

## AN ABSTRACT OF THE DISSERTATION OF

Michael W. Beets for the degree of Doctor of Philosophy in Public Health presented on April 6<sup>th</sup>, 2007.

Title: Factors Associated with the Implementation Fidelity of a School-Based Social and Character Development Program: Findings from the *Positive Action* Program, Hawai'i

Abstract approved:

---

Brian R. Flay

Implementation fidelity (dosage, adherence, quality of delivery, responsiveness, program differentiation) is the degree to which a program was carried out to its original intent. The following three studies examined the implementation fidelity of a school-based health promotion social and character development (SACD) program. The, *Positive Action* (PA) program was a 5yr effectiveness trial conducted in 20 elementary (10 control, 10 intervention) schools in Hawai'i from 2001-2006. Study 1 examined teacher- and school-level factors that influenced fidelity of program implementation using a Diffusion of Innovations framework. Implementation was defined as "the amount of the programs' curriculum delivered, and use of program-specific materials in the classroom and in relation to school-wide activities." Teachers' completed year-end process-evaluation reports for year 2 (N = 171) and 3 (N = 191). Classroom and school-wide material usage were influenced by the amount of the curriculum delivered, which were influenced by teachers' attitudes toward the program and teachers' beliefs about SACD. These, in turn, were shaped by teachers' perceptions of school climate. Study 2 examined the impact of student responsiveness

ratings on prosocial behaviors and feelings, and substance abuse rates. Elementary students ( $N = 2,926$ ) completed 4 year-end surveys assessing prosocial behaviors and feelings (yr1-4), responsiveness towards PA (yr3 and 4), and substance abuse (yr4,  $n = 458$ ). Longitudinal path models indicated responsiveness was related to positive behaviors and feelings, and reduced substance abuse rates at year 4. Study 3 examined the extent to which control schools self-initiated or received programming of similar content to the intervention being evaluated (program differentiation). One school leader from each school ( $N = 18$ ), responded to a questionnaire regarding information about the number of SCD programs conducted over the prior 3 academic years. Control schools reported an average of  $8.0 \pm 5.1$  SCD programs vs.  $3.6 \pm 3.6$  (in addition to PA) reported by intervention schools. In conclusion, program developers need to consider: a) the importance of a supportive school climate on implementation fidelity; b) why students may or may not like a given program; and c) the self-initiated programming occurring in control schools. Greater attention to these will assist in the interpretation of positive and negative outcomes from school-based prevention programs.

Factors Associated with the Implementation Fidelity of a School-Based Social and Character Development Program: Findings from the *Positive Action* Program, Hawai'i

by  
Michael W. Beets

A DISSERTATION

submitted to

Oregon State University

in partial fulfillment of  
the requirements for the  
degree of

Doctor of Philosophy

Presented April 6<sup>th</sup>, 2007  
Commencement June 2007

Doctor of Philosophy dissertation of Michael W. Beets presented on April 6<sup>th</sup>, 2007.

APPROVED:

---

Major Professor, representing Public Health

---

Chair of the Department of Public Health

---

Dean of the Graduate School

I understand that my dissertation will become part of the permanent collection of Oregon State University libraries. My signature below authorizes release of my dissertation to any reader upon request.

---

Michael W. Beets, Author

## ACKNOWLEDGEMENTS

Without out the guidance provided by three very special people in my life, this process would not have been nearly as enjoyable and rewarding. The first is my mother, Dr. Leanne Beets, who supported all my decisions and kindly lent her ear for all the peaks and valleys I experienced along this path. And to Dr. Kenneth Pitetti, who inadvertently became my major advisor during my Masters of Exercise Science, and since then has become more so by being my friend and a father figure. And lastly to Dr. Bradley Cardinal, who was my unofficial major advisor throughout my time at Oregon State University and has been that bright light, in the sometimes darkness, of doctoral work. I would like to further acknowledge the guidance and the “opening of doors” Dr. Brian Flay has provided during the last two years of my doctoral work. Without his appearance at OSU, I would likely not be on the path I am today.

## CONTRIBUTION OF AUTHORS

Michael W. Beets, M.Ed., M.P.H. conceptualized, drafted, and analyzed all data analyses presented.

Dr. Brian R. Flay was the Principal Investigator for the National Institute on Drug Abuse grant #R01-DA13474 titled: *The Positive Action Program: Outcomes and Mediators*. He and co-investigators conceptualized the study, measures and data collection procedures.

Dr. Samuel Vuchinich, Dr. Alan C. Acock, and Dr. Carol Allred provided editorial comments and suggestions on the interpretation of the findings. Dr. Allred also developed the *Positive Action* program and trained the teachers to implement it.

## TABLE OF CONTENTS

	<u>Page</u>
Chapter 1. Introduction to Implementation .....	1
Specific Aims .....	2
Research Questions .....	3
Background/Significance .....	5
What is implementation? .....	6
The importance of measuring implementation .....	11
Factors that affect implementation .....	13
Summary .....	16
Chapter 2. Manuscript 1 .....	18
School climate and teachers' beliefs and attitudes associated with implementation of the Positive Action program: A Diffusion of Innovations Model .....	19
Chapter 3. Manuscript 2 .....	42
Linking student responsiveness to prosocial outcomes: Findings from the multiyear effectiveness trial of the Positive Action program .....	43
Chapter 4. Manuscript 3 .....	68
Where have all the control groups gone? .....	69
Chapter 5. General Conclusions .....	81
Bibliography .....	93

## LIST OF FIGURES

<u>Figure</u>	<u>Page</u>
1. Teacher-student-outcomes program fidelity model of proposed effects.....	10
2. Conceptual model of teacher, administrative, and contextual characteristics influencing implementation fidelity .....	17
3. (Figure 1 of manuscript 1) Year 2 and 3 (estimates in parentheses) model depicting school climate (school connectedness and administrative support) and teacher characteristics (beliefs regarding their responsibility to teacher social and character development concepts and attitude towards <i>Positive Action</i> ) related to the amount of the <i>Positive Action</i> curriculum delivered and corresponding program material usage in the classroom and school-wide. ....	41
4. (Figure 1 of manuscript 2) Final estimated model of hypothesized relationships of student responsiveness, program exposure, and program outcomes, behaviors (Do) and feelings (Feel). ....	66
5. (Figure 2 of manuscript 2) Subgroup model depicting hypothesized relationships of student responsiveness, program exposure, and program outcomes, behaviors (Do) and feelings (Feel), to substance abuse usage rates. ....	67
6. Teacher-student-outcomes program fidelity model of proposed effects.....	83
7. Conceptual model of teacher, administrative, and contextual characteristics influencing implementation fidelity .....	85
8. Expanded conceptual model of teacher, administrative, and contextual characteristics influencing implementation fidelity .....	86



## LIST OF TABLES

<u>Table</u>	<u>Page</u>
1. Levels of program implementation fidelity dimensions* .....	7
2. Rationale for measuring implementation* .....	13
3. (Table 1 of manuscript 1) Alpha reliabilities of full and parceled scales for Year 2 and 3 .....	37
4. (Table 2 of manuscript 1) Descriptive statistics of scale scores of latent constructs for Year 2 and 3 .....	38
5. (Table 3 of manuscript 1) Summary of indirect effects from school climate to implementation of the <i>Positive Action</i> program.....	39
6. (Table 4 of manuscript 1) Descriptive characteristics of teachers divided into two groups (1) upper 25% and (2) lower 25% distribution of school climate and attitude towards <i>Positive Action</i> composite scores for implementation Years 2 and 3 .....	40
7. (Table 1 of manuscript 2) Descriptive statistics across assessment years 1 through 4 for boys and girls .....	61
8. (Table 2 of manuscript 2) Cross-sectional relationship between student responsiveness and behaviors (Do) and feelings (Feel) for year 3 and 4.....	62
9. (Table 3 of manuscript 2) Cross-sectional logistic regression estimates for substance abuse (year 4) and program outcomes ( $n = 442$ ) .....	63
10. (Table 4 of manuscript 2) Estimate indirect effects from the full and sub-sample longitudinal structural equation models (Figure 2 and 3).....	64
11. (Table 5 of manuscript 2) Effects of gender (boys) on behaviors (Do), feelings (Feel), student responsiveness, and substance abuse .....	65
12. (Table 1 Manuscript 3) Responses to self-initiated SACD activities questionnaire for control and intervention schools .....	79
13. (Table 2 Manuscript 3) Self-initiated SACD programs and activities reported by control and experimental schools .....	80

## Chapter 1. Introduction to Implementation

### Specific Aims

The ability to determine how to achieve successful outcomes from school-based primary prevention programs in naturalistic settings is a concern of health promotion and disease prevention researchers and a necessary criterion in establishing whether a program is ready for widespread dissemination (Dane & Schneider, 1998; Domitrovich & Greenberg, 2000; Durlak, 1998; Flay, Biglan, Boruch, Castro, Gottfredson, Kellam, Mocicki, Schinke, Valentine, & Ji, 2005). Of importance is the measurement of how a program was implemented and under what circumstances it works; in that such information can be used by future implementers to assist in effective adoption and reproducibility. Measures of implementation can further be used to determine why programs fail to produce expected outcomes and whether this was attributable to program failure or failure of implementation. The latter is known as a Type III Error where null findings are attributed to a failure of the program, when, in fact, the failure was due to poor implementation (e.g., reduced dosage, alterations in materials) (Basch, Sliepcevich, Gold, Duncan, & Kolbe, 1985; Harachi, Abbott, Catalano, Haggerty, & Fleming, 1999; Scanlon, Horst, Nay, Schmidt, & Waller, 1977).

The *Positive Action* program is a multicomponent school-based program designed to target the reduction of students' health-compromising and other negative behaviors (e.g., substance abuse, violence, disciplinary referrals, suspensions), while simultaneously enhancing health-enhancing and other positive behaviors (e.g., honesty, time management) and behavioral attributions directed at the self (e.g., self-responsibility, positive self-concept) and social relationships (e.g., conflict resolution, respect/kindness), with these leading to improved school-related performance (e.g., improved attendance, academic achievement). Results from prior studies examining school level archival data (i.e., student data aggregated at the school level), have indicated that *Positive Action* beneficially impacts academic achievement and improves student behaviors (Flay & Allred, 2003; Flay, Allred, & Ordway, 2001). Yet, despite the positive outcomes associated with *Positive Action*, it has not been systematically evaluated for program fidelity; that is, whether the program components that are supposed to be delivered are actually delivered (i.e., adherence),

what the dosage of the program administered has been (i.e., frequency/duration of program delivered), how well the program was received (i.e., responsiveness), and to what extent characteristics of the individual delivering the program (e.g., attitudes) and setting (e.g., administrative support, school climate) differentially effect the delivery of the program as intended (Dane & Schneider, 1998).

The overall objective of the following three studies is to analyze the implementation of the *Positive Action* program using data from a school-based randomized trial study (Hawai'i). Setting apart these studies from those previously conducted (Flay & Allred, 2003; Flay et al., 2001) are the individual-level data specific to the implementation of the *Positive Action* program (e.g., dosage, adherence, responsiveness) collected from multiple sources – teachers and students.

### Research Questions

The three studies address the following research questions and corresponding hypotheses:

#### Manuscript #1

*“What teacher- and school-related factors influence some teachers to implement Positive Action with greater fidelity than others?”* Based on the diffusion of innovations model, along with the findings from prior research, the following hypotheses were proposed:

- a) Utilization of the *Positive Action* program materials will be enhanced by the following teacher-level process: positive teacher attitudes toward *Positive Action* will positively influence teaching of the program's curriculum which, in turn, will lead to increased material usage, both within the classroom and across the school as a whole
- b) Teacher attitudes toward *Positive Action* will be influenced by their beliefs regarding their responsibility to teach prevention/SACD concepts
- c) School climate will directly affect teacher-level processes (i.e., beliefs and attitudes) and the delivery of the curriculum, along with utilization of *Positive*

*Action* materials school-wide, thereby exerting both direct and indirect effects on implementation

## Manuscript #2

*“What is the role of student responsiveness (i.e., customer satisfaction) to prosocial behaviors (and the feelings associated with performing those behaviors) and substance abuse outcomes for elementary students participating in four years of the school-based prevention program, Positive Action.”*

Based on prior studies of responsiveness and theories of consumer satisfaction, hypotheses were developed and tested in both cross-sectional and longitudinal models:

- a) student responsiveness will be positively related to program exposure
- b) student responsiveness, in turn, will be positively related to increased prosocial behaviors and feelings, and indirectly related (mediated by prosocial behaviors and feelings) to reduced substance abuse rates of elementary students

## Manuscript #3

*“To what extent were control schools self-initiating or receiving programming of similar content to the Positive Action program (i.e., program differentiation) during the multiyear effectiveness trial?”*

It was hypothesized that the control schools were self-initiating programs that may bring the control schools up to a level of performance comparable with the intervention schools. Thus, based on prior studies indicating control schools may self-initiate programming that mimics experimental conditions, even when asked to refrain from doing such, the following questions were proposed.

- a) to what extent were control schools conducting social and character development programs
- b) how often (weeks) and amount of time (hours devoted to teaching/week) were the programs being conducted
- c) if programming was taking place, were teachers receiving training to deliver the program(s) with fidelity

### Background/Significance

Over the past decades, considerable emphasis has been placed on the design, implementation, and evaluation of school-based primary prevention programs aimed at decreasing health-compromising behaviors (e.g., substance abuse, delinquency, physical aggression) while also improving positive student behaviors (e.g., academic performance, self-concept) and school level outcomes (e.g., climate) (Flay & Collins, 2005). Recently, criteria have been outlined for establishing the “value” of a program under controlled (i.e., efficacy) and naturalistic (i.e., effectiveness) settings which specify that in order for a program to be considered ready for widespread dissemination, detailed analyses must be conducted on *how* and *why* a given program works and under what conditions (Flay et al., 2005).

Numerous school-based programs have been developed (e.g., Adolescent Alcohol Prevention Trial, Comer’s School Development Program, Promoting Alternative THinking Strategies, Midwestern Prevention Project) (Cook, Murphy, & Hunt, 2000; Hansen, Graham, Wolkenstein, & Rohrbach, 1991; Pentz, Trebow, Hansen, MacKinnon, Dwyer, Johnson, Flay, Daniels, & Cormack, 1990) and “proven” to beneficially impact student/school level outcomes. Reports on these programs have included assessments specific to the delivery of program components to ensure it was implemented in accordance with the original intentions/design (Basch et al., 1985; Domitrovich & Greenberg, 2000). Information has also detailed under what circumstances a program either achieved success or did not (Durlak, 1998; Harachi et al., 1999). The *Positive Action* program, on the contrary, has yet to be systematically investigated in terms of implementation practices and their affect on program outcomes. Primary prevention programs differ fundamentally on their guiding theoretical basis, targeted mechanisms of change, intended audience, and program delivery. Because of this, it is essential that each individual program’s implementation be evaluated in order to attribute observed outcomes with the program (internal and external validity) and to guide future implementers in adoption and replication (Durlak, 1998).

### *What is implementation?*<sup>1</sup>

Implementation, in general terms, is the degree to which a planned program was carried out to its original intent. Treatment integrity, program integrity, fidelity, and adherence are terms used interchangeably throughout the literature to refer to implementation (Basch, 1984; Basch et al., 1985; Dane & Schneider, 1998; Domitrovich & Greenberg, 2000; Durlak, 1998), but as will be discussed below, implementation consists of multiple facets, each examining an integral part of the overall picture of how a program was conducted (Dane & Schneider, 1998).

When programs are conducted under naturalistic conditions (e.g., by teachers in schools), implementation information is used to inform evaluators about how well a program was actually delivered and why it may or may not have achieved desired results. In their extensive review of implementation measures in school-based primary prevention programs, Dane and Schneider (1998) conceptualized implementation fidelity<sup>2</sup> as a multidimensional construct, consisting of five components: *adherence*, *exposure*, *quality of delivery*, *responsiveness*, and *program differentiation* (see Table 1) (Dane & Schneider, 1998). Each dimension offers information for evaluators to determine what implementers ultimately provided to the audience (i.e., adherence), how much (i.e., exposure) and how well (i.e., quality of delivery) it was provided (i.e., exposure), what the intended audience thought of what was provided (i.e., responsiveness), and whether similar provisions were taking place in the control conditions (i.e., program differentiation).

---

<sup>1</sup> The author recognizes prevention programs can be conducted within numerous contexts, such as worksite health promotion programs, community recreation organizations, school-based prevention programs, among others. Reference to who the *implementers* and *intended audience* are can, therefore, take many forms given context-specific characteristics. Throughout this paper, when referring to implementers and the intended audience the author is specifically referring to Teachers and Students, respectively.

<sup>2</sup> *Fidelity* and *integrity*, are used interchangeably throughout the implementation literature to refer to the degree to which specific procedures outlined by a given program are delivered as intended. For the purpose of this discussion, *fidelity* will be used as a global construct to refer to the various components/dimensions of implementation as indicated by Dane and Schneider (1998) and presented in Table 1.

Table 1. Levels of program implementation fidelity dimensions\*

Level	Implementation Fidelity	Measures
Audience	Responsiveness	Participation, engagement, attitude and beliefs towards
Implementer	Adherence	The extent to which specified program components were delivered as prescribed
	Exposure	Number of sessions, duration of sessions, frequency with which program techniques were implemented
	Quality of delivery	“Buy-in”, attitude towards, perception of delivery, global estimates of session effectiveness
Unit of Randomization	Program Differentiation	Verification that only program schools received treatment and control schools were not implementing parallel/comparable curriculum/techniques

\*Adapted from: Dane, A. V., & Schneider, B. H. (1998). Program integrity in primary and early secondary prevention: Are implementation effects out of control? *Clinical Psychology Review*, 18, pg45.

Adherence refers to how closely the specific components of a program were delivered as originally intended. Modifications to program lessons/content, also referred to as reinvention (Pentz et al., 1990), can change their meaning and eventual impact, leading to either reduced or enhanced outcomes depending on what modifications were made. Measures of adherence, such as the degree to which adaptation of a planned curriculum occurs or the extent to which program material/concepts were omitted, can shed light on whether the program was implemented according to its original design. However, a certain degree of local adaptation (Basch, 1984) of program materials should be anticipated; thus providing flexible program curriculum, as opposed to rigidly adhered to materials, which is likely to enhance teacher acceptance and eventual delivery (Dane & Schneider, 1998). Nevertheless, quantifying these changes is necessary in determining whether such modifications compromise the program’s impact.

Exposure (student-level measure of how much of a program they received) and dosage (teacher-level measure of how much was delivered) consist of the amount of a program the intended audience is provided over the duration of an intervention. The most common measures of exposure are the number of classes taught, frequency of delivery over a specified segment of time (e.g., week, month, year), and/or the amount of time spent on program content. These provide an understanding of the amount of a program the audience requires in order to achieve intended results. The need for more or less of a program may be revealed. For example, dosage information may indicate that a program of a given length beneficially impacts outcomes with minimal



exposure, signifying to program developers that administering the entirety of a program may not be necessary to realize positive effects. Conversely, one of the more robust findings in the literature is the dose-response relationship -- with those receiving greater amounts of a program's content experiencing better outcomes (e.g., reduced substance abuse, hostile behaviors) (Aber, Brown, & Jones, 2003; Domitrovich & Greenberg, 2000; Flay, 1986; Pentz et al., 1990) .

However, just because a program's content is delivered does not mean it was delivered well. Quality of delivery pertains to the aspects of the delivery of the program related to the effectiveness of the implementer in teaching program materials (self-perception of abilities), their attitude towards program ideals/content, and their perceived/expected effects (positive or negative impact) of the program on the intended audience (Dane & Schneider, 1998). These can indirectly influence an implementer's willingness to deliver a given program as designed. Important in evaluating the dosage of a program and the quality of that delivery is whether the implementer has "buy-in" to the program ideals, and thus enthusiastically promotes, models, and reinforces fundamental program concepts (e.g., positive behaviors). If implementers disagree with a program's goals, then it is likely that qualitative aspects of program delivery will be diminished (e.g. low enthusiasm, lack of depth of program knowledge) (Han & Weiss, 2005). This, in turn, might lead to lower dosage and quality of program delivery and, thus, the program will have less effect on targeted outcomes. In the evaluation of the Adolescent Alcohol Prevention Trial, Rohrbach and colleagues (1993) found that teachers who delivered a greater number of program lessons were more enthusiastic and had higher self-perceptions of their abilities to deliver the materials in comparison to the teachers who failed to implement the program. Teacher effectiveness scores, which included observations of the quality of classroom delivery of program materials, also have been found to be positively associated with student health-related outcomes (Taggart, Bush, Zuckerman, & Theiss, 1990). Within treatment settings (i.e., schools receiving the experimental program) quality of delivery has been shown to moderate program impact (Hansen et al., 1991). Not surprisingly, these studies indicate that when program materials are delivered with

enthusiasm and teachers' believe in the program goals/objectives, positive outcomes are realized.

Responsiveness (e.g., customer satisfaction) is the reaction from the intended audience to the program that is actually delivered (Dane & Schneider, 1998). Depending on the theoretical constructs proposed by a given program, responsiveness may consist of several measures, such as audience participation, enthusiasm, beliefs and attitudes towards the program, audience perception of the materials/curriculum, and how involved they are in the classroom lessons. For instance, if a program proposes teacher-student positive interactions are an important component in realizing program outcomes, student responses as to whether these interactions are perceived as positive, or are observed to be so by a third party, would indicate whether the program is having the intended effect (Harachi et al., 1999). However, a program may be delivered as designed, in a high dosage, by competent implementers, yet if the audience does not take part in delivery (i.e., participation) or views the curriculum as inappropriate (e.g., low enthusiasm, negative attitude towards), expected program results are unlikely to be achieved.

Using a program resiliency framework, Gager and Elias (1997) found student program satisfaction to promote program "resiliency" as demonstrated by continued implementation despite a school's high risk status (low SES, degree of urbanization). Although the importance of the responsiveness of the audience to a program is crucial, relatively few studies have included such indicators (Gager & Elias, 1997; Hansen et al., 1991).

The responsiveness component of implementation relies on what the implementers do (i.e., adherence, dosage) and how well they do it (i.e., quality of delivery), creating a sequential chain from implementer actions to audience responsiveness to intended program outcomes (see Figure 1). Further, the responsiveness of the audience and the perceived positive outcomes observed by the implementers are likely to influence, via a feedback mechanism, how often a program is implemented in the future. Theoretically the proposed model is likely to work as such: Teachers deliver a program with enthusiasm and confidence towards topic comprehension or comfortableness (i.e., quality of delivery), do minimal

modifications to the materials/content (i.e., adherence), and do so continuously within the classroom (i.e., dosage). Students in turn enjoy the program's activities and message (i.e., responsiveness), reciprocating this back to the teacher in the form of engagement during the lessons and by exhibiting proposed program outcomes, such as intra- and inter-personal behaviors in the classroom (e.g., self-honesty, respect for others, improved academic performance). Teachers, in turn, observe that students are not only participating in the program activities, but also benefiting from the exposure, and decide to incorporate additional program materials and concepts into their curriculum.

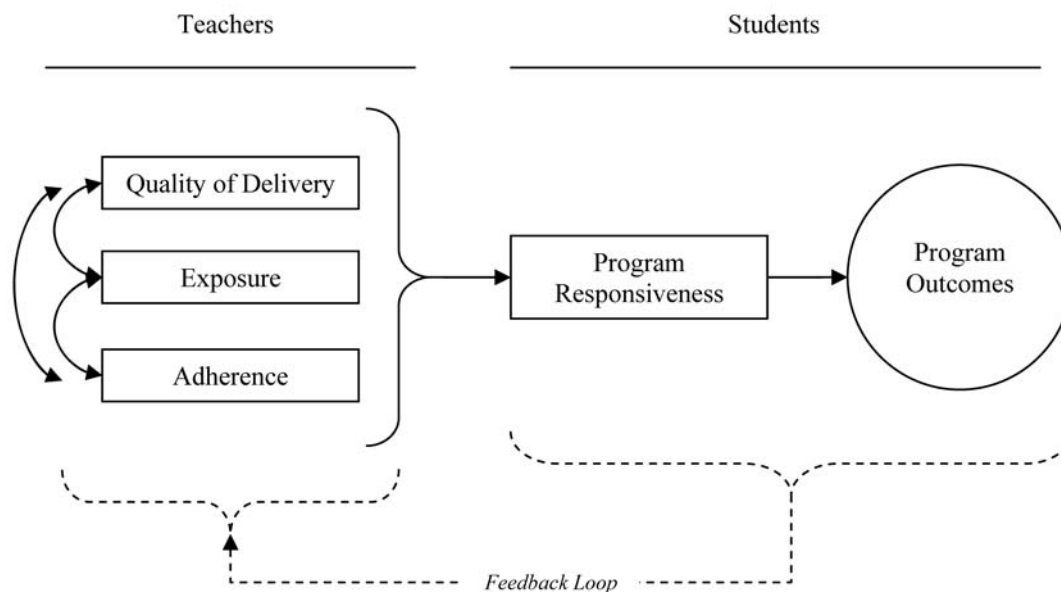


Figure 1. Teacher-student-outcomes program fidelity model of proposed effects.

*Note: The curved brackets are used to simplify the number of arrows in the diagram and represent the paths from each variable (e.g., exposure, adherence) to its respective casual influence (e.g., program responsiveness)*

Important in this feedback loop is that both student involvement and beneficial outcomes are observed. For instance, students may be highly enthusiastic about a program's content, but if no observable improvements or deleterious outcomes are observed (e.g., increased negative behavior towards others, increased substance abuse), the teacher is unlikely to continue delivering the program. Conversely, if

students dread the program sessions (if positive outcomes are realized) or fail to attend to the teacher during the delivery, cessation of the program will likely occur. Thus, the key elements in teacher-student-outcome model are responsiveness coupled with the attainment of beneficial outcomes.

The final component of implementation is program differentiation. Program differentiation deals with assuring that control conditions (e.g., classrooms, schools, districts) are not adopting or implementing programs/curriculum of similar content and techniques as specified in the program schools. Cook and Campbell (1979) refer to the idea of control groups adopting interventions similar to that provided to treatment groups as compensatory equalization of treatments or compensatory rivalry, and it is an important threat to internal validity. Part of program differentiation is verifying that the mechanisms of change, as defined by a program's theoretical perspective, are taking place only within the program schools. For example, to verify program differentiation in the Raising Healthy Children trial, Harachi and colleagues (1999) evaluated the extent to which the innovative teaching practices (theoretical mechanisms of change based on the social development model, Catalano & Hawkins, 1996) were occurring within both the control and program schools. Using observations of teacher-student interactions, the authors found that similar practices were occurring within both conditions, yet at a greater frequency in the program schools. They concluded that training on the specific teaching practices (e.g., engagement in learning tasks, cooperative learning opportunities) lead to improved student outcomes, but that such practices may be implemented without program specific training. Program differentiation measures also assist in understanding null program results. It is likely that nonsignificant treatment results could be explained by the existence of parallel teaching activities in control schools, indicating that the program was not a unique treatment.

#### *The importance of measuring implementation*

The primary purpose of measuring implementation is to lend internal and external validity to assessments and conclusions from prevention program trials (Durlak, 1998). Monitoring program implementation is crucial in understanding how a

program works and under what circumstances it is successful in real-world settings. Effectiveness trials, where programs are delivered by lay-professionals or paraprofessionals and delivered in situations where control over the degree to which program adoption and implementation varies, offer this “real-world” evaluation (Flay, 1986). While invaluable in determining program worth, outcomes from effectiveness trials are vulnerable to a greater rate of implementation variability or failure because of the lack of control over the delivery of the program. Further, a greater degree of variation in implementation between and within schools is likely to occur (Dane & Schneider, 1998). Such circumstances necessitate the need to closely monitor program fidelity (as defined above), with this information leading to greater assurance that the effects observed are attributable to the program itself.

Several rationale are proposed to substantiate the need for prevention scientists to monitor implementation of programs in naturalistic settings (Basch et al., 1985; Domitrovich & Greenberg, 2000). These are presented in Table 2. The two primary themes center on the notion that, in the absence of implementation information, it is impossible to determine what actually occurred during the intervention trial and, with this, the potential exists for misinterpretation of the programs’ efficacy/effectiveness. In addressing the former, both positive outcomes (i.e., beneficial outcomes presumably attributed to delivery of a given program) and negative outcomes (i.e., the failure to realize program benefits – attributed to failure of program design) cannot be credited to the program itself because it is unclear whether the program was delivered as intended, to what degree participants were exposed (e.g., frequency and duration), and how or if program materials and content were modified and adapted to fit with the intended audience (Durlak, 1998). Furthermore, when proposed program effects are realized, questions cannot be answered regarding the circumstances and contextual characteristics (both of the implementers and settings) that may, in part, account for the findings. Implementation information directly addresses whether the program itself, variations of a program, contextual variables, or any combination thereof explain why a program did or did not achieve the desired outcomes. Each piece of this information allows for firmer conclusions to be made when determining whether a program truly failed or was simply not delivered as intended.

Table 2. Rationale for measuring implementation\*

Source	Implementation Rationale
Basch et al. (1985)	Improve understanding about best practice techniques for promotion, long-term adoption, and future program dissemination
	Provide accountability to agencies that allocate resources
	Enhancement of validity of summative evaluations
	Information used to modify and improve program effectiveness
Domitrovich & Greenberg (2000)	Without implementation it is impossible to know what actually happened during an intervention
	Relating implementation quality to program outcomes is essential for establishing the internal validity of a program and strengthens any conclusions that are made about the programs role in producing change (reduce Type III Errors)
	Understand the internal dynamics and operations of an intervention
	Source of feedback for continuing improvement
	Advance knowledge of best practices for replicating, maintaining, and diffusing programs in naturalistic settings

Adapted from: Basch, C. E., Sliepcevich, E. M., Gold, R. S., Duncan, D. F., & Kolbe, L. J. (1985) and Domitrovich, C. E., & Greenberg, M. T. (2000).

Researchers have termed arriving at incorrect conclusions regarding the effectiveness of a program a Type III error (Scanlon et al., 1977) – incorrectly concluding a failure to achieve program goals/outcomes was due to program design, when in fact the program was not implemented as planned. When program components are not fully delivered or major portions modified or completely removed, adherence to the original intent of the program is compromised, thereby jeopardizing the fidelity of the program (Dane & Schneider, 1998). If modifications did occur, measures of implementation that detail what the specific program components should be, and what were eventually delivered, can provide valuable information as to whether the “spirit” of the program was maintained or delivery was completely “off-the-mark” in regards to the program’s original intentions. Thus, implementation measures can be used to avoid misinterpretation of program results, positive or negative.

### *Factors effecting implementation*

The reasons why some implementers (e.g., teachers) choose to implement a program without changes (i.e., adherence), do so on a frequent basis (i.e., dosage), and deliver program ideals with confidence and competence (i.e., quality of delivery),

while others choose not to do so, are complex. Chen's (1998) theory-driven evaluation model of program implementation and Han and Weiss (2005) describe the proximal factors that affect the treatment/intervention as delivered. These fall under two broad categories of characteristics of the implementer (e.g., teachers and other school staff) and contextual characteristics of the setting (e.g., school).

For school-based primary prevention programs, teachers' primary capacity is to serve as central change agents (Han & Weiss, 2005). Their position in the classroom places them in constant contact with students (i.e., the target audience) and with this comes the opportunity to promote/reinforce program ideals while providing students instances to practice new behaviors (e.g., respectful communication). Yet, not all teachers, even within the same school, implement a program equally. Contributing to this are "teacher-level processes" or teacher-specific factors found to influence what teachers actually do (Chen, 1998; Han & Weiss, 2005).

The beliefs and attitudes teachers hold towards a given program's message or ideals can influence whether they decide to implement a program. When a program is perceived as having favorable outcomes and is consistent with teachers' current sets of beliefs, greater implementation is likely to result (Kealey, Peterson, Gaul, & Dinh, 2000). Rohrbach and colleagues (1993) found that teachers who displayed greater enthusiasm for the program's materials – an indirect assessment of beliefs/attitudes – taught a greater number of the program's lessons. Changes in teachers' attitudes towards innovative teaching practices designed to promote cooperative learning, prosocial values, and intrinsic motivation have been related to declines in student positive behaviors and less misconduct at school (Battistich, Schaps, & Wilson, 2004; Solomon, Battistich, Watson, Schaps, & Lewis, 2000). In addition, programs typically target undesirable behaviors; however if a program focuses attention on to the implementers own negative behavioral attributes, perceptions towards the program may be unfavorable, ultimately affecting the level of implementation (Taggart et al., 1990). Thus, if teachers do not "buy-in" to the message a program is promoting or believe it does not coincide with their teaching philosophy, successful implementation is unlikely.

If teachers do not perceive support/commitment from administration to use the program, no amount of positive attitudes/beliefs will ameliorate this deficiency (Connell, Turner, & Mason, 1985; Smith, McCormack, Steckler, & McLeroy, 1993). The finding that administrative encouragement increases positive program maintenance and outcomes is robust (Kam, Greenberg, & Walls, 2003; Rohrbach et al., 1993; Sheldon, 2005; Smith et al., 1993). A supportive administration may also contribute to program outcomes through a synergistic effect. Kam and colleagues (2003) found that when both support from administration and teacher implementation of program material (e.g., number of lessons taught) were high, greater positive student outcomes were observed. They also found that when implementation by teachers was high and administrative support absent, positive student outcomes were not guaranteed. Given the consistency of these findings, it appears that administrators who champion a program are more likely to have programs that achieve their intended positive results.

The contextual characteristics of the setting include the size of the school, demographics (e.g., SES, percent free and reduced lunch), and location (e.g., rural, urban). The size of a school or district can exhibit variable effects on implementation fidelity. Larger schools and districts are more likely to have greater access to resources (Smith et al., 1993), yet are also subject to greater organizational complexity which may reduce effective decision making and communication of the program's core elements (Smith, Redican, & Olsen, 1992). On the other hand, smaller schools and districts have closer-knit communication among administrative staff and teachers, but may lack adequate access to the required resources to successfully implement a program (Smith et al., 1993; Smith et al., 1992). Schools serving economically disadvantaged families may experience greater difficulties in reaching families for participation in family-school program activities (Sheldon, 2005).

While the presence of contextual characteristics can negatively impact implementation fidelity, the detrimental effect of these can be alleviated. In their analysis of factors associated with successful implementation and adoption of programs in high-risk schools (based on SES and demographic status, percent minority), Gager and Elias (1997) found two dimensions related to success – program



institutionalization and implementer/recipient focus. Program institutionalization consisted of having a program champion, either a designated program coordinator or staff member, to promote the message of the program, high staff commitment, and high visibility of the program's activities. Implementer/recipient focus dealt with ensuring teachers were trained, appropriate materials available, and students' had positive perceptions of the program.

Taking the above teacher and administrative characteristics into account, the conceptual model of implementation fidelity presented in Figure 1 is expanded and presented in Figure 2. The additions to the model illustrate that the teacher dimensions of program fidelity (i.e., quality of delivery, exposure, and adherence) are moderated by their attitudes and beliefs towards the program's principles. Serving as an exogenous factor is the extent to which teachers perceive receiving support from their administration to implement the program. Affecting teachers' implementation and student response to a prevention program are the school-level contextual characteristics which can impact the entirety of the implementation system, yet can be enhanced or counteracted through levels of staff commitment and attitudes of students.

### Summary

Using Figure 2 as a guiding framework, three studies were conducted to examined the relationship of various components of implementation to the amount of the program utilized, its effects on program outcomes (e.g., prosocial behaviors), and the extent to which the control schools were self-initiating comparable programming. In regards to the former, the influences of teacher and school characteristics were modeled in relation to program material utilization in the classroom and school wide. This analysis addresses the left portion of Figure 2 – school climate influencing teacher characteristics (e.g., attitudes), which, in turn, influence adherence, exposure, and quality of delivery. Specific to the latter, a longitudinal analysis was conducted that examined the influence of the amount of the program received on student perceptions of the program (i.e., responsiveness). These were then specified to influence program-related outcomes. Finally, the amount of programming that mimics

experimental conditions (i.e., program differentiation) was examined at the final year of the trial.

In conclusion, the information gained from measuring the extent to which a program was implemented can be used to improve program content, delivery, and eventual outcome performance. Implementation is also an essential and necessary component to document in order to establish efficacy, effectiveness, and dissemination (Flay et al., 2005). Each primary prevention program deals with its own set of constructs, implementers, audiences, and settings in which the program takes place. Thus, each program should systematically examine the extent to which a program is carried out as intended, and when deviations occur, determine potential reasons for these and how these affect program outcomes.

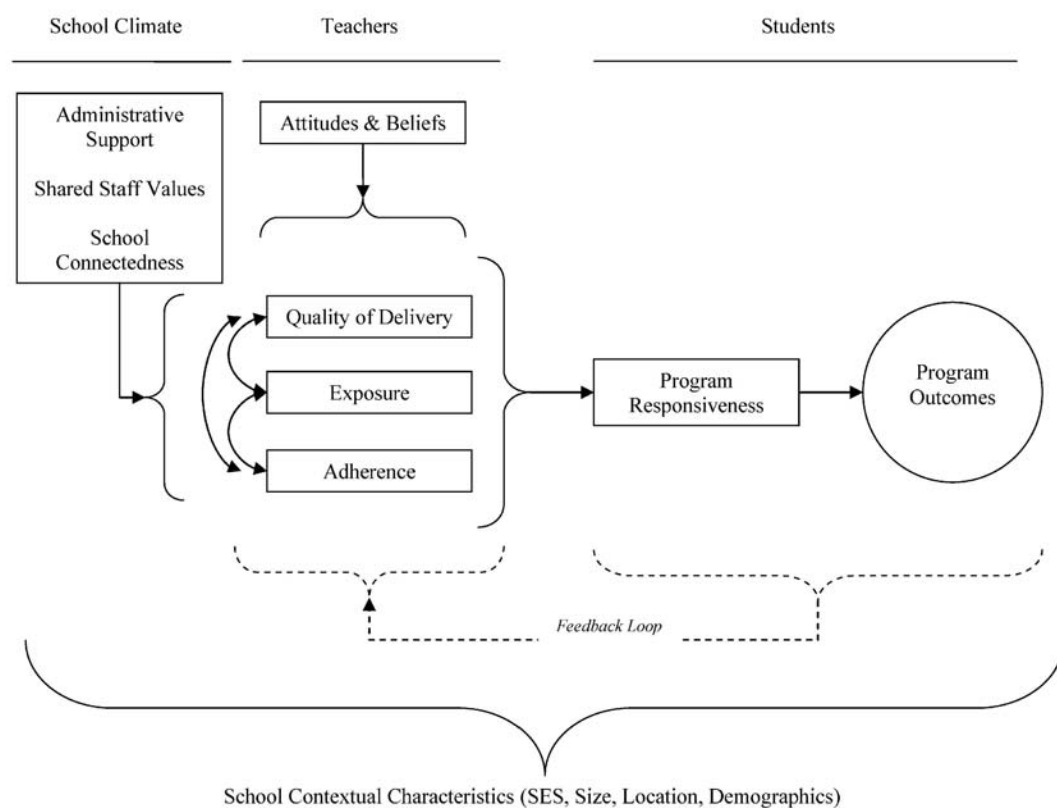


Figure 2. Conceptual model of teacher, administrative, and contextual characteristics influencing implementation fidelity

## Chapter 2. Manuscript 1

School climate and teachers' beliefs and attitudes associated with implementation of  
the *Positive Action* program: A Diffusion of Innovations Model

Michael W. Beets, Brian R. Flay, Samuel Vuchinich, Alan C. Acock, and Carol Allred

Journal: Prevention Science  
Address: Gilbert J. Botvin, Ph.D., Editor  
c/o Nicole Miller  
Managing Editor  
Prevention Science  
Weill Medical College of Cornell University  
411 E. 69th Street, KB-200-K1  
New York, NY 10021  
Status: In review

### Abstract

Teacher and school level factors influence the fidelity of implementation of school-based prevention and social character and development (SACD) programs. Using a diffusion of innovations framework, the relationships among (a) teacher beliefs and attitudes towards a prevention/SACD program, (b) the influence of a school's administrative support and perceptions of school connectedness, characteristics of a school's climate, were specified in two cross-sectional mediation models of program implementation. Implementation was defined as the amount of the programs' curriculum delivered (e.g., lessons taught), use of program-specific materials in the classroom (e.g., ICU boxes and notes) and use of school-wide activities (e.g., participation in assemblies). Teachers from 10 elementary schools completed year-end process evaluation reports for year 2 (N = 171) and 3 (N = 191) of a multi-year trial. Classroom and school-wide material usage were each favorably influenced by the amount of the curriculum delivered, which were influenced by teachers' attitudes toward the program which, in turn, were shaped by teachers' beliefs about SACD. These, in turn, were shaped by teachers' perceptions of school climate. Perceptions of school climate exerted an indirect effect on classroom material usage and both indirect and direct effects on the use of school-wide activities. Program developers need to consider the importance of a supportive environment for program implementation and attempt to incorporate models of successful school leadership and collaboration among teachers that foster a climate promoting cohesiveness, shared visions, and support.

**Keywords:** Fidelity; Primary Prevention; Elementary; Children

*School climate and teachers' beliefs and attitudes associated with implementation of the Positive Action program: A Diffusion of Innovations Model*

Over the past few decades, considerable emphasis has been placed on the design, implementation, and evaluation of school-based prevention/SACD programs aimed at decreasing behaviors and attributes that are health-compromising and detrimental to academic and social success (e.g., substance abuse, delinquency, physical aggression) (Flay & Collins, 2005). Some recent work also has focused on improving positive student behaviors (e.g., academic performance, prosocial values) and school level outcomes (e.g., climate). Recently, criteria have been outlined for establishing the “value” of a program<sup>3</sup> under controlled (efficacy) and naturalistic (effectiveness) settings (Flay, 1986). These include specifying that, in order for a program to be considered ready for widespread dissemination, detailed analyses must be conducted on the *how* and *why* a given program works and under what conditions (Flay et al., 2005).

One of the factors instrumental in determining whether a program “achieves” success, as demonstrated by beneficial outcomes (e.g., improved academic performance, reduced substance abuse rates), is the delivery of a program by teachers with fidelity (Battistich et al., 2004; Dusenbury, Brannigan, Falco, & Hansen, 2003; Games, Millsap, & Goodson, 2002; Harachi et al., 1999; Lillehoj, Griffin, & Spoth, 2004; Rohrbach et al., 1993; Smith et al., 1993; Solomon et al., 2000; Taggart et al., 1990). As central change agents, teachers typically serve as the main deliverers (i.e., implementers) of school-based prevention/SACD programs (Han & Weiss, 2005). Their constant presence in the classroom creates frequent and extended contact with students, the primary target group. This provides ample opportunity to teach and reinforce program concepts so that the utilization of a program’s curriculum, along with program-specific materials, can become a natural extension of everyday activities. Teachers are, therefore, viewed as an integral component in the implementation process. Yet, the extent to which they use a program in their classroom is primarily self-guided (Rohrbach, Ringwalt, Ennett, & Vincus, 2005). Teachers’ on-going decisions to use as much or as little of a program as they choose

---

<sup>3</sup> The term “program” is used herein to refer to school-based primary prevention programs

ultimately determines whether it is implemented with fidelity in terms of dosage, adherence and quality of delivery (Basch, 1984; Dane & Schneider, 1998; Domitrovich & Greenberg, 2000). Further, not all teachers, even within the same school, implement a program with the same fidelity (e.g., Rohrbach et al., 1993; Smith et al., 1993; Taggart et al., 1990).

Prior research has identified two broad categories of characteristics that affect a teachers' ability and willingness to implement a program (Chen, 1998; Han & Weiss, 2005): teacher-level characteristics (e.g., attitudes) and contextual characteristics of the setting (e.g., school). Based on theory-driven evaluations (Chen, 1998) and the diffusion of innovations model<sup>4</sup> (Rogers, 1995), researchers also have identified certain characteristics and qualities of implementers that influence the degree to which a program is carried out with fidelity. Rogers (2002) points out that key factors associated with the adoption of an innovation (i.e., program) are whether innovators (i.e., teachers) perceive the program to be more advantageous than alternatives and if it (i.e., the program) is consistent with existing values and beliefs. These teacher-level processes, therefore, are linked to decisions about adoption and the effort put towards program implementation (Chen, 1998). For instance, teachers who a) report greater enthusiasm for a program's content (Hahn, Noland, Rayens, & Christie, 2002; Rohrbach et al., 1993), b) self-identify with a program's goals and concepts (e.g., see themselves as good role models) (Taggart et al., 1990), c) view a program to be compatible with their current set of beliefs (Kealey et al., 2000), and d) hold favorable attitudes towards a program (Battistich et al., 2004; Solomon et al., 2000), are likely to deliver a program with greater adherence and dosage (Dane & Schneider, 1998; Domitrovich & Greenberg, 2000). These, in turn, should lead to improved student outcomes. However, while a teacher may implement a program with high fidelity, in the absence of a supportive school setting positive program outcomes are not always guaranteed (Kam et al., 2003; Sheldon, 2005).

---

<sup>4</sup> It is important to note that in the current study, the use of the diffusion model is specific to the internal dissemination processes (Pentz, 2004) taking place within a school and not the overall dissemination process which would incorporate a larger system, such as those that take place among schools and governing agencies within and across school districts.

The school setting or climate refers to contextual characteristics specific to schools that distinguish one from another (Gittelsohn, Merkle, Story, Stone, Steckler, Noel, Davis, Martin, & Ethelbah, 2003). Of these characteristics, several of the more robust associations with high fidelity of implementation are a supportive administration (Connell et al., 1985; Elliott & Mihalic, 2004; Kam et al., 2003; Payne, Gottfredson, & Gottfredson, 2006; Rohrbach et al., 1993; Smith et al., 1993) and support for a program received from one's peer group (i.e., teachers) and other school personnel (McCormick, Steckler, & McLeroy, 1995; Sheldon, 2005). These components of a school's climate are, in turn, posited to influence those who come into continuous contact with it (Gittelsohn et al., 2003; Rogers, 1995, 2002). Based on the diffusion of innovations model, the influence of administrators and co-workers on a teacher's attitudes and beliefs acts through a social process whereby individuals (i.e., teachers) evaluate a program, either favorably or unfavorably, through the subjective perceptions of the social system in which they are embedded, in this case the school (Gittelsohn et al., 2003; Rogers, 2002). Thus, for a program to realize its goals, the internal dissemination (Pentz, 2004) process would be characterized by an administration that champions a program, devotes time towards the program in the school's schedule, role models program ideals, and encourages program usage. This would occur in conjunction with staff that hold a shared vision regarding a program and whether the program's foundational concepts (e.g., caring, nurturing positive attributes) are instrumental in creating a cohesive climate among teachers. These, in turn, would begin to alter the teachers' attitudes towards a program and their beliefs regarding their responsibility to teach program-related concepts, with this leading to greater levels of implementation. Although this idea is intuitively appealing, little work has been conducted that specifically measures and tests these processes (Dusenbury et al., 2003).

A logical progression in the diffusion model is the adoption of a program and, with this, teachers developing an understanding of, and favorable attitude towards, the program's concepts. These processes lead to a critical outcome in program implementation – the teaching of the program's concepts through a program's curriculum. Further, some prevention/SACD programs, including *Positive Action*,



incorporate materials designed specifically to reinforce the core concepts central to the program's goals (e.g., prosocial values, equity, "everyone matters"). Thus, many programs rely on the use of specific materials to fully realize successful outcomes (e.g., reduced substance abuse). Program material utilization, therefore, implies that implementers make use of program-specific materials corresponding to lessons that focus on core concepts embedded within a program. A likely scenario is one where a teacher may deliver a program's curriculum without utilizing the corresponding materials designed to reinforce the concepts embedded within the lessons. Conversely, a teacher may utilize program materials, such as providing token reinforcements (in this case, to help students remember how good they feel about themselves when they perform positive behaviors), yet fail to teach the behavioral concepts in corresponding program curricula. Both of these scenarios would be considered low fidelity implementation. The former is a failure to strengthen the link between the curricula and requisite behaviors outside the context of lesson delivery, while the latter is a failure to capitalize on the "teachable moments" that correspond to *why* the youth was rewarded for their behavior. If teachers are utilizing program materials in the absence of teaching the core components of a program, or vice versa, a "disconnect" is created between what behaviors are ultimately viewed as appropriate and valued. Thus, when teachers teach concepts, this should occur with the use of materials specific to the program, thereby building a connection between the program and the behaviors it wishes to promote. A reasonable hypothesis in this framework would be that the attitudes teachers hold towards a program would be predictive of whether they teach the programs' curriculum, with this, in turn, related to the extent to which they utilized a program's materials (see Figure 1). The diffusion model presented here provides a theory-driven approach to measuring and testing these specific implementation processes.

--- Insert Figure 1 about here ---

Conceptually, implementation consists of five key components: dosage, quality of delivery, responsiveness, adherence, and program differentiation (Dane & Schneider, 1998; Domitrovich & Greenberg, 2000). The current investigation focused on three of these – dosage (i.e., teaching the program curriculum), quality of delivery

(i.e., attitude towards program), and adherence (i.e., usage of program-related materials). Although each is a component of implementation, considerable interconnectedness exists among them, with no one component existing within a vacuum. That is, the presence of high or low levels of one component is often, though not always, associated with the presence of high levels of the other components. As stated previously, teacher's with favorable attitudes towards a program are likely to deliver more of a given program than teachers with less favorable attitudes (e.g., Battistich et al., 2004; Solomon et al., 2000), thereby linking quality of delivery with dosage. A likely outcome of increased program delivery should be usage of corresponding program-specific materials. However, this is unlikely to occur if a teacher's attitudes toward a program are not complimentary. Further, while dosage and adherence are typically emphasized as part of a comprehensive and cohesive implementation of a program, teachers clearly have the option to utilize as much or as little of the program as deemed necessary. Given this and the findings from previous studies (e.g. Battistich et al., 2004; Games et al., 2002; Rohrbach et al., 1993; Solomon et al., 2000; Taggart et al., 1990), it was expected that varying rates of program implementation would be observed in the current study.

The present study examines these relationships in a comprehensive model that includes the influence of the school's climate and teacher beliefs regarding their responsibility to teach SACD concepts in a multiyear evaluation of a school-based prevention/SACD program *Positive Action* ([www.positiveaction.net](http://www.positiveaction.net)) conducted in 10 Hawai'i public elementary (K-6) schools. The current study addressed the following research question, "*What teacher- and school-related factors influence some teachers to implement Positive Action with greater fidelity than others?*" Based on the diffusion of innovations model, along with the findings from prior research and theory presented above, the following hypotheses were proposed:

- d) Utilization of the *Positive Action* program materials will be enhanced by the following teacher-level process: positive teacher attitudes toward *Positive Action* will positively influence teaching of the program's curriculum which, in turn, will lead to increased material usage, both within the classroom and across the school as a whole.

- e) Teachers' attitudes toward *Positive Action* will be influenced by their beliefs regarding their responsibility to teach prevention/SACD concepts.
- f) School climate will directly affect teacher-level processes (i.e., beliefs and attitudes) and the delivery of the curriculum, along with utilization of *Positive Action* materials school-wide, thereby exerting both direct and indirect effects on implementation

## Methods

### Program Overview

The *Positive Action* program is a comprehensive school-based prevention and youth development program designed to target the enhancement of positive behaviors and behavioral attributions directed at the self (e.g., self-responsibility, positive self-concept) and social relationships (e.g., conflict resolution, respect, kindness), while simultaneously reducing students' negative behavioral problems (e.g., substance abuse, disciplinary referrals, suspensions), with these leading to improved school-related performance (e.g., improved attendance, academic achievement). Prior quasi-experimental evaluations of *Positive Action* have demonstrated its beneficial impacts on student academic achievement and improved student behaviors (e.g., reduced disciplinary referrals) (Flay & Allred, 2003; Flay et al., 2001).

The program's theoretical foundation is broad and comprehensive, and may be summarized by the Theory of Triadic Influence (Flay & Petraitis, 1994). At its core, *Positive Action* is based on the intuitive philosophy that "you feel good about yourself when you do positive actions and there are always positive ways to do everything". The K-6 program focuses on the entire self, with 6 units consisting of approximately 140 15-minute, age appropriate lessons per grade level taught over the whole academic year. The lessons cover the topics of self-concept, mind and body positive actions (e.g., nutrition, physical activity, decision-making skills, motivation to learn), and social and emotional actions for managing yourself responsibly (e.g., emotional regulation, time management), getting along with others (e.g., empathy, respect), being honest with yourself and others, and self-improvement (e.g., goal setting, courage to try new things). Unique to the *Positive Action* program are the scripted, sequenced, and scoped lessons which include an extensive set of program materials:

ICU (“I see you”) doing something positive recognitions, posters, classroom and school-wide recognitions and announcements, as well as stickers and tokens to reinforce positive student/classroom behaviors.

### Study Background

The Hawai’i multiyear randomized trial of the *Positive Action* program took place in 20 public elementary schools on three islands in the Hawai’i school district. The sample of teachers examined in the current study comprises those who were employed in the 10 schools receiving the treatment (*Positive Action* program) during program implementation years 2 and 3. These years correspond to two cohorts of students receiving the program during grades 2 and 3 and grades 3 and 4. Schools were selected to receive the intervention based on random assignment from matched pairs based on a multivariate indicator of risk developed from school level data on the following characteristics: the proportion of students receiving free and reduced lunch, school size, pupil to teacher ratio, percent stability of student enrollment, ethnic distributions, and indicators of problem behaviors (e.g., suspensions) and achievement (e.g., standardized testing) (c.f., Dent, Sussman, & Flay, 1993).

Training. Prior to the beginning of each academic year, teachers, administrators, and support staff (e.g., counselors) attended *Positive Action* training sessions conducted by the developer of the program (Carol Allred). The training sessions lasted approximately 3-4 hours for the initial year, and 1-2 hours for each successive year (years 2 and 3 presented in this study). Booster sessions, conducted by the project coordinator, were provided at least once during the academic year for each school. These lasted approximately 30-50 minutes. Additionally, mini-conferences, held in February of each year, were conducted to bring together 5-6 leaders and staff (e.g., principals, counselors, teachers) from each of the 10 participating schools in order to share ideas and experiences as well as to get answers to any concerns regarding implementing the program.

Participants. Teachers’ ( $N_{Y2} = 171$  and  $N_{Y3} = 191$ ) from the 10 elementary schools completed year-end process evaluations. The majority of teachers at year 2 self-identified as being Japanese (41%), of other Asian decent (21%), and White non-Hispanic (18%). For year 3, teachers were primarily Japanese (36%), White non-

Hispanic (25%), and full or part Hawai'ian (12%). Approximately 17% and 10% of the teachers were male, with response rates of 60.2% and 56.7% for years 2 and 3, respectively. Response rates were calculated from School Status and Improvement Report estimates of full time regular teaching staff at each school for the corresponding study years (State of Hawaii Department of Education Systems of Accountability, 2006). The average number of teachers per school responding to the questionnaire for year 2 was 17.1 (range 1 to 27) and 19.1 (range 4 to 34) for year 3.

Questionnaire

Year-end process evaluation surveys were completed at the end of each academic year for implementation years 2 and 3 of the multiyear evaluation. In order to maintain confidentiality and to increase the rate of honest reports of low program implementation (e.g., amount of curriculum delivered, material usage or nonuse), teachers were only asked to identify their school and grade level taught. This, together with the fact that teachers of the student cohort followed in the study changed from year to year, precluded a longitudinal analysis of change in implementation across the two years. However, the teacher responses from the process evaluation provide two cross-sectional sets of measures of perceptions and implementation. This creates a cross-sectional design that allows two empirical tests of the diffusion model, one for each year. Approximately three-fourths of the same teachers are included in both years of the study, and they would have more experience with the program than teachers new to the program in year 3. This could create some differences in the magnitude of estimated effects across years. However, the same linkages predicted by the diffusion model would be predicted for both years, and this kind of dual cross-sectional design can be considered as a form of replication (Sidman, 1960).

The process evaluation survey was self-administered. The survey asked a series of questions (see complete descriptions below) pertaining to the teachers' perceptions of support from their administrators and connectedness with their school, their belief in their responsibility to teach social and character development concepts, their attitude towards *Positive Action*, the amount of the *Positive Action* curriculum delivered, and their usage of program specific materials for two contexts – the classroom and school-wide levels. All procedures were approved by the Human

Subjects Committee at the University of Illinois at Chicago and the Hawaii Department of Education.

School Climate. School climate was composed of two series of questions assessing perceived *administrative support* and *school connectedness*. The scale for administrative support consisted of 7 items referring to behaviors exhibited by administrators. Example items included, administrators: “accomplish their job with enthusiasm”, “reinforce student accomplishments”, and “are proud of this school.” Response options ranged from 1 “never” to 5 “always”. The second component of school climate consisted of 6 items referring to teachers’ perceptions of their connectedness with the school. These items included “I feel at home in this school”, “people at this school generally get along with each other”, and “this school is a good place for me to work.” Responses ranged from 1 “strongly disagree” to 4 “strongly agree”.

Beliefs about one’s responsibility to teach social and character development concepts. The beliefs construct was composed of 14 items referring to the extent to which teachers believed it was their responsibility to teach students social and character development concepts. Examples are “having positive self-concept”, “peer pressure resistance”, and “decision making skills”. Responses ranged from 1 “never” to 5 “always”.

#### Implementation measures

Quality of Delivery (Attitudes toward *Positive Action*). Teachers’ attitudes toward the *Positive Action* program were measured by 4 items. Example items were “the time required by *Positive Action* is well worth it in improved student behavior and classroom management” and “I personally benefit from teaching *Positive Action*”. Responses ranged from 1 “strongly disagree” to 4 “strongly agree”.

Dosage (Amount of the *Positive Action* curriculum delivered). The amount of the *Positive Action* curriculum delivered was measured by 6 items. A single item referred to each of the 6 units in the *Positive Action* curriculum (e.g., Thoughts-Actions-Feelings circle from unit 1) and asked about how often the teachers taught the concept throughout the school day. Responses ranged from 1 “never” to 5 “always”. Additionally, a single item assessed the amount of *Positive Action* workbooks and

activity sheets used during a typical day. Responses ranged from 1 “never” to 5 “always”.

Adherence (Program-specific material usage). Adherence to the *Positive Action* program was measured for two specific contexts. The first was a series of 3 items referring to the usage of classroom-specific materials – labeled Classroom Material Usage. Classroom materials consisted of “Word of the Week cards”, “ICU Box”, and “Positive Notes”. Items asked about how often the teacher utilized the classroom materials during a typical school day, with responses ranging from 1 “never” to 5 “always”. The second implementation construct, School-Wide Material Usage, was measured by 3 items referring to the extent of use of school level *Positive Action* program materials utilized throughout the school. Items were “Word of the Week announcements”, “*Positive Action* posters”, and “attending *Positive Action* assemblies”. Responses ranged from 1 “never” to 5 “always”.

#### Analysis

Using structural equation modeling, a conceptual model based on the hypothesized relationships among the latent constructs (see above) was specified (see Figure 1). A full information maximum likelihood estimator robust to non-normality was used to derive model estimates. All estimates were adjusted to control for clustering of teachers within schools.

Beginning with year 2, we used the entire set of items for each latent construct and applied a model trimming approach to reduce the number of observed indicators per latent factor (Marsh, Hau, Balla, & Grayson, 1998). Item correlations (analyzed in Stata v.9.2) for each construct (e.g., beliefs 14 items) were examined to create parceled items (i.e., average of items). Items with correlations at or above .60 were averaged to create a single average item. Correlations among the remaining single items and averaged (i.e., parceled) items within a scale were reexamined and the above process repeated. This iterative process was performed until all latent constructs had observed items reduced from their original number to no more than 4 per construct (Marsh et al., 1998). The decision to parcel items was based on the need to provide a more optimal ratio of variables to sample size and to reduce the number of indicators per latent variable and, subsequently, the number of paths estimated in the

final model (Bandalos & Finney, 2001). Full-scale and parceled-scale item alpha reliabilities are presented in Table 1.

--- Insert Table 1 about here ---

Both direct and indirect effects were estimated for both models. Model-data fit was assessed by the following indices: Chi-Square statistic, root mean square error of approximation (RMSEA), Comparative Fit Index (CFI), and the Tucker Lewis Index (TLI) (Hu & Bentler, 1999; Kline, 2005). Values for the RMSEA ranging from 0.05 to zero, and CFI and TLI above 0.90 and 0.95, respectively for both years, represent acceptable fit of the model (Hu & Bentler, 1999; Kline, 2005). Significance for paths was set at  $p \leq .05$  ( $z \geq 1.96$ ). All structural equation modeling was performed with *Mplus* v.4.2 (Muthén and Muthén - [www.statmodel.com](http://www.statmodel.com)).

### Results

Descriptive statistics (means and  $\pm$ standard deviations) for the scale scores corresponding to the latent constructs are presented in Table 2. The final model with the parceled items, correlated residuals, and structural paths, for years 2 and 3, are presented in Figure 1. Overall, model data fit was good for both years; fit indices for year 2 were  $\chi^2_{df\ 179} (N = 171) = 305.92, p < .001$ , CFI = .952, TLI = .944, RMSEA = .064 and year 3  $\chi^2_{df\ 179} (N = 191) = 299.03, p < .001$ , CFI = .943, TLI = .933, RMSEA = .059. Correlated residuals were specified for the parceled items for perceived administrative support (Ad1 and Ad2) and beliefs on the responsibility to teach SACD concepts (Be1 and Be2) (see Figure 1) since significant correlations between the variables remained after parceling. The following subscript notation is used to refer to model estimates across years: Year 2 estimates ( $_{Y2}$ ) and Year 3 estimates ( $_{Y3}$ ).

--- Insert Table 2 about here ---

In support of our first and second hypotheses, teacher beliefs regarding their responsibility to teach SACD concepts were significantly ( $z > 1.96$ ) related to their attitudes towards *Positive Action* (standardized effects  $\beta_1 = .42_{Y2}$  and  $.40_{Y3}$ ); attitudes towards the *Positive Action* program were positively related to the amount of the *Positive Action* curriculum delivered ( $.67_{Y2}$  and  $.64_{Y3}$ ), and the amount of the curriculum delivered was positively related to material utilization in both the classroom ( $.47_{Y2}$  and  $.40_{Y3}$ ) and school-wide ( $.47_{Y2}$  and  $.31_{Y3}$ ) (see Figure 1).



The hypothesis regarding the impact of perceptions of school climate was also supported. During year 2, perceptions of school climate were directly related to the beliefs teachers held about prevention/SACD (standardized effect  $\beta_1 = .66_{Y2}$ ) and the attitudes teachers had towards the *Positive Action* program (.37<sub>Y2</sub>). At year 3, school climate was related to teachers' beliefs about their responsibility to teach prevention/SACD (.34<sub>Y3</sub>), but not significantly related to teacher attitudes towards *Positive Action* (.15ns<sub>Y3</sub>). School climate was significantly related to school-wide material usage during year 2 (.34) and year 3 (.30).

The summary of the direct and indirect effects of school climate on school-wide material usage are presented in Table 3. Perceived school climate demonstrated an indirect effect, mediated through attitudes towards *Positive Action* and the amount of the *Positive Action* curriculum delivered to school-wide material usage at year 2, whereas the only indirect effect observed at year 3 was with classroom material usage. The total effect, both direct and indirect, of school climate on material usage for classroom and school-wide at year 2 was .20 and .54, respectively, and .07 and .35 for classroom and school-wide, respectively, for year 3. The proportion of variance explained in the latent constructs for years 2 and 3, respectively, was .43 and .11 for beliefs about the responsibility to teach SACD concepts, .53 and .22 for attitude towards *Positive Action*, .45 and .41 for the amount of the *Positive Action* curriculum delivered, and .22 and .16 for classroom and .47 and .21 for school-wide material usage.

--- Insert Table 3 about here ---

To illustrate the relationships observed in the estimated models (see Figure 1), respondents were divided into two groups based on the lower and upper 25% of the distribution of the school climate and attitude towards *Positive Action* summary scale scores for year 2 and 3, separately (see Table 4). Descriptive statistics and effect sizes (ES), were computed (using a pooled standard deviation) to illustrate the influence of school climate and attitudes on the summary scores corresponding to the other latent constructs (e.g., beliefs, amount of curriculum delivered, material usage) estimated in the model. Overall, teachers who indicated the highest ratings of a favorable school climate (i.e., upper 25% distribution) also reported greater material usage, delivered

more of the curriculum, and had more favorable beliefs about their responsibility to teach SACD concepts and attitudes towards *Positive Action* than the teachers reporting the lowest levels of school climate (i.e., lower 25% distribution). The comparison of the teachers reporting the highest levels of favorable attitudes towards *Positive Action* indicated they also delivered more of the curriculum, and utilized more program-specific materials at both the classroom and school-wide.

--- Insert Table 4 about here ---

### Discussion

In order to be considered ready for widespread dissemination, a program must establish the key components that determine its success (Flay et al., 2005). Furthermore, when examining the decisions guiding the adoption, delivery, and sustainability of school-based primary prevention/SACD programs, it is essential to examine the factors that influence teachers to implement a given program with fidelity (Glasgow, Vogt, & Boles, 1999). An understanding of modifiable factors related to program delivery can inform program developers about the requisite strategies that shape implementers' attitudes towards a program and beliefs regarding their responsibility to teach program-related concepts, along with those for altering the climate of a school. By integrating these findings into a programs' training and practice, innovative programs can refine the program dissemination process, which is likely to lead to greater program outcomes (Botvin, 2004).

Using a diffusion of innovations model (Rogers, 1995, 2002) and guided by a theory-driven evaluation framework (Chen, 1998), this study examined the teacher- and school-related factors associated with the implementation of the *Positive Action* program. Important in the findings were the substantial direct and indirect effects of teacher perceptions of school climate on teacher beliefs about the importance of teaching SACD concepts and the attitudes they held regarding the utility of the program itself. These findings illustrate that in order to ensure teachers implement the program with fidelity, school leadership must develop a culture that encourages a shared and collective vision among staff and administration, is supportive of new innovations, and is aligned with the core values and concepts a given program is promoting. It seems reasonable to suggest that prior to adoption or full implementation

of a program, efforts should be directed towards ensuring that administration and staff members within a school are united in their efforts and perceptions towards a program considered for adoption. Such preliminary work may include obtaining widespread agreement for adoption before entering a school into a trial (Slavin & Fashola, 1998), pre-training sessions for school administrators, and staff cohesion activities during pre-implementation program training sessions (e.g., McCormick et al., 1995; Rohrbach, D'Onofrio, Backer, & Montgomery, 1996).

The indirect effects of perceptions of school climate on teacher attitudes and beliefs are consistent with the diffusion of innovations model. This model posits that a social system, which is in favor of a given innovation, will begin to modify and shape the salient attributes (e.g., attitudes) of those within it, with this leading to greater implementation of the innovation (Rogers, 2002). This process was demonstrated by the relation of these attributes to higher levels of the program curriculum delivered and the use of corresponding program-specific materials (see Figure 1). Although cross-sectional, thereby precluding the ability to infer causality, the replication of the overall models across years provides preliminary evidence that these processes were occurring over time (Sidman, 1960).

In regards to our first hypothesis, teacher attitudes toward the program impacted the extent to which the program's curriculum was delivered, and this led to greater program utilization, both in the classroom and school-wide. These findings are consistent with prior studies indicating that teachers who perceive a program to be effective and observe positive results hold favorable attitudes toward the program and perceive the time put into delivering the program to be worthwhile, with this ultimately resulting in higher levels of implementation (e.g., Battistich et al., 2004; Han & Weiss, 2005; Rohrbach et al., 1993; Solomon et al., 2000; Taggart et al., 1990).

Implications of these findings would be that the assessment of teacher attitudes at the early stages of program adoption, possibly directly after pre-implementation training, would assist researchers in identifying which teachers may be more likely to deliver a program with lower levels of fidelity. Specific efforts, such as intensive program support from an on-site coordinator or the provision of additional evidence of a program's effectiveness, could then be tailored to those individuals identified as

“high-risk” for implementation infidelity; potentially alleviating issues with poor program implementation. Evidence of teacher beliefs regarding their responsibility to promote and teach social and character development concepts – being thoughtful of others, physical health, social skills – influencing teacher attitudes was observed, supporting our second hypothesis. Thus, considerable utility is likely to result from extended pre-implementation evaluation of teacher beliefs and attitudes. In addition to identifying potentially low-fidelity implementers, the information could be used to track changes in perceptions over time to determine the extent to which these are modifiable and whether a program wields influences over these.

While replication was observed from year 2 to 3, the magnitude of the structural paths from school climate to beliefs, attitudes, and school-wide material usage were reduced, in some cases almost in half, by year 3 (e.g., .66 reduced to .34, see Figure 1). These reductions may speak to a potential leveling-off effect of school climate. Specifically, at the earlier stages of adoption (in this case year 2), the presence of a supportive climate may be of greater necessity for shaping the beliefs and attitudes of those within the system, whereas once the climate has been established, responsibility to implement a program may be more reliant upon teachers. Further, the mean composite scores for school climate increased from year 2 to 3 (see Table 2), along with a reduction in the variance and range, which may account for the large reduction in the estimated path from school climate to beliefs and attitudes at year 3. Conversely, this reduction may have occurred from the program having a positive affect on the schools’ climate, thereby changing the social system within which the program is embedded. This scenario is also supported by the increase in the composite school climate score from year 2 to 3 (see Table 2).

In light of these findings, several limitations need to be considered. First, the data presented were cross-sectional representations of the hypothesized relationships. Thus, causality cannot be established. The decision to not link teachers across years was based on the rationale that greater anonymity would lead to teachers providing more “truthful” answers regarding the amount of program utilization and responses regarding their perception of their peers and administration. Prior studies and reviews have indicated that teacher self-report can be substantially biased (Dane & Schneider,

1998; Lillehoj et al., 2004), with teachers unwilling to indicate they failed to teach or deliver a program with fidelity. Additionally, the processes modeled may be more appropriately considered in a multilevel context, with school climate serving as a level 2 predictor. This was not performed in the current study due to the limited statistical power associated with the number of schools in the sample ( $N = 10$ ) (Raudenbush & Liu, 2000). Conducting such an analysis with the current data would also raise questions surrounding the appropriateness of aggregating individual perceptions of peers and administration to a higher order (i.e., level 2). Future studies should consider collecting objective information on school climate with a larger number of schools to more accurately model these associations. Finally, it is recognized that implementation consists of other components (i.e., responsiveness and program differentiation) that were not analyzed in the current study (Dane & Schneider, 1998; Domitrovich & Greenberg, 2000). Thus, the findings presented here represent only a portion of the entirety of what constitutes implementation.

In conclusion, support was found for the importance of a school's climate in shaping teachers' beliefs and attitudes that ultimately impacted the amount of the program they utilized. We recognize, however, that modifying a school's climate, and with this the beliefs and attitudes of those embedded within it, may present obstacles and, as the cliché goes "be easier said than done". Nevertheless, our findings indicate that these processes play a critical role in whether and how well teachers teach a program's lessons and utilize a program's materials, indicators of implementation fidelity. Future studies should focus on the initial pre-implementation manipulation of school climate factors to determine the extent to which these factors are mutable and can eventually improve program implementation.

#### Funding Disclosure

This project was funded by the National Institute on Drug Abuse, grant #R01-DA13474.

Table 3 (Table 1 of manuscript 1). Alpha reliabilities of full and parceled scales for Year 2 and 3

Latent Construct "Sample Items"	Year 2 Scales		Year 3 Scales		No. Indicators	
	Full	Parceled	Full	Parceled	Full	Parceled
Administrative Support "Are proud of this school" "Accomplish their job with enthusiasm"	.85	.80	.90	.88	7	2
School Connectedness "I feel at home at this school" "People at this school generally get along with each other"	.79	.66	.83	.78	6	2
School Climate (School Connectedness & Admin Support)	.89	.77	.91	.80		4
Attitudes Towards Positive Action "The more effort I put into Positive Action the more effective it is" "I personally benefit from teaching Positive Action"	.91	.90	.93	.91	4	3
Beliefs on Responsibility to Teach SACD concepts "Having positive self-concept" "Peer pressure resistance skills"	.90	.90	.96	.95	14	4
Amount of Positive Action Curriculum Delivered "Unit 1 Thoughts-Actions-Feelings Circle" "Unit 2 Positive Actions for your Body and Mind"	.93	.86	.94	.87	7	4
Classroom Material Usage	.74	.74	.78	.78	3	3
School Wide Material Usage	.64	.64	.61	.61	3	3

Table 4 (Table 2 of manuscript 1). Descriptive statistics of scale scores of latent constructs for Year 2 and 3

Year	Variable	M	±SD	Range
2	Material Usage (Adherence)			
	Classroom	2.8	±1.0	(1.0 - 5.0)
	School-wide	3.6	±1.0	(1.0 - 5.0)
	Amount of <i>Positive Action</i> Curriculum Delivered (Dosage)	3.4	±0.9	(1.0 - 5.0)
	Attitude towards <i>Positive Action</i> (Quality of Delivery)	3.1	±0.6	(1.0 - 4.0)
	Beliefs about one's responsibility to teach SACD concepts	4.0	±0.8	(1.0 - 5.0)
	School climate <sup>a</sup>	3.4	±0.6	(1.0 - 4.5)
3	Material Usage (Adherence)			
	Classroom	2.8	±1.1	(1.0 - 5.0)
	School-wide	3.8	±1.0	(1.0 - 5.0)
	Amount of <i>Positive Action</i> Curriculum Delivered (Dosage)	3.5	±0.9	(1.0 - 5.0)
	Attitude towards <i>Positive Action</i> (Quality of Delivery)	2.9	±0.7	(1.0 - 4.0)
	Beliefs about one's responsibility to teach SACD concepts	4.0	±0.7	(1.9 - 5.0)
	School climate <sup>a</sup>	3.7	±0.5	(2.3 - 4.6)

<sup>a</sup> Composite score composed of perceived administrative support and school connectedness subscales

Table 5 (Table 3 of manuscript 1). Summary of indirect effects from school climate to implementation of the *Positive Action* program

Year	Effects							$\beta$	$z$			
2	Total <sup>a</sup>											
	Climate <sup>b</sup>				→	Classroom Material Usage		.20	7.33			
	Climate				→	School Wide Material Usage		.54	5.88			
	Total Indirect <sup>c</sup>											
	Climate				→	Classroom Material Usage		.20	7.33			
	Climate				→	School Wide Material Usage		.21	2.71			
	Specific Indirect Effects											
	Climate		→	Attitudes	→	Curriculum Delivered	→	Classroom Material Usage	.12	2.92		
	Climate		→	Beliefs	→	Attitudes	→	Curriculum Delivered	→	Classroom Material Usage	.09	2.59
	Climate			→	Attitudes	→	Curriculum Delivered	→	School Wide Material Usage	.12	2.86	
3	Total <sup>a</sup>											
	Climate <sup>b</sup>				→	Classroom Material Usage		.07	2.64			
	Climate				→	School Wide Material Usage		.35	3.55			
	Total Indirect <sup>c</sup>											
	Climate				→	Classroom Material Usage		.07	2.64			
	Climate				→	School Wide Material Usage		.06	1.79			
	Specific Indirect Effects											
	Climate		→	Attitudes	→	Curriculum Delivered	→	Classroom Material Usage	.04	1.64		
	Climate		→	Beliefs	→	Attitudes	→	Curriculum Delivered	→	Classroom Material Usage	.03	3.48
	Climate			→	Attitudes	→	Curriculum Delivered	→	School Wide Material Usage	.03	1.42	
Climate		→	Beliefs	→	Attitudes	→	Curriculum Delivered	→	School Wide Material Usage	.03	1.82	

<sup>a</sup>Total indirect and direct effect estimates; <sup>b</sup>No direct effects specified in the model, estimate includes only indirect effects; <sup>c</sup>Total indirect effect estimates



Table 6 (Table 4 of manuscript 1). Descriptive characteristics of teachers divided into two groups (1) upper 25% and (2) lower 25% distribution of school climate and attitude towards *Positive Action* composite scores for implementation Years 2 and 3

Variables	School Climate												
	Year 2							Year 3					
	Upper 25% Distribution			Lower 25% Distribution			ES	Upper 25% Distribution			Lower 25% Distribution		
	M	±SD	Range	M	±SD	Range		M	±SD	Range	M	±SD	Range
Material Usage (Adherence)													
Classroom	3.1	±1.0	(1.8 - 5.0)	2.6	±1.1	(1.0 - 4.3)	0.46	3.5	±1.1	(2.0 - 5.0)	2.5	±0.9	(1.0 - 4.5)
School-wide	3.9	±0.8	(2.0 - 5.0)	3.3	±1.0	(1.0 - 5.0)	0.67	4.1	±0.8	(2.3 - 5.0)	3.6	±1.1	(1.0 - 5.0)
Amount of <i>Positive Action</i> Curriculum Delivered (Dosage)	3.7	±0.8	(1.0 - 5.0)	3.0	±1.0	(1.0 - 5.0)	0.72	3.9	±0.7	(2.7 - 5.0)	3.3	±1.0	(1.0 - 5.0)
Attitude towards <i>Positive Action</i> (Quality of Delivery)	3.2	±0.5	(2.0 - 4.0)	2.8	±0.7	(1.0 - 4.0)	0.67	3.2	±0.7	(2.0 - 4.0)	2.7	±0.5	(1.0 - 4.0)
Beliefs about one's responsibility to teach SCD concepts	4.3	±0.7	(3.0 - 5.0)	3.6	±1.0	(1.0 - 5.0)	0.85	4.3	±0.7	(3.0 - 5.0)	3.8	±0.7	(1.9 - 5.0)
Variables	Attitude towards <i>Positive Action</i>												
	Year 2							Year 3					
	Upper 25% Distribution			Lower 25% Distribution			ES	Upper 25% Distribution			Lower 25% Distribution		
	M	±SD	Range	M	±SD	Range		M	±SD	Range	M	±SD	Range
Material Usage (Adherence)													
Classroom	3.3	±0.9	(1.5 - 5.0)	2.6	±1.0	(1.0 - 4.5)	0.71	3.5	±1.0	(1.0 - 5.0)	2.7	±1.3	(1.0 - 5.0)
School-wide	4.1	±0.8	(1.0 - 5.0)	3.5	±1.0	(1.0 - 5.0)	0.62	4.2	±1.0	(1.0 - 5.0)	3.5	±1.0	(1.0 - 5.0)
Amount of <i>Positive Action</i> Curriculum Delivered (Dosage)	4.1	±0.6	(3.0 - 5.0)	3.3	±0.9	(1.0 - 5.0)	1.03	4.2	±0.6	(3.0 - 5.0)	3.0	±1.1	(1.0 - 5.0)

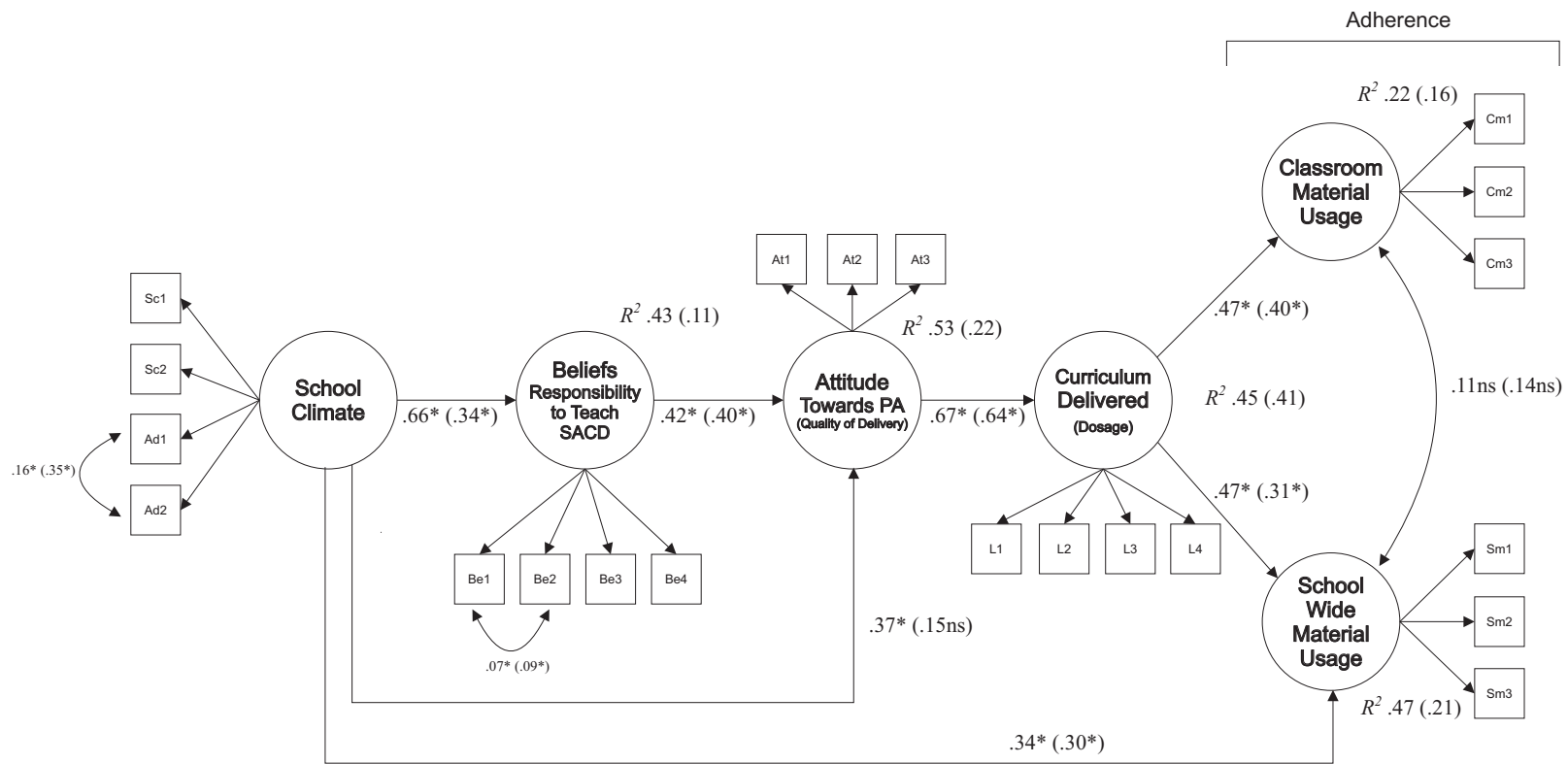


Figure 3 (Figure 1 of manuscript 1). Year 2 and 3 (estimates in parentheses) model depicting school climate (school connectedness and administrative support) and teacher characteristics (beliefs regarding their responsibility to teacher social and character development concepts and attitude towards *Positive Action*) related to the amount of the *Positive Action* curriculum delivered and corresponding program material usage in the classroom and school-wide.

Model fit indices: Year 2  $\chi^2_{df 179} (N = 171) = 305.92, p < .001, CFI = .952, TLI = .944, RMSEA = .064$  and Year 3  $\chi^2_{df 179} (N = 191) = 299.03, p < .001, CFI = .943, TLI = .933, RMSEA = .059$  “\*” indicates significant estimates ( $z \geq 1.96$ ). Abbreviations: SCD = social and character development; PA = *Positive Action*; Latent variable indicators: Sc1 and Sc2 correspond to the parceled items representing school connectedness; Ad1 and Ad2 correspond to the parceled items representing perceived administrative support; Be1-Be4 correspond to the parceled items representing beliefs about the responsibility to teach SCD concepts; L1-L4 correspond to the parceled items representing the amount of the *Positive Action* curriculum delivered; Cm1-Cm3 correspond to the items representing classroom material usage; and Sm1-Sm3 correspond to the items representing school-wide material usage

## Chapter 3. Manuscript 2

Linking student responsiveness to prosocial outcomes: Findings from the multiyear effectiveness trial of the *Positive Action* program

Michael W. Beets, Samuel Vuchinich, Alan Acock, Carol G. Allred, and Brian R. Flay

Journal: Health Education & Behavior  
Address: Society for Public Health Education  
750 First St. NE, Suite 910  
Washington, DC 20002-4242  
Status: In review

### Abstract

Audience responsiveness consists of the perceptions, positive and negative, from the intended audience about a given program during and after it has been delivered. In school-based programs, students are the audience. This study examined the impact of student responsiveness ratings on prosocial behaviors and feelings, and substance abuse rates in a multiyear effectiveness trial of the school-based prevention program *Positive Action*. Elementary students ( $N = 2,926$ ), beginning in first and second grades, completed 4 year-end surveys assessing prosocial behaviors and feelings (years 1-4), student responsiveness towards *Positive Action* (years 3 and 4), and substance abuse (year 4, 5<sup>th</sup> grade only,  $n = 458$ ). Longitudinal path models indicated that student responsiveness had direct positive effects on behaviors and feelings, with this leading to reduced substance abuse rates at year 4. Consistent with theory, student responsiveness was directly related to program outcomes, indicating that those students who perceived *Positive Action* as worthwhile achieved more positive outcomes. Examining factors related to student responsiveness and its relation to program outcomes should play a central role in the understanding of *how* a school-based program is effective.

Keywords: Customer Satisfaction, Implementation, Children, Prevention

*Linking student responsiveness to prosocial outcomes: Findings from the multiyear effectiveness trial of the Positive Action program*

In the context of school-based prevention programs, *customer satisfaction* consists of the perceptions of the intended audience to a given program after it has been delivered (Calvert & Johnston, 1990). During the program delivery process, students serve as the active consumers of a prevention program. Students' judgments concerning whether the program is a) useful, b) age appropriate, c) likely to result in beneficial effects for them, and d) acceptable, play a critical role in determining whether a program will realize its goals (e.g., reduced substance abuse, improved prosocial behaviors) (Kirigin, Braukmann, Atwater, & Wolf, 1982; Tolan, Hanish, McKay, & Dickey, 2002). The term *customer satisfaction* is synonymous with *participant responsiveness*, one of the core components of implementation fidelity (Dane & Schneider, 1998; Domitrovich & Greenberg, 2000). The assessment of whether the program is viewed favorably by consumers is, therefore, an indispensable indicator for understanding why a program does or does not result in both short- and/or long-term maintenance of behavior change (Kennedy, 2002; Schwartz & Baer, 1991).

One of the overarching goals of designing effective prevention programs is to not only change targeted behaviors, but also to create a product (i.e., program) that is liked by the consumer. This latter point is of considerable interest, since it is more likely that treatments will be adhered to, concepts and skills utilized after removal of treatment, and long-term outcomes realized from those programs that consumers enjoy (Kennedy, 2002; Kirigin et al., 1982; Reimers & Wacker, 1992; Reimers, Wacker, & Koeppel, 1987; Tolan et al., 2002). Although it is intuitively logical that those consumers who "like" a program are most likely to personally benefit from it, little work has specifically linked student responsiveness ratings of school-based prevention programs to program-related outcomes. In the few studies that have used responsiveness as a predictor of program outcomes, the results are promising. Kirigin and colleagues (1982), in an evaluation of the Teaching-Family (Achievement Place) program, designed to reduce juvenile criminal behavior, found that youth who indicated greater satisfaction with the program (i.e., higher levels of responsiveness)

were less likely to be involved in criminal activity by treatment's end. Likewise, larger improvements in parenting-related practices (e.g., care, fair discipline) were observed by those parents that rated participation in family-based therapy more favorably (Tolan et al., 2002).

Apart from these examples, most studies typically include student reports of responsiveness as part of a composite measure of program fidelity (Hansen et al., 1991; Rohrbach et al., 1993). By including student responsiveness with other indicators of implementation (e.g., dosage, quality of delivery) (Dane & Schneider, 1998; Domitrovich & Greenberg, 2000), practitioners are unable to determine its independent contribution to program outcomes. Even fewer studies have examined why students do or do not like a prevention program. Of those that have, high levels of responsiveness towards a program (termed "acceptance" by Rohrbach et al., 1993) were related to high levels of program fidelity (Rohrbach et al., 1993). Program fidelity, in this case, consisted of teacher enthusiasm, the number of program lessons taught, and appropriateness of delivery (Rohrbach et al., 1993). Although limited work exists to contrast these results, it appears that when a program is delivered with minimal modifications and done so with considerable buy-in from the teachers (i.e., high fidelity), students report more favorable attitudes towards it than when a program is delivered poorly (i.e., low fidelity).

Utilizing consumer satisfaction and value concepts from the service, marketing, and business fields of study, two components are theorized to predict consumer satisfaction. These are the consumers' *expected results* from using the product and the *quality* with which the product was delivered (Heskett, W. E. Sasser, & Schlesinger, 1997). In relation to school-based prevention programs, the *expected results* refer to the outcomes expected from engaging in and using the core concepts embedded within a program. For instance, if a program focuses on the development of drug resistance skills, the perceived outcomes for the consumers, this being whether using the skills worked or did not, are tied directly to the value they place on the program. Subsequently, the results of using the skills will determine whether they will utilize them in the future.

In reference to the latter component, in a school-based setting the deliverers are usually teachers (Han & Weiss, 2005). Thus, the way in which teachers present program materials (e.g., with enthusiasm), the amount of the curriculum they deliver (e.g., number of lessons taught), whether they are good role models of the program's core concepts, and how much they utilize program-related materials (e.g., student recognition rewards) are critical to the success of the intervention (Rohrbach et al., 1993). Given this research and theory, we hypothesized that better delivery of a program (i.e., higher fidelity) by a teacher will lead to more favorable student response to the program (i.e., increased responsiveness). These should lead to improved program outcomes (e.g., increased prosocial behaviors, reduced substance abuse). Hence, theoretically, increases in responsiveness should arise from consumers realizing positive results from participation in and using program-related skills in addition to teachers delivering the program with a high degree of fidelity (Heskett et al., 1997). However, there is a dearth of formal testing of these hypotheses in the literature on school-based prevention programs.

In the present study, we examined the role of student responsiveness (i.e., customer satisfaction) to influencing prosocial behaviors, feelings associated with performing prosocial behaviors, and substance abuse outcomes for elementary students participating in three years of the school-based prevention program, *Positive Action*. Based on prior studies of responsiveness and theories of consumer satisfaction, hypotheses were developed and tested in both cross-sectional and longitudinal models. Specifically, we hypothesized that student responsiveness would be positively related to program exposure. Student responsiveness, in turn, was hypothesized to be positively related to increased prosocial behaviors and feelings, and indirectly related (mediated by prosocial behaviors and feelings) to reduced substance abuse rates of elementary students.

## Methods

### Design

The Hawai'i randomized trial of the *Positive Action* program took place in 20 public elementary schools on three islands in the Hawai'i school district. The sample of students examined in the current study comprises those who attended the 10 schools



receiving the treatment (*Positive Action* program) from baseline (year 1, academic year 2000-01 – 1<sup>st</sup> and 2<sup>nd</sup> grade) to year 4 (academic year 2004-05 – 4<sup>th</sup> and 5<sup>th</sup> grade). Years 2 (2<sup>nd</sup> and 3<sup>rd</sup> grade), 3 (3<sup>rd</sup> and 4<sup>th</sup> grade), and 4 (4<sup>th</sup> and 5<sup>th</sup> grade) correspond to the years in which the *Positive Action* program was implemented by the time of this study; totaling three years of exposure to the *Positive Action* program. Since the focus of the current study was on student responsiveness, only students in the treatment schools were included in the analyses reported here.

Schools were selected to receive the intervention based on random assignment from matched pairs based on a multivariate indicator of risk developed from school-level data on the following characteristics: the proportion of students receiving free/reduced lunch, school size, pupil/teacher ratio, percent stability of student enrollment, ethnic distributions, and indicators of problem behaviors (e.g., suspensions) and achievement (e.g., standardized testing) (c.f., Dent et al., 1993). Additionally, parental written consent, both active (for the substance abuse sub-questionnaire, see below for details) and passive, and student verbal assent (up to grade 3) and written consent (grade 4 onwards) were obtained prior to participation. All assessments and procedures were approved by the IRB at the University of Illinois at Chicago, the Hawaii Department of Education and, for year 4, Oregon State University.

#### Intervention

The *Positive Action* program ([www.positiveaction.net](http://www.positiveaction.net)) is a multicomponent social and character development school-based prevention program. It is grounded in a broad theory of self-concept (Purkey, 1970) and its comprehensive approach can be best described by the Theory of Triadic Influence (Flay & Petraitis, 1994). The former theory posits that people determine their self-concepts by what they do; that actions, more than thoughts or feelings, determine self-concept; and that making positive and healthy behavioral choices results in feelings of self-worth. The *Positive Action* program teaches children what actions are positive, that they feel good when they do positive actions, and that they then have more positive thoughts and future actions. The theory of triadic influence suggests a holistic approach to reducing behavior problems by addressing distal and proximal influences within the same program. The

areas targeted in the *Positive Action* program are school climate, student-teacher relationships, parent involvement, instructional practices, and the development of the self-concept of all parties involved (e.g., students, teachers, parents).

The *Positive Action* program is available for K-12 students (only the elementary curriculum was utilized in the current evaluation trial). The curriculum is scoped and sequenced and at the elementary level consists of almost daily, 15-20 minute lessons (140 lessons per grade, per academic year) based on six units. The six units cover topics related to self-concept, mind and body positive actions (e.g., nutrition, physical activity, decision-making skills, motivation to learn), and social/emotional actions for managing yourself responsibly (e.g., emotional regulation, time management), getting along with others (e.g., empathy, respect), being honest with yourself and others, and self-improvement (e.g., goal setting, courage to try new things). The program includes an extensive set of program materials: posters, classroom and school-wide recognitions and announcements, as well as token rewards to reinforce classrooms for positive student behaviors throughout the school day.

At the beginning of each academic year, teachers from the 10 treatment schools attended *Positive Action* training sessions conducted by the developer of the program (Carol Allred). The training sessions lasted approximately 3-4 hours for the initial year (year 2), and 1-2 hours for each successive year (years 3 and 4). Booster sessions, conducted by the site coordinator, were provided at least once during the academic year for each school. These lasted approximately 30-50 minutes. Additionally, one-day min-conferences, held in February of each year, were conducted to bring together 5-6 staff members from each of the 10 participating schools in order to share ideas and experiences as well as answer any questions regarding implementing the program.

### Sample

The sample of students varies across academic year, with the variation attributable to the addition/loss of students due to relocation (i.e., moving in and out of the treatment schools), absence during survey administration and receipt or withdrawal of parental consent and/or verbal assent. The average number of students surveyed in the treatment schools at any one year was 1,523, with the number ranging from 1,399

to 1,574. Approximately 50% of the sample was girls, with the majority of students self-identifying as Hawaiian or part Hawaiian (25%), Filipino (15%), White non-Hispanic (16%), Japanese (6%), or other (e.g., Chinese, African American, Korean, Samoan, Portuguese). The age range of the students was 7 to 9 years at baseline and 10 to 12 at year 4 survey administration.

#### Year-end survey

#### Main Program Outcomes

**Behaviors and feelings.** At the end (May/June) of each of the four school years, students completed a year-end survey asking about their involvement in positive and negative behaviors and how they feel about themselves if or when they do each of the behaviors. Behavior items (referred to as “Do” items) consisted of 19 core items administered across all 4 years of the study. Sample items were “do you work hard in school?”, “do you respect others?”, and “do you get into fights?”. Response categories were based on a 3 point scale ranging from “no”, “sometimes”, to “yes”. After reversal of scores for negative behaviors, the mean of the 19 items was computed, representing an overall Do (i.e., positive behavior) score for each survey year. Alpha reliabilities for the Do scale were .70, .79, .76, and .82, for years 1 through 4, respectively.

Feeling items (referred to as “Feel” items) consisted of how the students felt about themselves if or when they performed the corresponding 19 core Do items. Examples include “how do you feel when you work hard in school?”, “how do you feel when you respect others?”, and “how do you feel when you get into fights?”. Response categories were based on a 4 point scale ranging from “very bad” to “very good”. Similar to the Do items, after reversal of scores for negative behaviors, the Feel items were averaged, representing an overall Feel (i.e., feelings about positive behaviors) score for each year. Alpha reliabilities for the Feel scale were .77, .87, .84, and .88, for years 1 through 4, respectively. For both the Do and Feel scales, higher scores represent more positive and less negative behaviors and feelings.

**Substance abuse.** Once students reached grade 5, active parental consent was obtained to ask students ( $n = 458$ , 50% girls) about substance use/abuse. Students were asked whether they had ever “smoked cigarettes”, “drank alcohol”, “gotten drunk”, and “used illegal drugs (e.g., marijuana)”. Responses ranged from “no, never”,

“yes, once”, to “yes, more than once”. For the cross-sectional analyses (described below), because the prevalence of usage was low for any one item, the 4 items were summed and the summed variable dichotomized to represent students who never used drugs (“no, never”, 88.7%) versus those that reported usage on any of the 4 items (“yes, once” and “yes, more than once”, 11.3%). For the longitudinal analyses (described below), the four items were left in their original 3-pt scale format (i.e., “no, never”, “yes, once”, and “yes, more than once”) and modeled as observed indicators of a latent construct, “substance abuse” in the longitudinal structural equation models.

### Predictors of Program Outcomes

Program exposure. Included in the student survey at year 3 and 4 were a series of 13 items referring to the amount of the *Positive Action* program the students received from their teacher or engaged in as part of school-wide activities. The questions asked the student to report the number of times their teacher used classroom materials (e.g., ICU Box, tokens), how often they participated in school-wide *Positive Action* activities (e.g., attendance at *Positive Action* assemblies), and the number of lessons they recall the teacher delivering on an average week. All items, except the number of lessons taught, were answered on a 4-pt scale ranging from “0” to “3 or more”. The number of lessons taught on most days of the week was answered using a 5-pt scale ranging from “0” to “4 or more days per week”. The responses on these items were averaged, by year, to create a single score representing program exposure for each year. Alpha reliability was .77 and .84 for year 3 and 4, respectively.

Student responsiveness. Beginning at year 3, a 6-item inventory was included in the year-end survey that assessed students’ views of the acceptability of, and judgments about the perceived benefits attained from participation in the *Positive Action* program. Items asked about (a) overall satisfaction with the program, (b) whether they liked what they talked about and did in the program, (c) whether the program had helped them make better decisions, and (d) behave better at home or (e) at school, and (f) whether they felt better about themselves. Response categories were based on a 4-point scale ranging from “none of the time” to “all of the time”. The responses for these items were averaged for each year, with the resulting score

representing overall student responsiveness. Alpha reliabilities of the scale were .85 and .91 for year 3 and 4, respectively.

Survey administration. Paper and pencil surveys were administered to students during regularly scheduled classroom time near the end of each academic year (May/June). Project staff administered the survey to eligible students (i.e., those who returned parental consent and provided assent) in the classroom while the teacher was asked to sit aside and do other work and/or supervise students who did not participate in the survey. Staff worked in teams of two, with one serving as the reader, the other as a proctor. This procedure was used in order to ensure comprehension of items and that all students finished the survey together. Average time to complete the survey was 45 minutes.

### Analyses

Means, standard deviations, and frequencies were computed for all variables where appropriate (see Table 1). To establish the relationship of responsiveness to program outcomes (i.e., behaviors [Do], feelings [Feel], and substance abuse) an ordered series of analyses were conducted, beginning with cross-sectional analyses followed by longitudinal models.

--- Insert Table 1 about here ---

### Comparison of respondents vs. non-respondents to substance abuse

Since not all students at year 4 responded to the substance abuse survey, independent *t*-tests were calculated comparing behavior and feeling subscales and responsiveness reports for those students who completed the substance abuse and *Positive Action* survey versus those students who completed the *Positive Action* survey only.

### Testing for cross-sectional relationships

Student responsiveness with behaviors and feelings. For years 3 and 4, separately, ordinary least squares (OLS) regression models were specified with the composite behavior (Do) and feeling (Feel) scale scores serving as the dependent variable. Prior to analysis, the Do and Feel composite scores were linearly transformed to the percent of maximum possible score (POMP) (Cohen, Cohen, Aiken, & West, 1999). This transformation allowed for the interpretation of the unstandardized

coefficient in terms of percent change on a scale ranging from 0 to 100%. Student responsiveness scale scores for each year were trichotomized into three groups – low responsiveness (lower 25% of distribution), moderate responsiveness (25 to 75% of distribution), and high responsiveness (upper 25% of distribution). Each category was dummy coded and the low and high responsiveness categories specified as predictors in the OLS regression models, with the moderate responsiveness category serving as the comparison group. All OLS model estimates were adjusted for non-normality of variables, using robust standard errors, and for the clustering of students within classrooms. Postestimation analyses, testing the equivalence of the point estimates of low and high responsiveness coefficients, were conducted to test the hypothesis that the difference between the low and high responsive coefficients was zero. Thus, one coefficient gives the percent difference between the moderate and high response groups, and the other coefficient gives the percent difference between the moderate and low response group.

Substance abuse. Using logistic regression models, year 4 Do and Feel summary scores were regressed onto the binary substance abuse variable. Included in the logistic model were gender (boys = 1) and the dummy variables of low (< 25% distribution) and high (> 25% distribution) responsiveness at year 4. Cross-sectional odds ratio estimates were adjusted for non-normality of variables and for the clustering of students within classrooms. All cross-sectional models were conducted using Stata v9.2 (StataCorp, College Station, TX).

#### Longitudinal path analysis

Based on the cross-sectional findings, two path models were specified using Do and Feel items from years 1 through 4, responsiveness reports and program exposure from year 3 and 4, and gender (boys = 1) as a covariate (see Figure 1). The second model, for the sub-sample of students ( $n = 458$ ) who completed the substance abuse survey and had information on the remaining model variables, included a latent substance abuse outcome at year 4 (see Figure 2). For each model, Do and Feel composite scores were specified to affect subsequent composite scores (e.g., Do year 1

predicting Do year 2).<sup>5</sup> At year 3 and 4, program exposure and responsiveness composite scores were included. According to our hypotheses, program exposure was specified to predict student responsiveness. Additionally, program exposure and student responsiveness were hypothesized to affect their corresponding (within the same year) Do and Feel composite scores. Further, student responsiveness at year 3 was hypothesized to have a lagged effect on year 4 reports of Do and Feel, and responsiveness. For the second model, year 4 Do and Feel composite scores were specified to affect reports of substance abuse rates. All estimates are standardized path coefficients. Significance was denoted by a  $z$  value  $\geq 1.96$ .

Both longitudinal models were estimated using full information maximum likelihood with an estimator robust to non-normality. Since students changed classrooms, and thus experienced different student groupings across years, nesting was specified at the school level. Longitudinal analyses were performed using *Mplus* v4.2, (Muthén and Muthén, [www.statmodel.com](http://www.statmodel.com)). Model-data fit was assessed by the following indices: Chi-Square statistic, root mean square error of approximation (RMSEA), Comparative Fit Index (CFI), and the Tucker Lewis Index (TLI) (Hu & Bentler, 1999; Kline, 2005). Values for the RMSEA ranging from 0.05 to zero, and CFI and TLI above 0.90 and 0.95, respectively, represent acceptable fit of the model (Hu & Bentler, 1999; Kline, 2005).

## Results

### Comparison of respondents vs. non-respondents to substance abuse

Results indicated that students who responded to the substance abuse portion of the questionnaire reported more positive feelings at year 1 ( $3.60 \pm 0.33$  vs.  $3.51 \pm 0.34$ ) and less positive feelings at year 3 ( $3.55 \pm 0.33$  vs.  $3.62 \pm 0.33$ ), less positive behaviors at year 4 ( $2.58 \pm 0.26$  vs.  $2.62 \pm 0.25$ ), and lower levels of responsiveness at year 3 ( $2.93 \pm 0.78$  vs.  $3.10 \pm 0.80$ ) in comparison to those students who chose not to respond to the substance abuse items.

### Cross-sectional estimates

---

<sup>5</sup> Initially, cross-lagged models were specified for behaviors predicting subsequent feelings and vice versa (e.g., Do year 1 predicting Feel year 2, Feel year 2 predicting Do year 3). No significant relations were observed so these paths were removed from the final model.

The results from the cross-sectional OLS regression for year 3 and 4 are presented in Table 2. For the behaviors (Do) and feeling (Feel) items, a linear effect for responsiveness was observed, with students that reported high responsiveness at year 3 or 4 exhibiting approximately 13 to 15% (range 11.9 to 17.8%; POMP transformed scale scores) higher scores on behaviors and their corresponding feelings than students who reported low responsiveness (see Table 3).<sup>6</sup> The relationship of behaviors, feelings, and responsiveness to the binary substance abuse use variable at year 4 are presented in Table 3. A 79% reduction in substance abuse rate (odds ratio 0.21, 95% confidence interval, .08, .58) was observed for every one unit increase in positive feelings (scale ranging from 1 to 4). No other associations were significant.

--- Insert Tables 2 & 3 about here ---

#### Longitudinal path analyses

The results of the longitudinal path analyses are presented in Figures 1 and 2, with indirect model estimates for both models presented in Table 4. Model fit indices indicated good model-data fit, with estimates for the overall model (Figure 1) of:  $\chi^2$  (df = 38, N = 2,926) = 126.66,  $p < .001$ , CFI = .968, TLI = .937, and RMSEA = .028; and for the sub-group analysis (Figure 2):  $\chi^2$  (df = 91, N = 458) = 155.94,  $p < .001$ , CFI = .951, TLI = .928, and RMSEA = .039.

--- Insert Figures 1 & 2 and Table 4 about here ---

For the overall model (see Figure 1), in support of our hypotheses, student responsiveness was positively associated with behaviors (Do) and feelings (Feel) at their corresponding year (e.g., year 3 student responsiveness predictive of year 3 behaviors and feelings, and similarly for year 4). Additionally, the hypothesis that responsiveness would be predictive of subsequent behaviors and feelings was supported, with student responsiveness at year 3 positively associated with behaviors and feelings at year 4.

Behaviors at year 2 predicted student responsiveness at year 3, and this pattern was repeated from year 3 to year 4, while feelings at year 2 and 3 were not related to year 3 or 4 student responsiveness. Support was also observed for the hypothesis that

---

<sup>6</sup> OLS models were estimated using student responsiveness as a continuous variable; significant effects were observed for all models, supporting the categorical comparisons



specified that program exposure would be predictive of student responsiveness, suggesting that the more of the program the students were exposed to the greater their reported level of satisfaction. The level of student responsiveness reported at year 3 was positively related to subsequent student responsiveness at year 4. Indirect estimates indicated that the effect of student responsiveness at year 3 on behaviors and feelings at year 4 was partially mediated through student responsiveness at year 4, and behavior and feelings at year 3 (see Table 4).

Estimates for the sub-group analysis, with substance abuse rates assessed at year 4, are presented in Figure 2. Overall, many of the hypothesized associations supported in the overall model (see Figure 1) are replicated in the sub-sample. However, variations were observed, with non-significant lagged effects of student responsiveness at year 3 on year 4 behaviors and feelings. As with the main model, feelings at prior years failed to predict student responsiveness at successive years, whereas behaviors maintained their lagged-effect. For the main outcome, the latent construct of substance abuse, student reported feelings at year 4 had a significant negative affect on substance abuse, indicating that more positive feelings about positive behaviors were related to lower rates of substance abuse. Student responsiveness at year 3 had an indirect effect on the reduction of substance abuse at year 4 (see Table 4). This effect was mediated through responsiveness and feelings at year 4, and feelings at year 3 and 4. The effect of student responsiveness at year 4 on reduced substance abuse was mediated through feelings at year 4. The effect of gender (boys) on the model estimates (Figure 1 and 2) are presented in Table 5. Overall, boys exhibited lower rates of positive behaviors and feelings about doing them across all 4 years. However, there were no significant effects of gender on responsiveness or substance abuse.

--- Insert Table 5 about here ---

### Discussion

Student responsiveness plays a critical role in the outcomes realized from prevention programs (Hansen et al., 1991; Kirigin et al., 1982; Tolan et al., 2002). Additionally, program responsiveness, synonymous with customer satisfaction in the marketing literature, is identified as one of the key sources of information to

understand *why* and *how* a prevention programs work (Dane & Schneider, 1998). Students' views are considered essential in modifying current program implementation practices, as well as guiding future program development. However, few studies of school-based prevention programs have attempted to specifically link program-related outcomes to reports of students' satisfaction with a program. In the current study, customers were defined as those students who received the school-based prevention program, *Positive Action*. In accordance with our hypotheses and theoretical linkages between customer satisfaction and value from service-based industries (Heskett et al., 1997), student responsiveness/satisfaction was positively associated with increased prosocial behaviors and their corresponding feelings. Further, student responsiveness at prior years exhibited indirect effects on lowering substance abuse rates at year 4 of the study. Thus, consistent with other studies (Kirigin et al., 1982; Tolan et al., 2002), the program produced greater results for those students who liked it more.

Of interest, student responsiveness was related to the amount of the program the students received at years 3 and 4 of the trial. Rohrbach and colleagues (1993) observed similar associations. In their evaluation, they observed that higher levels of program acceptance (analogous to customer satisfaction and student responsiveness) were reported from those students in classrooms where the teacher delivered the program with a high level of fidelity. Fidelity, in this case, was defined by Rohrbach and colleagues as a composite score that consisted of classroom management, teacher and student enthusiasm, and quality of delivery. In the current study, implementation fidelity was conceptualized as the amount of the program-specific materials the students reported the teacher using in the classroom and school-wide, along with the average number of lessons they recalled their teacher delivered during a typical week. Yet, despite the differences between the measures of fidelity in the current study and Rohrbach et al. (1993), the similarity of the findings provide substantial evidence that when programs are properly delivered, done so with enthusiasm and on a regular basis, and delivered in conjunction with the use of their corresponding materials to reinforce the program's core concepts, students *like* the program to a greater extent and with this, increased program outcomes are attained.

One of the important findings was the link between student responsiveness and its affect on subsequent satisfaction and feelings, with these leading to reduced rates of substance abuse (see Figure 2). The conceptualization of the feelings (Feel items) variable in the current study is comparable to affective states and can be considered a measure of the students' general feelings about the self, with items related to how students feel about themselves when they think about how others feel (empathy), work hard in school and set goals (future orientation), and try to make themselves a better person (self-esteem). Consistent with the multistage social learning theory (MSLT) (Simons, Conger, & Whitbeck, 1988) and the Theory of Triadic Influence (Flay & Petraitis, 1994), deficiencies in these intrapersonal characteristics are likely to place youth at greater risk for experimental substance abuse during adolescence. Specifically, within these theories, youth are theorized to engage in substance abuse behaviors through interactions with deviant peer which is related to lack of social skills (e.g., being unempathetic) and a present-oriented value system. While the current study did not examine deviant peer interactions, the parallels between the hypothesized linkages of more positive intrapersonal characteristics and reduced substance abuse in the current study and those proposed by these theories are indicative that these processes may have occurred. Conversely, in a comprehensive review of illicit substance abuse among adolescents, Petraitis and colleagues (1998) found intrapersonal-level affective states (e.g., anxiety, depressed mood, low self-esteem) to exhibit minimal influence on illicit substance usage in the majority of the prospective studies reviewed. The present study, however, indicated a direct link between higher levels of positive feelings about the self and reduced substance abuse ratings (see Figure 2). A possible explanation may lie in the non-clinical measure of feelings in the current study, with the focus on a broader or overall measure of general feelings, as stated previously, rather than a focus on one or more specific constructs of intrapersonal feelings (e.g., self-esteem solely).

### *Limitations*

The model accounted for a sizable portion of variance in the prosocial behavior and feeling scores (upwards of 33%). However, while the findings illustrated the link between having positive views about a prevention program and program-related

outcomes, the current analysis cannot account for a considerable portion of the variance in student responsiveness ratings, with the models tested accounting for approximately 10 to 19% (see Figure 1 and 2) of the variance. Notably, however, was the absence of a relationship of gender to responsiveness ratings in either year 3 or 4 (see Table 5). This provides an indication that the *Positive Action* program is equally appealing to both sexes, thereby reducing the potential issues of the program favoring one gender over the other. This suggests other factors, not measured in the current study, are influencing student perceptions about the program. For an answer to this, future studies could utilize focus group methodology, with these composed of known groups – students who are satisfied and unsatisfied with the program. This approach might lead to greater understanding of why program components, such as lesson content and the use of token rewards, are viewed as appropriate or inappropriate. Further, the students assessed in the current study were from a single state (Hawai'i) and therefore the findings may not generalize to the entire population of elementary students.

#### *Implications for Practitioners*

The findings from this study highlight the importance of eliciting feedback from the intended audience (i.e., students) about the working processes of a program. This includes active monitoring of participant response to the program while the program is being delivered. When participant response is negative, some adjustments should be made in order to increase beneficial program effects. Key processes to monitor include method of delivery of the program by teachers, the programs content (e.g., lessons), use of program-specific materials, student perceptions of whether the program is viewed as age-appropriate, culturally relevant, and consistent with the prevailing norms adopted by a generation. Thus, systematic and ongoing evaluation of student views regarding a program may prove useful in modifying a program for maximum impact.

#### *Conclusions*

In conclusion, this study showed that participant response to a preventive intervention made a substantial contribution to the impact of the program on behaviors, feelings and substance use in at-risk elementary school children.

Responsiveness, in turn, was predicted by positive behavior engagement and led to increased positive behaviors and feelings towards those behaviors, both concurrently and successively. The results of this study suggest that participant response should play a central role in the evaluation and on-going modification of school-based program delivery.

#### Funding Disclosure

This project was funded by the National Institute on Drug Abuse, grant #R01-DA13474.

Table 7 (Table 1 of manuscript 2). Descriptive statistics across assessment years 1 through 4 for boys and girls

Variable	Year	Girls		Boys	
		M	±SD	M	±SD
Behaviors	1	2.67	±0.21	2.58	±0.25
	2	2.66	±0.23	2.56	±0.29
	3	2.69	±0.20	2.58	±0.25
	4	2.67	±0.22	2.56	±0.26
Feelings	1	3.58	±0.30	3.48	±0.36
	2	3.60	±0.33	3.50	±0.39
	3	3.68	±0.26	3.52	±0.37
	4	3.64	±0.30	3.55	±0.35
Positive Action Program Exposure	3	1.37	±0.70	1.28	±0.69
	4	1.24	±0.72	1.24	±0.75
Student Responsiveness	3	3.10	±0.76	2.94	±0.82
	4	2.92	±0.84	2.78	±0.84

Table 8 (Table 2 of manuscript 2). Cross-sectional relationship between student responsiveness and behaviors (Do) and feelings (Feel) for year 3 and 4

Variables	Do (POMP)				Feel (POMP)			
	Unstandardized Coefficient	(95CI)	SE	<i>t</i> *	Unstandardized Coefficient	(95CI)	SE	<i>t</i> *
Year 3 ( <i>n</i> = 1,285)								
Low SR <sup>a</sup> ( <i>n</i> = 363)	-8.50	(-10.13, -6.86)	0.82	-10.36	-8.43	(-9.93, -6.93)	0.75	-11.22
High SR <sup>b</sup> ( <i>n</i> = 357)	5.37	(4.13, 6.61)	0.62	8.62	6.41	(5.48, 7.34)	0.47	13.74
Constant	82.37	(81.46, 83.29)	0.46	180.10	87.40	(86.65, 88.15)	0.38	232.47
Low = High SR <sup>c</sup>	-13.86	(-15.74, -11.99)	0.94	-14.76	-14.84	(-16.40, -13.28)	0.78	-18.95
<i>R</i> <sup>2</sup>				.17				.21
Year 4 ( <i>n</i> = 1,376)								
Low SR <sup>a</sup> ( <i>n</i> = 343)	-7.93	(-9.65, -6.21)	0.86	-9.21	-6.86	(-8.58, -5.15)	0.86	-8.00
High SR <sup>b</sup> ( <i>n</i> = 342)	7.45	(5.72, 9.19)	0.87	8.57	6.98	(5.73, 8.24)	0.63	11.11
Constant	80.61	(79.57, 81.64)	0.52	155.91	86.25	(85.27, 87.23)	0.49	176.34
Low = High SR <sup>c</sup>	-15.38	(-17.77, -12.99)	1.19	-12.89	-13.85	(-15.77, -11.93)	0.96	-14.42
<i>R</i> <sup>2</sup>				.18				.19

Abbreviations: POMP = percent of maximum possible score; SR = student responsiveness

\* significant at  $p < .001$

<sup>a</sup> Low student responsiveness (lower 25% distribution)

<sup>b</sup> High student responsiveness (upper 25% distribution)

<sup>c</sup> Postestimation test for equality of Low and High student responsiveness coefficients

Table 9 (Table 3 of manuscript 2). Cross-sectional logistic regression estimates for substance abuse (year 4) and program outcomes ( $n = 442$ )

Variable	OR	(95CI)	<i>z</i>	<i>p</i>
Feelings	.21	(.08, .58)	3.04	.002
Behaviors	.36	(.07, 1.81)	0.93	.355
Student Responsiveness				
Low	.98	(.49, 1.99)	0.04	.964
High	.75	(.39, 1.44)	0.87	.382
Boys	.94	(.77, 1.15)	0.59	.554



Table 10 (Table 4 of manuscript 2). Estimate indirect effects from the full and sub-sample longitudinal structural equation models (Figure 2 and 3)

Model	Estimated Indirect Effects	Standardized Effect	<i>z</i>
Full	SR3 → SR4 → Do4	.09	3.32
	SR3 → Do3 → SR4 → Do4	.01	2.60
	SR3 → Do3 → Do4	.08	5.10
	SR3 → SR4 → Feel4	.11	4.18
	SR3 → Do3 → SR4 → Feel4	.01	2.53
	SR3 → Feel3 → Feel4	.06	3.77
Sub-sample	SR4 → Feel4 → Drug	-.11	-2.28
	SR3 → SR4 → Feel4 → Drug	-.03	-2.51
	SR3 → Feel3 → Feel4 → Drug	-.02	-2.00

Abbreviations: SR = Student Responsiveness; Do = Behaviors; Feel = Feelings corresponding to the behaviors

Table 11 (Table 5 of manuscript 2). Effects of gender (boys) on behaviors (Do), feelings (Feel), student responsiveness, and substance abuse

Variables (numbers refer to year survey administered)	Overall Model <sup>a</sup>		Sub-sample Model <sup>b</sup>	
	Standardized Effects	$z^{\dagger}$	Standardized Effects	$z^{\dagger}$
Student Responsiveness				
Year 3	-.06	-1.60	-.05	-1.08
Year 4	-.03	-1.16	-.04	-0.84
Behaviors				
Do1	-.20	-6.81	-.15	-2.11
Do2	-.14	-6.72	-.09	-1.56
Do3	-.16	-4.87	-.15	-3.29
Do4	-.13	-5.42	-.11	-2.66
Feelings				
Feel1	-.15	-7.91	-.21	-3.12
Feel2	-.12	-5.33	-.03	-0.56
Feel3	-.17	-6.08	-.15	-3.37
Feel4	-.07	-2.92	-.01	-0.22
Substance Abuse				
Year 4	-	-	.07	1.22

<sup>a</sup> Estimates correspond with Figure 1

<sup>b</sup> Estimates correspond with Figure 2

<sup>†</sup>  $z$  value  $\geq 1.96$  denotes significant at  $p \leq .05$

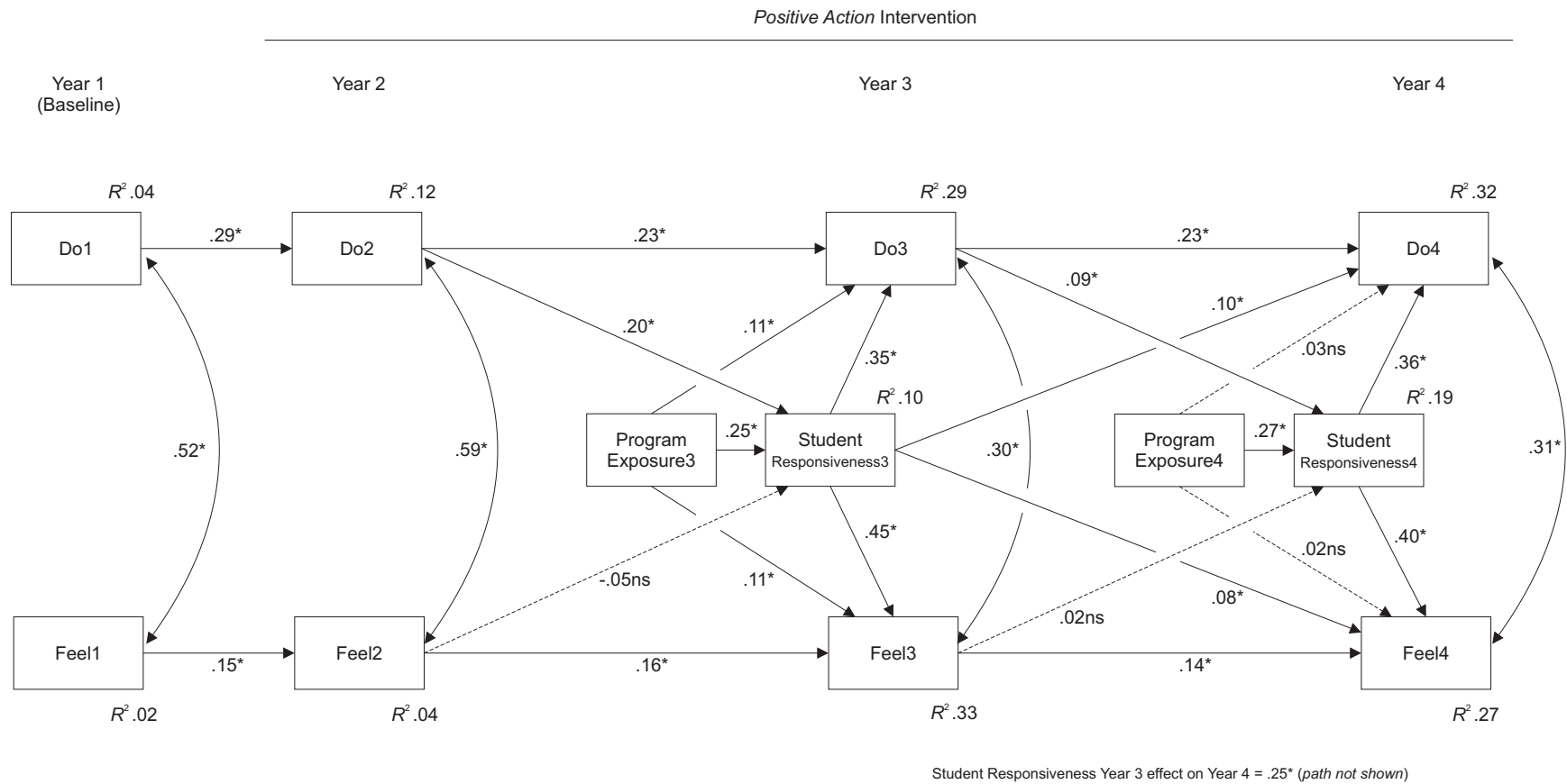


Figure 4 (Figure 1 of manuscript 2). Final estimated model of hypothesized relationships of student responsiveness, program exposure, and program outcomes, behaviors (Do) and feelings (Feel). Gender (boys) entered as a covariate in the model, but not presented in the model illustration (see Table 5). *Dashed lines indicate non-significant paths ( $z < 1.96$ )*

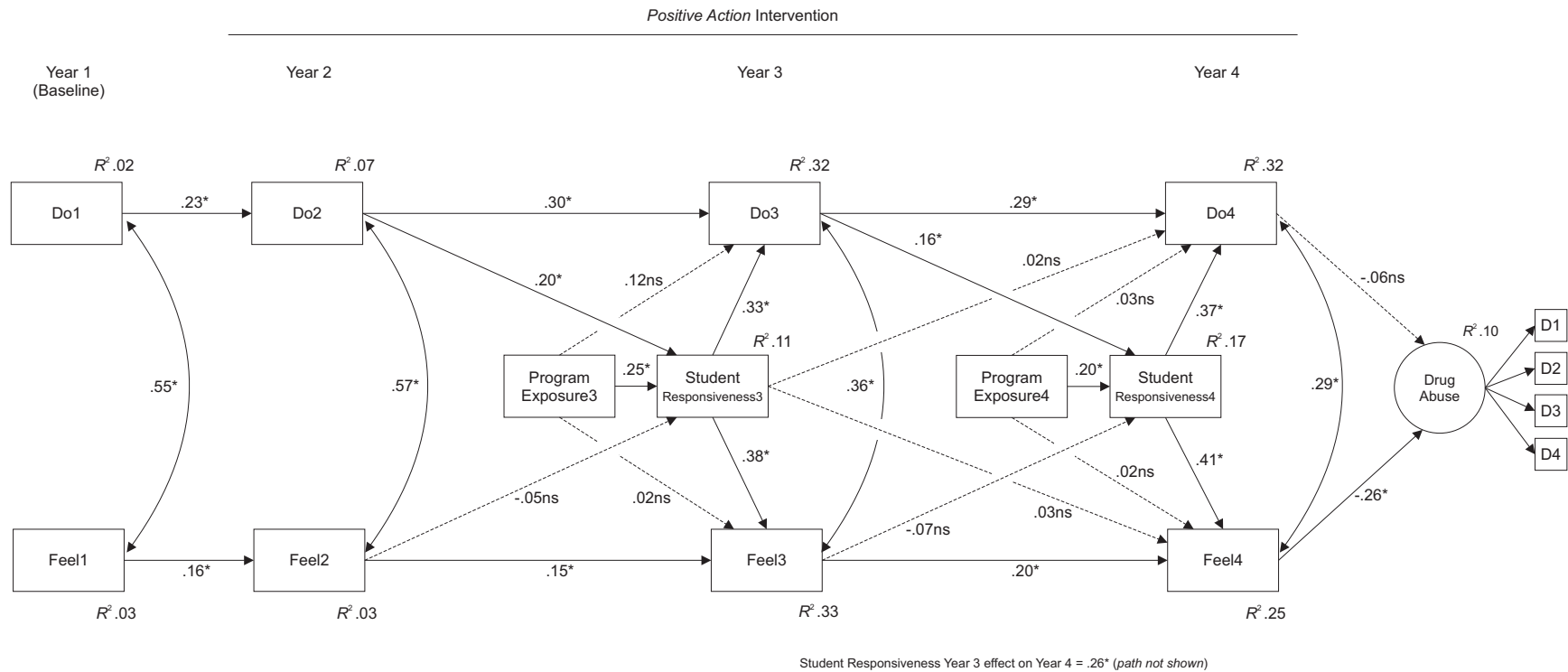


Figure 5 (Figure 2 of manuscript 2). Subgroup model depicting hypothesized relationships of student responsiveness, program exposure, and program outcomes, behaviors (Do) and feelings (Feel), to substance abuse usage rates. Gender (boys) entered as a covariate in the model, but not presented in the model illustration (see Table 5). *Dashed lines indicate non-significant paths ( $z < 1.96$ )*

## Chapter 4. Manuscript 3

Where have all the control groups gone?  
When “business as usual” mirrors experimental conditions

Michael W. Beets and Brian R. Flay

Journal: Health Education and Behavior  
Address: Society for Public Health Education  
750 First St. NE, Suite 910  
Washington, DC 20002-4242  
Status: In review

### Abstract

The extent to which control schools self-initiate or receive programming of similar content to an intervention under evaluation is referred to as program differentiation. This remains one of the least studied aspects of implementation fidelity. One school leader from 18 elementary schools in Hawai'i, that participated in a effectiveness trial of a school-based social and character development program (SACD), responded to a survey regarding: (a) the number of self-initiated SACD programs; (b) the number of hours and weeks devoted to SACD; and (c) whether teachers received formal training to deliver the SACD programs. Control schools ( $n = 8$ ) reported an average of 8 SACD programs vs. 3 (in addition to the program being evaluated) reported by intervention schools. Control schools spent approximately 23wks/yr and 33hrs/program, yet teachers only received training for 45% of the programs adopted. Widespread self-initiation of SACD in control schools may weaken or neutralize the possible effect that can be detected when evaluating school-based interventions.

**Keywords:** Implementation; evaluation; fidelity; children; primary school

*Where have all the control groups gone?**When “business as usual” mirrors experimental conditions*

Systematic documentation of the fidelity of implementation, the degree to which a planned program was carried out as intended (Dane & Schneider, 1998; Durlak, 1998), of school-based prevention programs is a primary criteria in establishing whether a program is ready for widespread dissemination (Flay et al., 2005). The most credible evidence of a program's effectiveness comes from randomized controlled trials. Randomization is used to ensure that factors known and unknown are equally distributed between experimental and control conditions, allowing changes in the outcome(s) of interest (e.g., prosocial behaviors, substance abuse reduction) to be attributed to the program under evaluation. In school-based prevention program evaluations, schools or classrooms are randomized to receive either (a) an experimental program designed to reduce (e.g., violence, substance abuse) or enhance (e.g., empathy, self-esteem) specific behaviors or (b) requested to continue with what is referred to as “business as usual”; essentially to continue current practices without initiating substantial program-related reforms.

It is specifically this “business as usual” that program evaluators have neglected as potential reasons why program-related outcomes may have been smaller than expected (Cordray & Jacobs, 2007; Susser, 1995; Tobler, Roona, Ochshorn, Marshall, Streke, & Stackpole, 1998). When programs are observed to have minimal and, in some cases, nonexistent, effects, this is often attributed to program failure (i.e., failure of theory). However, an often overlooked rationale behind minimal or null program findings is the extent to which control conditions may receive or self-initiate programming similar to that specified in the experimental condition (Dane & Schneider, 1998; Domitrovich & Greenberg, 2000). This is referred to as *program differentiation* (Dane & Schneider, 1998; Domitrovich & Greenberg, 2000) and *compensatory equalization of treatments* or *compensatory rivalry* by Cook and Campbell (1979) and remains one of the least studied aspects of implementation fidelity.

Schools or classrooms are likely to self-initiate a number of programs that assist students in developing positive behavioral attributes. In fact, recent estimates of



a nationally representative sample of middle schools indicated that the self-initiation of substance abuse prevention curricula was widespread, with almost half of the schools surveyed utilizing 3 or more programs (Ringwalt, Ennett, Vincus, Thorne, Rohrbach, & Simons-Rudolph, 2002). Self-initiated programs adopted by schools, unfortunately, are less likely to be evidence-based (Ringwalt et al., 2002), and therefore are presumed to have minimal influence on student behaviors (e.g., reduced substance abuse). However, given the plethora of prevention programs, many of which remain untested (Green, 1998), it is unlikely that programming of any kind would not have some potential benefit on, at minimum, a portion of the students exposed. Such occurrence is likely to abate the comparison with an experimental program (Cordray & Jacobs, 2007; Tobler et al., 1998). Further, the willingness of schools to participate in a randomized controlled trial suggests that the issues addressed in the trial (e.g., substance abuse, violent behaviors) are of considerable interest to those inclined to be involved and, hence, may indicate a school's desire to attend to these whether randomized to a control condition or not (Luepker, Perry, McKinlay, Nader, Parcel, Stone, Webber, Elder, Feldman, Johnson, & et al., 1996).

The addition of SACD programs and activities, those beyond the academically focused programming, are what many parents expect schools to initiate and are already woven into the everyday delivery of curricula content within the classroom and throughout the school (Elias, Zins, Graczyk, & Weissberg, 2003). Thus, it may be unrealistic to expect schools assigned to control conditions to be completely void of programming that may parallel concepts addressed in an intervention under investigation (Cook, 2001). Not surprisingly, the concepts touted by many intervention programs are found to commonly appear in the mission statements of schools and such programming is increasingly being encouraged by state departments of education and for schools who receive funds through Safe and Drug-Free Schools and Communities Act (U.S. Department of Education Safe and Drug-Free Schools Program, 1998). A critical aspect in understanding the effectiveness of a prevention program, therefore, lies in the documentation of the activities and/or strategies occurring in control conditions that may parallel and, subsequently, mask or weaken the effects of an intervention under evaluation.

## Methods

The data presented in this paper were collected during the last year of a four-year cluster-randomized control trial of a comprehensive prevention and social and character development (SACD) program, *Positive Action*, implemented under “real-world” conditions (i.e., an effectiveness trial, Flay, 1986) in 20 elementary schools on three islands in the Hawai'i school district. Schools were selected to receive the intervention ( $n = 10$ ) or continue “business as usual” ( $n = 10$ ) based on random assignment from matched-pairs. Pairs were determined by a multivariate indicator of risk developed from school-level data on the following characteristics: the proportion of students receiving free and reduced lunch, school size, pupil to teacher ratio, percent stability of student enrollment, ethnic distributions, and indicators of problem behaviors (e.g., suspensions) and achievement (e.g., standardized testing) (c.f., Dent et al., 1993).

Intervention. The *Positive Action* program ([www.positiveaction.net](http://www.positiveaction.net)) is a comprehensive school-based prevention and youth development program designed to target the enhancement of positive behaviors and behavioral attributes directed at the self (e.g., self-responsibility, positive self-concept) and social relationships (e.g., conflict resolution, respect, kindness), while simultaneously reducing students' negative behavioral problems (e.g., substance abuse, disciplinary referrals, suspensions), as well as improving school-related performance (e.g., attendance, academic achievement). The K-6 program focuses on the entire self, with 6 units consisting of approximately 140 15-minute, age-appropriate lessons per grade level taught over the whole academic year (Flay & Allred, 2003; Flay et al., 2001).

Study Design. During the spring of the final year of the effectiveness trial, one school leader (principal, vice principal, counselor) from each of the 20 study schools was invited to respond to an online survey regarding the SACD programs and/or activities that were conducted in their school during the prior 3 academic years.

Outcome Measures. Social and character development programs/activities were defined as those that (a) promote social and emotional competence (e.g., self-regulation, conflict resolution); (b) develop positive character traits (e.g., responsibility, honesty); (c) foster prosocial attitudes and behaviors (e.g., altruism);

and (d) address problem behaviors (e.g., violence, delinquency). All questions were adapted from the SACD-activities Principal and Teacher interview surveys design by the Social and Character Development Research Consortium (2004). Respondents (school leaders) were asked to list up to 16 SACD programs. For each program, respondents indicated the grade level or levels (all, lower K-2<sup>nd</sup>, upper 3<sup>rd</sup>-5<sup>th</sup>) that received it, the number of weeks the program was offered, the amount of time (minutes) devoted to the program per week, and whether or not teachers attended/received training to deliver the program (yes/no). Respondents also indicated how well their school conducted informal (without a curriculum) SACD activities and/or strategies (e.g., character education, violence prevention, citizenship, tolerance); items (17 total) were answered on a 0 “Do not do” to 4 “Outstanding” scale.

Analysis. Given the limited sample size, we report simple descriptive statistics (mean, standard deviation, frequencies) to illustrate the similarities and differences between the SACD programming activities conducted in the control vs. intervention schools. We provide intervention estimates for SACD programs without *Positive Action* and *Positive Action*, solely. The time estimates reflect the average number of hours per SACD program. These were calculated by taking the number of minutes (transformed into hours) devoted to the program multiplied by the number of weeks the program was conducted. These were averaged across all SACD programs for control and intervention schools, separately.

### Results

Responses to the questionnaire are presented in Table 1. Responses rates were 80% (8 out of 10) and 100% (10 out of 10) for control and intervention schools, respectively. Two intervention schools reported conducting no other SACD programs, apart from *Positive Action*. One intervention school failed to report the number of weeks and time devoted to *Positive Action*, so the estimates associated with the intervention program are based on the average of 9 schools. On average, control schools self-initiated 8 SACD programs, whereas intervention schools reported an average of 3.6 additional SACD programs (excluding *Positive Action*). A list containing the types of SACD programming occurring in both control and

experimental schools is presented in Table 2. The number of weeks the self-initiated SACD programs were performed during an academic year was greater for control schools ( $23.6 \pm 2.9$  wks/program) than intervention schools ( $16.6 \pm 1.5$  wks/program). The average amount of time devoted to a SACD program (not including *Positive Action*) during an academic year was  $32.9 \pm 27.8$  hrs/program for control schools vs.  $22.1 \pm 15.1$  hrs/program in intervention schools. Teacher training specific to delivering the SACD programs accompanied 44.6% and 36.1% of the control and intervention school programs (in addition to *Positive Action*), respectively. Control schools reported conducting SACD programs that did not have a formal curriculum at “good” (54.1%) and “excellent” (28.8%) levels, while intervention schools indicated they conducted the SACD programs “somewhat well” (33.6%) and “good” (47.1%). The majority of programs were for students at all grade levels (53.6% control and 38.9% intervention) or were programs designed specifically for students in upper grades (30.4% control and 25% intervention). Intervention schools spent approximately 34.4 wks ( $\pm 0.8$  wks) for 32.1 hrs ( $\pm 22.8$  hrs) per school year conducting the *Positive Action* program.

----- Insert Tables 1 and 2 here -----

## Discussion

Both control and intervention schools self-initiated a large number of SACD programs that promoted concepts and ideas comparable to the program under investigation (*Positive Action*). Of particular concern was the high rate of parallel programming occurring in the control schools (see Table 1). The findings highlight that “business as usual” is a condition that may consist of activities that are not solely separate from the intervention program. In keeping with prior school-based assessments of self-initiated programming (Ringwalt et al., 2002), almost all of the self-initiated programming was non-evidence-based, suggesting that schools, when seeking out programs to adopt, are not considering research findings to guide the selection. Additionally, prior program evaluations have indicated that schools assigned to control conditions slowly adopt intervention-like practices (Cook et al., 2000) and that elements of the intervention were carried out in control schools, albeit at a lesser frequency than observed in experimental schools (Harachi et al., 1999).

It is clear from these data that “business as usual” consists of approximations of experimental conditions that promote and nurture positive behavioral attributes of students. Researchers, therefore, must conscientiously attend to the fact that any programming that may prove beneficial to student outcomes, achievement or otherwise, cannot be withheld, and that control schools are unlikely to postpone initiating any kind of SCD programming simply to allow for an experimental test of one program’s effectiveness (Cook, 2001; Games et al., 2002). Thus, it appears that self-initiation of programming is rampant throughout schools, and therefore, must be accounted for in evaluation trials.

Experimental conditions imposed in the evaluation of school-based programs are, therefore, in need of accounting and correcting for what constitutes “business as usual”. Clearly, the presence of these programs is likely to reduce the differences between experimental and control conditions by raising the controls up to a level comparable to that of the intervention, where, this counterfactual condition begins to imitate the treatment condition more so than what was intended or expected (Cordray & Jacobs, 2007). Conversely, not all schools assigned to an experimental condition implement a program with equal fidelity (Aber et al., 2003; Harachi et al., 1999; Pentz et al., 1990; Rohrbach et al., 1993; Rohrbach et al., 2005; Smith et al., 1993; Taggart et al., 1990). The “under-implementation” of a program would lower the experimental condition to a level comparable to the absence of an intervention (Botvin, Griffin, & Nichols, 2006), the condition expected of the control. If both of these scenarios take place, a “squeezing” of the detectable effect would occur from both sides (from the top and bottom), leading to an underestimation of the overall effect of an intervention program. Given the upper limit to the feasible effectiveness of a well-implemented program is modest (Tobler et al., 1998), unaccounted for programming in control conditions, coupled by poorly implemented programming in the intervention conditions, may lead to program developers rejecting potentially effective programs (i.e., Type III error) (Basch et al., 1985; Scanlon et al., 1977).

Despite these findings, several limitations need to be addressed. The authors were unable to obtain information from two of the control schools, even after repeated attempts. This lack of response may be systemic of other issues surrounding the desire

to conduct the intervention or reflect the views of having to serve in a control capacity (several schools assigned to the control condition expressed dissatisfaction with their allocation status – perhaps accounting for the increased self-initiation of SADC programming). Additionally, this information was collected at the end of the trial and is subject to recall bias and the authors cannot account for changes in self-initiated SADC programming over the course of the trial. Further, only a single respondent from each school was asked to indicate current SADC programming. Even though SADC programming was defined in the survey, it is likely that people have varying concepts of what represents SADC programming and thus may have provided responses about more or less SADC programs than actually took place.

Ultimately, what transpired over the course of the four-year evaluation was not specifically a test of the hypothesis that the intervention would result in improved student outcomes in comparison to a no-treatment condition, but rather, the evaluation was a test of whether the provision of a coordinated, teacher trained, and ongoing technically supported SADC program would result in improved student outcomes when compared to alternative SADC programs that were primarily non-evidence based, were conducted with minimal formal training (less than half in the control schools, see Table 1), consisted of programming of varying length, and/or may not have been implemented well.

## Conclusions

Systematic evaluations of ongoing programs that approximate experimental conditions should be included in the evaluation of the fidelity of implementation (Cordray & Jacobs, 2007; Dane & Schneider, 1998; Domitrovich & Greenberg, 2000). This necessitates that prior to designing evaluation trials, researchers specify and delineate the “active ingredients” and activities that make an intervention unique (Chen, 1998; Cordray & Jacobs, 2007). Once outlined, this information could be linked to program-related outcomes using various statistical modeling approaches, such as a school-level covariate in multilevel analyses or, based on various levels of compliance, propensity score estimation, randomization as an instrumental variable, or estimating complier average causal effects (CACE) models (Foster, 2003; Gennetian, Bos, & Morris, 2002). Given the lack of attention program differentiation has received

to date, a great deal of inquiry is still required to define what approach or combination of approaches will best account for program differentiation.

#### Funding Disclosure

This project was funded by the National Institute on Drug Abuse, grant #R01-DA13474.

Table 12 (Table 1 Manuscript 3). Responses to self-initiated SCD activities questionnaire for control and intervention schools

Variables	Control ( <i>n</i> = 8)		Intervention – <i>Positive Action</i> ( <i>n</i> = 10)			
			SCD <sup>a</sup>		<i>Positive Action</i> <sup>b</sup>	
	M	SD	M	SD	M	SD
Average number of SCD programs per school <sup>c</sup>	8.0	±5.1	3.6	±3.6	1	
Average number of weeks per SCD program conducted	23.6	±2.9	16.6	±1.5	34.4 <sup>d</sup>	±0.8
Average time devoted per SCD programming (hours) <sup>e</sup>	32.9	±27.8	22.1	±15.1	32.1 <sup>d</sup>	±22.8
Teacher received training (percent “yes”)	44.6%		36.1%		100% <sup>f</sup>	
How well school conducted SCD programs (percentage) <sup>g</sup>						
Did not do	1.2%		2.5%			
Somewhat	13.5%		33.6%			
Good	54.1%		47.1%			
Excellent	28.8%		14.3%			
Outstanding	2.4%		2.5%			
Grade level SCD programs delivered						
All grades	53.6%		38.9%		100%	
Lower (K to 2 <sup>nd</sup> )	7.1%		5.6%			
Upper (3 <sup>rd</sup> to 5 <sup>th</sup> )	30.4%		25.0%			
Kindergarten	0.0%		0.0%			
First	0.0%		5.6%			
Second	0.0%		2.8%			
Third	0.0%		2.8%			
Fourth	1.8%		5.6%			
Fifth	7.1%		13.9%			

Abbreviation: SCD = social and character development

<sup>a</sup> Represents the average of self-initiated SCD programs, solely

<sup>b</sup> Represents the average of *Positive Action*, solely

<sup>c</sup> Range: Control 1 to 16 SCD programs; Intervention 1 to 10 programs including *Positive Action*

<sup>d</sup> Estimates based on reporting of *Positive Action* activities in 9 schools

<sup>e</sup> Average amount of time reported devoted to a SCD program

<sup>f</sup> All *Positive Action* schools participated in yearly training and booster sessions to implement the program

<sup>g</sup> Percentage of total responses. Responses represent how well SCD programs were conducted that did not have a formal curriculum. This implies intervention schools were reporting on self-initiated SCD, thus their responses do not reflect their perceived level of quality regarding the implementation of *Positive Action*



Table 13 (Table 2 Manuscript 3). Self-initiated SCD programs and activities reported by control and experimental schools

Program Name	Control	<i>Positive Action</i>
5 R's (Respect, Responsibility, Resourcefulness, Resiliency, Relationships)	X	X
Catch Them Being Good	X	X
Character Education	X	X
Character Education/Core Ethical Values	X	X
Character Word of the Week	X	X
Class rewards stores		X
Community Parades	X	
Complier Mediators	X	
D.A.R.E.	X	X
Drug Free Week	X	
E Ola Pono (Substance Abuse Awareness and Preventive Program)		X
Ethnic Dances	X	
Flag Raisers	X	
G.R.E.A.T.	X	
Getting Along with Others	X	
Gotcha rewards		X
Hawaiian Studies		X
Jump rope for Heart	X	
Junior Police Officers	X	X
Library Clubs	X	
M. Curriculum Project	X	
Parent Meetings	X	X
Peace Week	X	
Peer Mediation	X	X
Peer Tutorials	X	
Pennies for Patients	X	
Pono		X
Positive Action		X
Positive Behavior Support	X	X
Primary School Adjustment Program	X	X
Project Charley	X	
Parent-Teacher Student Ohana	X	
Respect campaign		X
School Counseling	X	X
School Level Supports	X	X
School Student of the Month	X	
Second Step		X
Service Learning Tech	X	
Tribes		X
Violence prevention videos	X	

## Chapter 5. Conclusions

An abundance of theory driven and socially valid prevention programs currently exist. These programs were developed by practitioners and prevention scientists with extensive background and knowledge about effective practices hypothesized to improve the behavior(s) of interest (e.g., reduced substance abuse, improved prosocial behaviors). Yet, despite these efforts, many programs fail to achieve their goals. One of the questions put forward to understand this incongruity is whether programs are, in fact, delivered as intended and, if not, what may account for the differences in program design and what ultimately transpires. Additionally, in the design phase of a program, a key issue is whether the audience for whom the program was intended will believe the program to be of utility, with this utility leading to program utilization followed by beneficial outcomes. The manuscripts presented addressed these questions by examining the implementation fidelity of a school-based social and character development program, *Positive Action*.

Implementation fidelity, in simplest terms, refers to the degree to which a planned program was carried out to its original intent. When a program, found to be beneficial during an efficacy trial, is evaluated in a naturalistic setting (i.e., effectiveness trial), many threats are present that are likely to “derail” the attainment of comparable outcomes. Specifically, it is issues related to implementation that practitioners can examine that may shed light on unanticipated, and many times unwelcome (e.g., no program effect), results from real-world evaluations.

Reporting of implementation is conspicuously absent in intervention evaluation studies (Mayo-Wilson, 2007) and even fewer studies specifically linking implementation (in quantifiable terms) to program-related outcomes (Dane & Schneider, 1998; Domitrovich & Greenberg, 2000; Durlak, 1998). In a recent review on the status of reporting implementation in randomized controlled trials, it was found that across a number of meta-analyses, few, if any, studies operationalized, verified, and/or linked program implementation status to program-related outcomes (Mayo-Wilson, 2007). Guidelines for reporting non-randomized controlled trials (TREND) (Bellg, Borrelli, Resnick, Hecht, Minicucci, Ory, Ogedegbe, Orwig, Ernst, & Czajkowski, 2004) and, recent calls for increasing the reporting of implementation in

randomized controlled trials (CONSORT) (Mayo-Wilson, 2007; Moher, Schulz, & Altman, 2001), place implementation as one of the key pieces of information that is essential for the successful replication and dissemination of potentially beneficial prevention programs. Additionally, this information can be used to refine and modify existing interventions, with hopes of eventually arriving at an intervention of sufficient quality that is effective within and across multiple settings.

In the prevention science literature, implementation is generally regarded as consisting of five components addressing *how much* (i.e., exposure) of *what* (i.e., adherence) was delivered in *what ways* (i.e., quality of delivery) and *how it was received* (i.e., responsiveness) and *in comparison* to what (i.e., program differentiation) (Dane & Schneider, 1998; Domitrovich & Greenberg, 2000). While each embodies an essential aspect of what transpired during an effectiveness trial, no single component exists in a vacuum (see Figure 6). That is to say, each component exists only in relation to the level of each of the other components and it is the amount of each that can have a substantial impact on the others. Adding an additional layer of complexity is the notion that the level of one component does not always predict or interact with the level of the others in foreseeable ways.

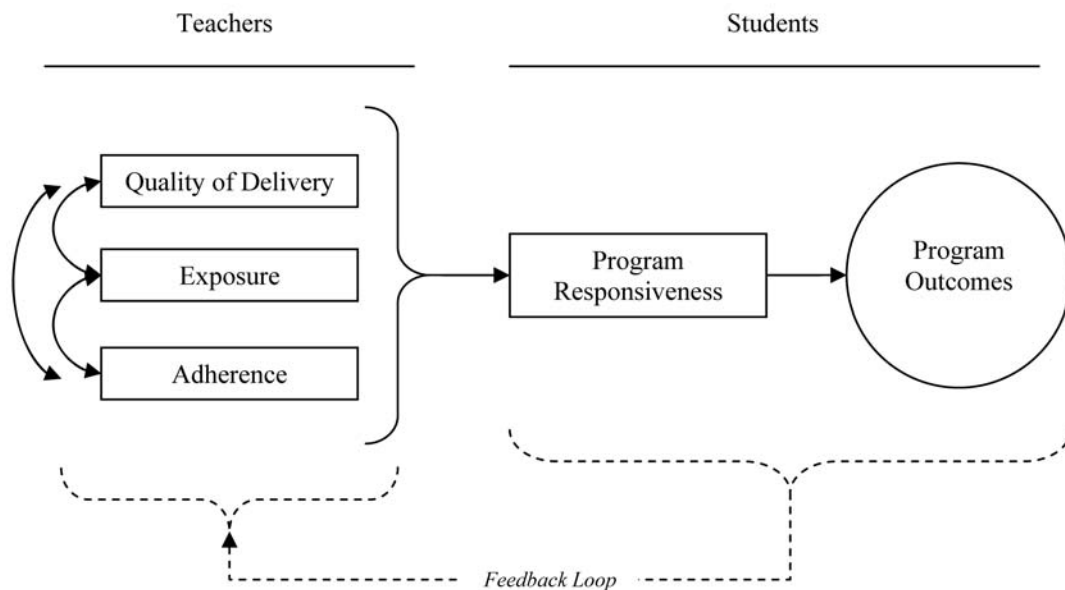


Figure 6. Teacher-student-outcomes program fidelity model of proposed effects

Using the conceptual model depicted in Figure 6, a number of scenarios can be generated. For instance, one may deliver a program at a high frequency, yet if it is not delivered well (i.e., high levels of exposure coupled with poor quality of delivery), it is likely the audience will view the program as unimportant or frivolous (i.e., low responsiveness), leading to a minimal impact on program outcomes. Conversely, an audience may show considerable “liking” of a program (i.e., high responsiveness), yet the program may have been noticeably modified (i.e., lack of adherence), accounting for the high level receptivity. A plausible scenario in this case would be the cultural adaptation of program materials. Such changes may constitute considerable program modification (i.e., low adherence), yet these changes may ultimately lead to higher levels of responsiveness and, subsequently, greater program impact. In the absence of this information, evaluators would be unable to replicate the findings in other settings. Additionally, a program may be delivered at a high rate of exposure, adherence, and quality, but again, if it is not viewed favorably by the audience (e.g., age or developmentally inappropriate), anticipated changes are unlikely to occur.

These examples are not exhaustive, and any number of scenarios can and do take place throughout effectiveness trials. Hence, it is with the documentation of the implementation components that evaluators can fully understand what transpired, and with that, what elements of a program are in need of modification, continuation, or elimination, and what circumstances are the most favorable for the program to work effectively. Implementation measures, therefore, assist in the proper interpretation of program outcomes, both those anticipated (positive) and unwanted (negative) (Durlak, 1998).

In reference to the misinterpretation of evaluation trials, implementation measures are essential in minimizing the possibility of committing a Type III error – incorrectly concluding a program to be ineffective when, in fact, it was not delivered as intended (Basch et al., 1985). If a program was designed to be delivered 3 times a week for 20 minutes each session, with each session including a story and student-teacher interactions, but if these were done less (or more) or with modifications (no stories or student-teacher interactions), the “real” program was never delivered, and thus, the outcomes (helpful or harmful) cannot be attributed to the program.

When information on the aspects of implementation are collected, there exists an opportunity to link these to program outcomes. Such analyses can help inform evaluators on how changes may have affected or are related to program performance. In contrast with the sometimes unexpected relationships observed among the components of implementation, the factors found to influence implementation are robust (Basch, 1984; Basch et al., 1985; Dane & Schneider, 1998; Domitrovich & Greenberg, 2000; Durlak, 1998; Dusenbury et al., 2003; Gager & Elias, 1997; Games et al., 2002; Han & Weiss, 2005). The most studied of these include supportive administrators, attachment to workplace, and personal attitudes and beliefs regarding a program. The relationships among these factors to the components of implementation are revisited in Figure 7.

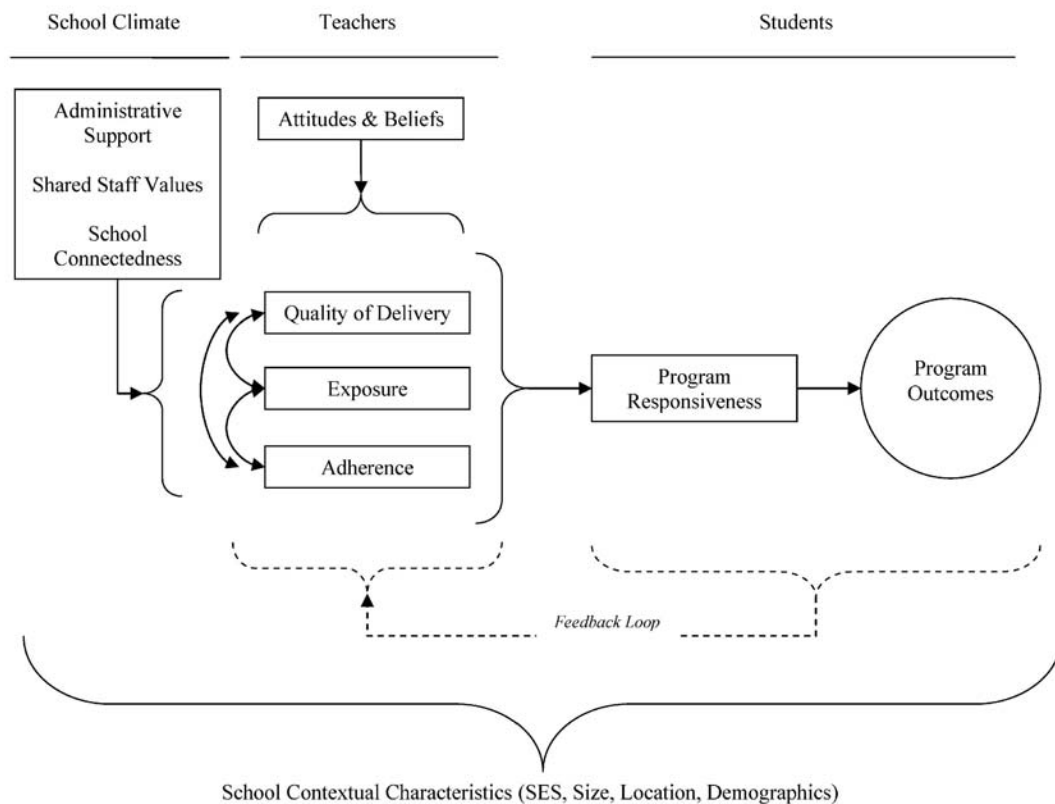


Figure 7. Conceptual model of teacher, administrative, and contextual characteristics influencing implementation fidelity

It can be seen from this model that school climate factors play a direct role in influencing *how much* (i.e., exposure) of *what* (i.e., adherence) was *how it was delivered* (i.e., quality of delivery). However, a limitation in this model is the non-specified relationship among school characteristics and those originating from the teacher. An expanded, and perhaps more realistic representation, of these relationships is presented in Figure 8. In this model, the direct influence of school characteristics on teacher characteristics is shown, with this influence only having an effect on implementation as it is mediated via the teacher.

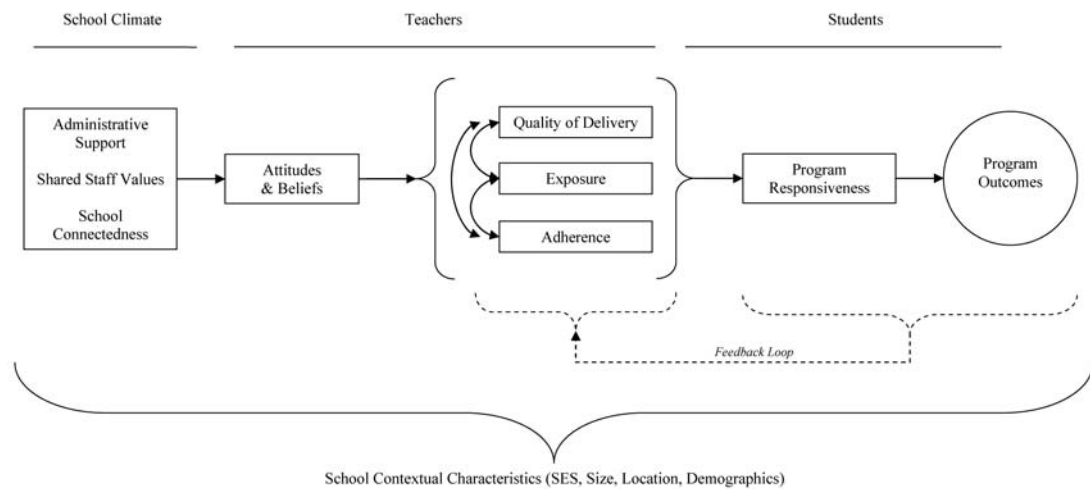


Figure 8. Expanded conceptual model of teacher, administrative, and contextual characteristics influencing implementation fidelity

Referring to the conceptual model presented in Figure 8, the manuscripts addressed the various dimensions of implementation by placing them in a causal framework which specified:

In order for teachers to deliver a program with a high level of fidelity (i.e., dosage, quality of delivery, adherence), the school climate must be supportive and have a shared vision in the beliefs regarding the school's responsibility to teacher social and character development concepts.

As a consequence of this (albeit not tested in the second manuscript, but hypothesized to have occurred) teacher implementation fidelity leads to student responsiveness which, in turn, is related to improved program-related outcomes (e.g., increased prosocial behaviors, reduced substance abuse).

Further, while not depicted in the conceptual model, program differentiation was found to be compromised (i.e., control schools adopted considerable amounts of parallel programming) which is likely to minimize the detectable effect of a well delivered (i.e., high dosage, adherence and quality of delivery) and received (i.e., students enjoy/like/are satisfied) SACD program.

Consistent with prior research and theory, the manuscripts demonstrate that multiple factors play essential roles in whether a program realizes its goals.

### Limitations

The findings presented need to be prefaced in the context of several methodological limitations. For the teacher and school climate study (manuscript 1), the hypotheses tested were cross-sectional. Unfortunately, this precludes drawing any conclusions on potential changes over time, where one would expect the intervention to have a beneficial effect on teacher attitudes towards *Positive Action*, leading to increased implementation. In fact, *Positive Action* is a comprehensive program with a school climate component; thus, one would also expect teachers' attitudes toward it to become more favorable over time as well. Additionally, a longitudinal analysis, and hence an assessment of whether attitudes, beliefs, and the school climate were changing over time, is perhaps a more appropriate test of these relationships. However, linking teachers across years was precluded by the design. Additionally, the processes modeled may be more appropriately considered in a multilevel context, with school climate serving as a level 2 predictor, which was limited in the study due to power issues with the number of schools in the sample ( $N = 10$ ) (Raudenbush & Liu, 2000). Conducting such an analysis with the data also would raise questions surrounding the appropriateness of aggregating individual perceptions of peers and



administration to a higher order (i.e., level 2). Future studies should consider collecting objective information on school climate with a larger number of schools to more accurately model these associations.

In relation to student responsiveness (manuscript 2), one limitation is that all the measures were self-reported by the students. This may increase the correlation among the constructs due to common method variance, thereby accounting for the relationship observed among the variables. Additionally, while the findings illustrated the link between having positive views about a prevention program and program-related outcomes, the analysis could not account for a considerable portion of the variance in student responsiveness ratings. This suggests other factors, not measured in the study, were influencing student perceptions about the program. For an answer to this, future studies could utilize focus group methodology, with these composed of known groups – students who are satisfied and unsatisfied with the program. This approach might lead to greater understanding of why program components, such as lesson content and the use of token rewards, are viewed as appropriate or inappropriate. Further, the students and teachers assessed in the reported studies were from a single state (Hawai'i) and therefore the findings may not generalize to the entire population of elementary students and teachers.

Finally, manuscript 3 (program differentiation) was limited by the timeframe for recall (i.e., the prior 3 academic years) and the lack of response from two of the control schools. The latter is of importance because it suggests that, for reasons unknown, the two non-responding schools were unwilling to answer the questions concerning alternative programming, even after repeated attempts. This may be due to their lack of comfort with being randomized to serve as a control condition, thereby creating animosity towards the project and possibly the program.

An overall limitation for all three studies is the use of self-report of the implementation components examined. While attempts to increase honest reports were made, it came at the price of not being able to link teachers across years to examine these changes longitudinally.

## Public Health Implications

The importance of these studies for public health practitioners resides not only in the ability to design effective interventions, but also to ensure that when a program is designed and tested, that this is done in a context where “fair” and “appropriate” judgments can be made. With implementation information linked to program outcomes, the “worth” of a program (that is, is it ready for widespread dissemination or are additional modifications required) can be based on what was actually delivered (e.g., adherence, dosage), with this compared to a known group consisting of either alternative programming or “business as usual” (i.e., program differentiation).

Without implementation information practitioners may mistakenly conclude that a beneficial program is not performing to initial expectations, when in fact it was not delivered as intended. Discarding such a program because of faulty implementation would result in failure to capitalize on the health gains that would have been realized from its widespread dissemination. Conversely, a program that is originally ineffective may be delivered with considerable modification and found to be effective, yet without these changes documented (i.e., information on implementation) future adopters would be unaware that modifications are required in order to realize beneficial outcomes. In the absence of this information, practitioners would wrongly conclude that the ineffective program is effective as designed and call for wide-spread dissemination. Only when future adopters implement the program and fail to realize the “claimed” results, would questions arise about why beneficial results were found initially, but not in successive trials. This final scenario is of concern, since considerable cost is associated with developing materials, training practitioners, and disseminating programs across the nation and the world. Only through added costs and painstaking efforts from additional trials are the flaws in the original program identified and rectified.

These studies indicate that a school’s decision to adopt a program is not as simple as selecting a program and carrying through with its implementation. This decision has extensive implications for all involved, administrators, teachers and students alike. At the initial stages of program selection, all parties should be informed of the benefits (e.g., improved student outcomes, school climate) and barriers (e.g.,

time, resources) associated with a given program. When schools, and the teachers within, decide to adopt a program, consensus needs to be garnered to ensure that the costs associated with implementing the program (e.g., materials, training, ongoing support) are not wasted simply because the climate was non-supportive of this new innovation (Rogers, 1995, 2002). This is referred to as *community readiness* (Chilenski, Greenberg, & Feinberg, 2007; Oetting, Donnermeyer, Plested, Edwards, Kelly, & Beauvais, 1995; Slater, Edwards, Plested, Thurman, Kelly, Comello, & Keefe, 2005), and is an emerging construct in need of scrutiny during the pre-implementation phase of school-based prevention programming. From this assessment, an understanding of whether the school and personnel are supportive of the proposed changes, those associated with a given program, can be developed in order to avoid poor implementation or identify schools where additional resources are required to ensure high levels of fidelity.

Recommendations of effective evidence-based programs are founded upon credible information from high quality scientific studies (Flay et al., 2005). Information from these studies provides future program adopters with an understanding that, when they implement a program, the declared results should be obtained. This is a critical link in the diffusion process, whereby science informs practice on what “works” and what efforts should not be pursued. If prevention scientists fail to adequately account for how a program was delivered during the efficacy and effectiveness trials, they risk misinforming the public about where resources should be allocated. When such events occur, not only is a culture developed where science loses credibility, but also where the public, when asked to adopt a truly beneficial program, becomes skeptical and slowly, if at all, adopts the program. This is likely to result in marginal gains in health. Thus, information on the fidelity of implementation is, for prevention scientists, a “check and balance” method of ensuring that when recommendations are made, they are done so with a greater assurance of success.

## Future Directions

While the findings reported in the three manuscripts answer important *post hoc* questions regarding fidelity, they also stimulate new questions as to what can be done to ensure fidelity in future implementation of the program. In relation to school climate and teacher beliefs and attitudes, the question presented is whether these can be altered prior to implementation (pre-program implementation) to create a staging ground whereby a program will be accepted, viewed favorably, and carried out with the utmost fidelity? To answer this, studies need to be designed with teachers and administrators surveyed before implementation occurs, with those expressing negative views of the program targeted with additional program materials, evidence of beneficial effects, and technical support (i.e., site coordination).

For student responsiveness (i.e., customer satisfaction), studies need to be designed that attempt to understand why some students liked the program, while others did not. Are certain characteristics of the students indicative of whether the program is viewed favorably? Is responsiveness dictated by characteristics of the program itself? The use of token rewards and recognition may be attractive to some students, yet unappealing to others – should these be modified to be more developmentally, culturally, and generationally appropriate? The same critique could be given to the lesson and unit plans.

With program responsiveness, the study indicates that no longer can prevention researchers simply collect pre- and post-measures on control students, assuming they were unexposed to programming of similar content. That is, greater attention needs to be paid to what is specifically going on in the units serving as controls (e.g., schools, classroom), with detailed documentation, comparable to the attention given to the intervention condition, in the type and amount of self-initiated programming.

While more questions than answers have grown out of these studies, they do point towards several essential themes in program implementation. First, without a conducive school climate characterized by supportive administration and shared staff values, implementing a program with fidelity is less likely to occur. The beliefs and attitudes teachers hold towards a program, in turn, are modified by exposure to a school's climate, indicating that the climate dictates and shapes the individuals and activities that are embedded within it. Secondly, although intuitive, designing

programs that audiences' like is an essential part of producing the desired effects. Interestingly, this is often overlooked in the prevention science literature. This basic notion grows out of the marketing discipline, whereby substantial resources are devoted to *why* individuals initially choose and continue use of various products. Perhaps greater insight into responsiveness will come from adopting a marketing-based approach to understand the "purchasing" (adoption) behaviors of schools and students, and using this information to make necessary changes in program content and delivery. And finally, without accounting for what's occurring within the control condition, a beneficial program may be found to be "ineffective", merely because of a increase in comparable programming by controls.

In conclusion, greater emphasis on the assessment and the experimental manipulation of these factors in future studies will assist in refining a prevention program for widespread adoption and, ultimately, informing the public on what programs are most likely to result in the greater health gains.

## Bibliography

- Aber, J. L., Brown, J. L., & Jones, S. M. (2003). Developmental trajectories toward violence in middle childhood: Course, demographic differences, and response to school-based intervention. *Developmental Psychology*, 39(2), 324-348.
- Bandalos, D. L., & Finney, S. J. (2001). Item parceling issues in structural equation modeling. In G. A. Marcoulides (Ed.), *New Developments and Techniques in Structural Equation Modeling* (pp. 269-296). Mahwah, NJ: Lawrence Erlbaum Associates.
- Basch, C. E. (1984). Research on disseminating and implementing health education programs in schools. *Journal of School Health*, 54(6), 57-66.
- Basch, C. E., Sliepecevic, E. M., Gold, R. S., Duncan, D. F., & Kolbe, L. J. (1985). Avoiding Type III Errors in health education program evaluations: A case study. *Health Education Quarterly*, 12(4), 315-331.
- Battistich, V., Schaps, E., & Wilson, N. (2004). Effects of an Elementary school intervention on students' "connectedness" to school and social adjustment during Middle school. *The Journal of Primary Prevention*, 24(3), 243-262.
- Bellg, A. J., Borrelli, B., Resnick, B., Hecht, J., Minicucci, D. S., Ory, M., Ogedegbe, G., Orwig, D., Ernst, D., & Czajkowski, S. (2004). Enhancing treatment fidelity in health behavior change studies: best practices and recommendations from the NIH Behavior Change Consortium. *Health Psychol*, 23(5), 443-451.
- Botvin, G. J. (2004). Advancing prevention science and practice: Challenges, critical issues, and future directions. *Prevention Science*, 5(1), 69-72.
- Botvin, G. J., Griffin, K. W., & Nichols, T. D. (2006). Preventing youth violence and delinquency through a universal school-based prevention approach. *Prev Sci*, 7(4), 403-408.
- Calvert, S. C., & Johnston, C. (1990). Acceptability of treatments for child behavior problems: Issues and implications for future research. *Journal of Clinical Child Psychology*, 19(1), 61-74.
- Catalano, R. F., & Hawkins, J. D. (1996). The Social Development Model: A theory of antisocial behavior. In J. D. Hawkins (Ed.), *Delinquency and crime: Current theories* (pp. 149-197). New York, NY: Press Syndicate of the University of Cambridge.
- Chen, H.-t. (1998). Theory-driven evaluations. In A. J. Reynolds & H. J. Walberg (Eds.), *Advances in educational productivity: Evaluation research for educational productivity* (Vol. 7). US: Elsevier Science/JAI Press.

- Chilenski, S. M., Greenberg, M. T., & Feinberg, M. E. (2007). Community readiness as a multidimensional construct. *Journal of Community Psychology*, 35(3), 347-365.
- Cohen, P., Cohen, J., Aiken, L. S., & West, S. G. (1999). The problem of units and the circumstance for POMP. *Multivariate Behavioral Research*, 34(3), 315-346.
- Connell, D. B., Turner, R. R., & Mason, E. F. (1985). Summary of findings of the School Health Education Evaluation: Health promotion effectiveness, implementation, and costs. *Journal of School Health*, 55(8), 316-321.
- Cook, T., & Campbell, D. (1979). *Quasi-experimentation : Design & analysis issues for field settings*: Rand McNally College.
- Cook, T. D. (2001). Sciencephobia: Why education researchers reject randomized experiments. *Education Next*, 1(3), 62-68.
- Cook, T. D., Murphy, R. F., & Hunt, H. D. (2000). Comer's School Development Program in Chicago: a theory-based evaluation.
- Cordray, D. S., & Jacobs, N. (2007). Identifying core-components of prevention programs: Conceptual and methodological improvements are needed. *unpublished manuscript*.
- Dane, A. V., & Schneider, B. H. (1998). Program integrity in primary and early secondary prevention: Are implementation effects out of control? *Clinical Psychology Review*, 18, 23-45.
- Dent, C. W., Sussman, S., & Flay, B. R. (1993). The use of archival data to select and assign schools in a drug prevention trial. *Evaluation Review*, 17(2), 159-181.
- Domitrovich, C. E., & Greenberg, M. T. (2000). The study of implementation: Current findings from effective programs that prevent mental disorders in school-aged children. *Journal of Educational & Psychological Consultation*, 11(2), 193-221.
- Durlak, J. A. (1998). Why program implementation is important. *Journal of Prevention and Intervention in the Community*, 17(2), 5-18.
- Dusenbury, L., Brannigan, R., Falco, M., & Hansen, W. B. (2003). A review of research on fidelity of implementation: Implications for drug abuse prevention in school settings *Health Education Research*, 18(2), 237-256.
- Elias, M. J., Zins, J. E., Graczyk, P. A., & Weissberg, R. P. (2003). Implementation, sustainability, and scaling up of social-emotional and academic innovations in public schools. *School Psychology Review*, 32(3), 303-319.

- Elliott, D. S., & Mihalic, S. (2004). Issues in disseminating and replicating effective prevention programs. *Prevention Science*, 5(1), 47-53.
- Flay, B. R. (1986). Efficacy and effectiveness trials (and other phases of research) in the development of health promotion programs. *Preventive Medicine*, 15, 451-474.
- Flay, B. R., & Allred, C. G. (2003). Long-term effects of the Positive Action program. *American Journal of Health Behavior*, 27(S1), S6-S21.
- Flay, B. R., Allred, C. G., & Ordway, N. (2001). Effects of the Positive Action program on academic achievement and discipline: Two matched-control comparisons. *Prevention Science*, 2(2), 71-89.
- Flay, B. R., Biglan, A., Boruch, R. F., Castro, F. G., Gottfredson, D., Kellam, S., Mocicki, E. K., Schinke, S., Valentine, J. C., & Ji, P. (2005). Standards of evidence: Criteria for efficacy, effectiveness and dissemination. *Prevention Science*, 6(3), 151-175.
- Flay, B. R., & Collins, L. M. (2005). Historical review of school-based randomized trials for evaluating problem behavior prevention programs. *Annals of the American Academy of Political and Social Science*, 599, 115-146.
- Flay, B. R., & Petraitis, J. (1994). The theory of triadic influence: A new theory of health behavior with implications for preventive interventions. In G. S. Albrecht (Ed.), *Advances in Medical Sociology, Vol IV: A Reconsideration of models of health behavior change* (pp. 19-44). Greenwich, CN: JAI Press.
- Foster, E. M. (2003). Propensity score matching: An illustrative analysis of dose response. *Medical Care*, 41, 1183-1192.
- Gager, P. J., & Elias, M. J. (1997). Implementing prevention programs in high-risk environments: Application of the resiliency paradigm. *American Journal of Orthopsychiatry*, 67(3), 363-373.
- Games, B., Millsap, M. A., & Goodson, B. (2002). When implementation threatens impact: Challenging lessons from evaluating educational programs. *Peabody Journal of Education*, 77(4), 146-166.
- Gennetian, L. A., Bos, J. M., & Morris, P. A. (2002). *Using instrumental variables analysis to learn more from social policy experiments*: Manpower Demonstration Research Corporation.
- Gittelsohn, J., Merkle, S., Story, M., Stone, E. J., Steckler, A., Noel, J., Davis, S., Martin, C. J., & Ethelbah, B. (2003). School climate and implementation of the Pathways study. *Preventive Medicine*, 37(6 pt 2), S97-S106.



- Glasgow, R. E., Vogt, T. M., & Boles, S. M. (1999). Evaluating the public health impact of health promotion interventions: The RE-AIM framework. *American Journal of Public Health, 89*(9), 1322-1327.
- Green, M. B. (1998). Most youth violence prevention programs remain untested. Retrieved May 5th, 2007, from <http://www.eurekalert.org>
- Hahn, E. J., Noland, M. P., Rayens, M. K., & Christie, D. M. (2002). Efficacy of training and fidelity of implementation of the life skills training program *Journal of School Health, 72*(7), 282-287.
- Han, S. S., & Weiss, B. (2005). Sustainability of teacher implementation of school-based mental health programs. *Journal of Abnormal Child Psychology, 33*(6), 665-679.
- Hansen, W. B., Graham, J. W., Wolkenstein, B. H., & Rohrbach, L. A. (1991). Program integrity as a moderator of prevention program effectiveness: Results for fifth-grade students in the Adolescent Alcohol Prevention Trial. *Journal of Studies on Alcohol, 52*(6), 568-579.
- Harachi, T. W., Abbott, R. D., Catalano, R. F., Haggerty, K. P., & Fleming, C. B. (1999). Opening the black box: Using process evaluation measures to assess implementation and theory building. *American Journal of Community Psychology, 27*(5), 711-731.
- Heskett, J. L., W. E. Sasser, J., & Schlesinger, L. A. (1997). *The Service Profit Chain: How Leading Companies Link Profit and Growth to Loyalty, Satisfaction, and Value*. New York, NY: The Free Press, a Division of Simon & Schuster, Inc.
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indices in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*, 1-55.
- Kam, C.-M., Greenberg, M. T., & Walls, C. T. (2003). Examining the role of implementation quality in school-based prevention using the PATHS curriculum. *Prevention Science, 4*(1), 55-63.
- Kealey, K. A., Peterson, A. V. J., Gaul, M. A., & Dinh, K. T. (2000). Teacher training as a behavior change process: Principles and results from a longitudinal study. *Health Education and Behavior, 27*, 64-81.
- Kennedy, C. H. (2002). The maintenance of behavior change as an indicator of social validity. *Behavior Modification, 26*(5), 594-604.
- Kirigin, K. A., Braukmann, C. J., Atwater, J. D., & Wolf, M. M. (1982). An evaluation of Teaching-Family (Achievement Place) group homes for juvenile offenders. *Journal of Applied Behavior Analysis, 15*(1), 1-16.

- Kline, R. B. (2005). *Principles and practice of structural equation modeling* (2nd ed.). New York: Guilford Press.
- Lillehoj, C. J., Griffin, K. W., & Spoth, R. (2004). Program provider and observer ratings of school-based preventive intervention implementation: Agreement and relation to youth outcomes. *Health Education and Behavior, 31*(2), 242-257.
- Luepker, R. V., Perry, C. L., McKinlay, S. M., Nader, P. R., Parcel, G. S., Stone, E. J., Webber, L. S., Elder, J. P., Feldman, H. A., Johnson, C. C., & et al. (1996). Outcomes of a field trial to improve children's dietary patterns and physical activity. The Child and Adolescent Trial for Cardiovascular Health. CATCH collaborative group. *Jama, 275*(10), 768-776.
- Marsh, H. W., Hau, K.-T., Balla, J. R., & Grayson, D. (1998). Is more ever too much? The number of indicators per factors in confirmatory factor analysis. *Multivariate Behavioral Research, 33*(2), 181-220.
- Mayo-Wilson, E. (2007). Reporting Implementation in Randomized Trials: Proposed Additions to the Consolidated Standards of Reporting Trials Statement. *Am J Public Health.*
- McCormick, L. K., Steckler, A. B., & McLeroy, K. R. (1995). Diffusion of innovations in schools: a study of adoption and implementation of school-based tobacco prevention curricula. *American Journal of Health Promotion, 9*(3), 210-219.
- Moher, D., Schulz, K. F., & Altman, D. G. (2001). The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomized trials. *Ann Intern Med, 134*(8), 657-662.
- Oetting, E. R., Donnermeyer, J. F., Plested, B. A., Edwards, R. W., Kelly, K., & Beauvais, F. (1995). Assessing community readiness for prevention. *Int J Addict, 30*(6), 659-683.
- Payne, A. A., Gottfredson, D. C., & Gottfredson, G. D. (2006). School predictors of the intensity of implementation of school-based prevention programs: Results from a national study. *Prevention Science, 7*(2), 225-237.
- Pentz, M. A. (2004). Form follows function: Designs for prevention effectiveness and diffusion research. *Prevention Science, 5*(1), 23-29.
- Pentz, M. A., Trebow, E. A., Hansen, W. B., MacKinnon, D. P., Dwyer, J. H., Johnson, C. A., Flay, B. R., Daniels, S., & Cormack, C. (1990). Effects of implementation on adolescent drug use behavior: The Midwestern Prevention Project (MPP). *Evaluation Review, 14*(3), 264-289.

- Petratis, J., Flay, B. R., Miller, T. Q., Torpy, E. J., & Greiner, B. (1998). Illicit substance use among adolescents: a matrix of prospective predictors. *Subst Use Misuse*, 33(13), 2561-2604.
- Purkey, W. W. (1970). *Self-concept and school achievement*. Englewood Cliffs, NJ: Prentice-Hall.
- Raudenbush, S. W., & Liu, X. (2000). Statistical power and optimal design for multisite randomized trials. *Psychological Methods*, 5(2), 199-213.
- Reimers, T. M., & Wacker, D. P. (1992). Acceptability of behavioral treatments for children: Analog and naturalistic evaluation by parents. *School Psychology Review*, 21(4).
- Reimers, T. M., Wacker, D. P., & Koepl, G. (1987). Acceptability of behavioral interventions: A review of the literature. *School Psychology Review*, 16(2), 212-227.
- Ringwalt, C. L., Ennett, S., Vincus, A., Thorne, J., Rohrbach, L. A., & Simons-Rudolph, A. (2002). The prevalence of effective substance use prevention curricula in U.S. middle schools. *Prev Sci*, 3(4), 257-265.
- Rogers, E. M. (1995). *Diffusion of Innovations* (4th ed.). New York: Free Press.
- Rogers, E. M. (2002). Diffusion of preventive innovations. *Addictive Behaviors*, 27, 989-993.
- Rohrbach, L. A., D'Onofrio, C. N., Backer, T. E., & Montgomery, S. B. (1996). Diffusion of school-based substance abuse prevention programs. *American Behavioral Scientist*, 39(7), 919-934.
- Rohrbach, L. A., Graham, J. W., & Hansen, W. B. (1993). Diffusion of a school-based substance abuse prevention program: Predictors of program implementation. *Preventive Medicine*, 22, 237-260.
- Rohrbach, L. A., Ringwalt, C. L., Ennett, S. T., & Vincus, A. A. (2005). Factors associated with adoption of evidence-based substance use prevention curricula in US school districts. *Health Education Research*, 20(5), 514-526.
- Scanlon, J. W., Horst, P., Nay, J. N., Schmidt, R. E., & Waller, A. E. (1977). Evaluability assessment: Avoiding Type III and IV Error in a field experiment. In G. R. Gilbert & P. J. Conklin (Eds.), *Evaluation Management: A Source Book of Readings*. Charlottesville, VA: US Civil Service Commission.
- Schwartz, I. S., & Baer, D. M. (1991). Social validity assessments: Is current practice state of the art? *Journal of Applied Behavior Analysis*, 24(2), 189-204.

- Sheldon, S. B. (2005). Testing a structural equation model of partnership program implementation and parent involvement. *The Elementary School Journal*, 106(2), 171-187.
- Sidman, M. (1960). *Tactics of Scientific Research*. New York: Basic Books.
- Simons, R. L., Conger, R. D., & Whitbeck, L. B. (1988). A multistage social learning model of the influences of family and peers upon adolescent substance abuse. *Journal of Drug Issues*, 18, 293-315.
- Slater, M. D., Edwards, R. W., Plested, B. A., Thurman, P. J., Kelly, K. J., Comello, M. L., & Keefe, T. J. (2005). Using community readiness key informant assessments in a randomized group prevention trial: impact of a participatory community-media intervention. *J Community Health*, 30(1), 39-53.
- Slavin, R. E., & Fashola, O. S. (1998). *Show me the evidence: Proven and promising programs for America's schools*. Thousand Oaks, CA: Corwin.
- Smith, D. W., McCormack, W. P., Steckler, A., & McLeroy, K. R. (1993). Teachers' use of health curricula: Implementation of Growing Healthy, Project SMART, and the Teenage Health Teaching Modules. *Journal of School Health*, 63(8), 349-354.
- Smith, D. W., Redican, K. J., & Olsen, L. K. (1992). The longevity of Growing Healthy: An analysis of the eight original sites implementing the School Health Curriculum Project. *Journal of School Health*, 62(3), 83-87.
- Social and Character Development Research Consortium. (2004). *Social and Character Development Research Program Evaluation Instrument*. Washington, DC: Institute of Education Sciences, U.S. Department of Education.
- Solomon, D., Battistich, V., Watson, M., Schaps, E., & Lewis, C. (2000). A six-district study of educational change: Direct and mediated effects of the Child Development Project. *Social Psychology of Education*, 4, 3-51.
- State of Hawaii Department of Education Systems of Accountability. (2006). School Accountability: School Status & Improvement Report. Retrieved 11/28, 2006, from <http://arch.k12.hi.us/school/ssir/ssir.html#>
- Susser, M. (1995). The tribulations of trials--intervention in communities. *Am J Public Health*, 85(2), 156-158.
- Taggart, V. S., Bush, P. J., Zuckerman, A. E., & Theiss, P. K. (1990). A process evaluation of the District of Columbia "Know Your Body" project. *Journal of School Health*, 60(2), 60-66.

- Tobler, N. S., Roona, M. R., Ochshorn, P., Marshall, D. G., Streke, A. V., & Stackpole, K. M. (1998). School-Based Adolescent Drug Prevention Programs: 1998 Meta-Analysis *Journal of Primary Prevention*, 20(4), 275-336.
- Tolan, P. H., Hanish, L. D., McKay, M. M., & Dickey, M. H. (2002). Evaluating process in child and family interventions: aggression prevention as an example. *J Fam Psychol*, 16(2), 220-236.
- U.S. Department of Education Safe and Drug-Free Schools Program. (1998). *Notice of final principles of effectiveness*: Federal Register, 63, 29901-29906.