kaftarian S, Robinson E, Compton W, et al. Blending prevention research and practice in schools: critical issues and suggestions. *Prev Sci*. 2004;5(1):1-3.

37. Greenberg MT. Current and future challenges in school-based prevention: the researcher perspective. *Prev Sci*. 2004;5(1):5-13.

38. Bavarian N, Lewis KM, Acock A, et al. Effects of a social-emotional and character development program on adolescent health behaviors and outcomes: a matched-pair, cluster-randomized controlled trial. Oregon State University. In review.

Table 1. **Youth and teacher reports of academic outcomes: Correlations at Wave 1 (above the diagonal, N=603) and Wave 8 (below the diagonal, N=335)**

Variables	1	2	3	4	5	6	7
Student Self Reports							
1. Disaffection with Learning	–	-0.04	-0.31**	-0.29**	-0.32**	-0.29**	-0.27**
2. Self Reported Grades	-0.23**	–	0.24**	0.17**	0.21**	0.21**	0.20**
Teacher Ratings of Students							
3. Reading	-0.03	0.33**	–	0.84**	0.89**	0.93**	0.71**
4. Math	-0.06	0.37**	0.93**	–	0.84**	0.87**	0.67**
5. Intellectual Functioning	-0.01	0.29**	0.91**	0.89**	–	0.91**	0.71**
6. Academic Performance	-0.07	0.34**	0.93**	0.93**	0.92**	–	0.73**
7. Academic Motivation	-0.09	0.44*	0.67**	0.67**	0.64**	0.68**	–

*p<0.05;**p<0.01

Table 2. **Multilevel growth-curve model estimates for student-level measures (N=1,170 students) and aggregated school-level (N=14 schools) archival measures**

Measure	Model Run	Intercept	Time	Time^2	Condition (0 = Non-*PA* 1=*PA*)	Condition × Time	Condition × Time^2
		B (SE)	B (SE)	B (SE)	B (SE)	B (SE)	B (SE)
<u>Student Self Reports</u>							
Disaffection with Learning	Random Intercept	1.69 (0.06)**	0.03 (0.04)	0.01(0.01)	0.15 (0.08)*	-0.20(0.06)**	0.03 (0.01)**
Self Reported Grades	Random Intercept	7.89 (0.12)**	-0.81(0.07)**	0.11(0.01)**	0.10(0.17)	0.01(0.03)	---
<u>Teacher Ratings of Students</u>							
Academic Performance [a]	Random Coefficients	2.62 (0.06)**	-0.05(0.03)*	0.02(0.005)**	-0.06(0.08)	0.02(0.02)	---
Academic Motivation	Random Coefficients	3.01(0.07)**	0.04(0.04)	-0.01(0.01)	0.05(0.10)	-0.12(0.06)*	0.03(0.01)**
<u>School Level Archival Data</u> [b]							
Absenteeism	Random Intercept	6.76 (0.56)**	0.03 (0.05)	---	0.43 (0.65)	-0.16 (0.07)*	---

a: For the random-intercept model, the condition × time interaction is significant at the .05 level ($B = 0.03$, $p < .05$).

b: For school level measures, time variable created using academic year, rather than time since implementation of intervention. Also, the *one-tailed* p-value is reported for school-level measures.

+p < .10; * p < .05; **p < .01

Table 3. **Multilevel random-intercept growth-curve model estimates for standardized academic test scores [a] (N=12 schools)**

Variables	Intercept [b]	Time [b]	$Time^{2}$ [b]	Condition (0 = Non-PA; 1= PA)	Condition × Time	Condition × Time
	B (SE)	B (SE)	B (SE)	B (SE)	B (SE)	One-tailed *p*-value
Reading						
All Students (Grades 3-8 Combined)	2.26 (0.07)	0.17 (0.01)	-0.02 (0.002)	0.04 (0.10)	0.01 (0.01)	0.16
Sub-Groups						
Males	2.22 (0.07)	0.16 (0.02)	-0.02 (0.003)	-0.001 (0.10)	0.01 (0.01)	0.10
Females	2.30 (0.07)	0.17 (0.02)	-0.02 (0.003)	0.07 (0.10)	0.004 (0.01)	0.35
African Americans	2.20 (0.06)	0.15 (0.02)	-0.01 (0.003)	0.05 (0.08)	0.01 (0.01)	0.10
African American Females	2.21 (0.05)	0.17 (0.02)	-0.02 (0.003)	0.13 (0.07)	-0.01 (0.01)	0.23
African American Males	2.17 (0.07)	0.16 (0.03)	-0.02 (0.005)	-0.02 (0.10)	0.03 (0.01)	0.02
Free or Reduced Price Lunch	2.25 (0.07)	0.17 (0.01)	-0.02 (0.002)	0.03 (0.09)	0.01 (0.01)	0.18
Math						
All Students (Grades 3-8 Combined)	2.15 (0.08)	0.24 (0.02)	-0.03 (0.003)	0.04 (0.12)	0.01 (0.01)	0.07
Sub-Groups						
Males	2.12 (0.09)	0.24 (0.02)	-0.03 (0.004)	0.04 (0.12)	0.01 (0.01)	0.12
Females	2.18 (0.08)	0.24 (0.02)	-0.03 (0.004)	0.04 (0.11)	0.02 (0.01)	0.09
African Americans	2.06 (0.06)	0.23 (0.02)	-0.02 (0.004)	0.06 (0.08)	0.02 (0.01)	0.11
African American Females	2.09 (0.07)	0.23 (0.03)	-0.02 (0.004)	0.10 (0.09)	0.02 (0.01)	0.11
African American Males	2.02 (0.07)	0.25 (0.03)	-0.03 (0.005)	0.04 (0.09)	0.02 (0.01)	0.12
Free or Reduced Price Lunch	2.15 (0.08)	0.24 (0.02)	-0.03 (0.003)	0.04 (0.11)	0.01 (0.01)	0.07

a: The average of values from 2000/2001 through 2002/2003 was used as the estimate of baseline levels.

b: The coefficients for Intercept, Time, and $Time^2$ were all significant at the .01 level, except the $time^2$ coefficient for African American Males, which was significant at the .05 level.

Table 4. **Estimated means and effect sizes for student- and school-level data**

			Wave 1		Wave 8		
Measure	Response Options	Model Run	Control	PA	Control	PA	Effect Size [a]
<u>Student Self Reports</u>							
Disaffection with Learning	1 to 4	Random Intercept	1.69	1.85	2.19	2.19	-0.19
Self Reported Grades	1 to 9	Random Intercept	7.89	7.98	6.67	6.81	0.02
<u>Teacher Ratings of Students</u>							
Academic Ability	1 to 5	Random Coefficients	2.63	2.57	2.84	2.91	0.14
Academic Motivation	1 to 5	Random Coefficients	3.01	3.06	2.80	3.24	0.39
<u>School Level Archival Data[b]</u>							
Absenteeism	0 to 100	Random Intercept	6.76	6.33	6.95	5.58	-0.78
ISATs-Reading	1 to 4						
All Students (Grades 3-8 Combined)	1 to 4	Random Intercept	2.26	2.29	2.64	2.72	0.22
Males	1 to 4	Random Intercept	2.22	2.22	2.60	2.66	0.33
Females	1 to 4	Random Intercept	2.30	2.37	2.68	2.78	0.11
African Americans	1 to 4	Random Intercept	2.20	2.25	2.62	2.74	0.50
African American Females	1 to 4	Random Intercept	2.21	2.34	2.66	2.74	-0.54
African American Males	1 to 4	Random Intercept	2.17	2.15	2.57	2.72	1.50
Free or Reduced Price Lunch	1 to 4	Random Intercept	2.25	2.28	2.63	2.70	0.23
ISATs-Math	1 to 4	Random Intercept					
All Students (Grades 3-8 Combined)	1 to 4	Random Intercept	2.15	2.19	2.67	2.79	0.38
Males	1 to 4	Random Intercept	2.12	2.17	2.67	2.79	0.31
Females	1 to 4	Random Intercept	2.18	2.22	2.68	2.81	0.41
African Americans	1 to 4	Random Intercept	2.06	2.12	2.62	2.77	0.55
African American Females	1 to 4	Random Intercept	2.09	2.19	2.61	2.80	0.69
African American Males	1 to 4	Random Intercept	2.02	2.07	2.62	2.76	0.63
Free or Reduced Price Lunch	1 to 4	Random Intercept	2.15	2.19	2.67	2.79	0.42

a: Effect size calculations made using estimated means. Namely, the estimated mean difference at the baseline was subtracted from the estimated mean difference at the end point to obtain the difference of differences, and this value was then divided by the pooled standard deviation at baseline.

b: For school level measures, time variable created using academic year, rather than time since implementation of the *Positive Action* intervention.