Is More Better? Outcome and Dose of a Universal Drug Prevention Effectiveness Trial

Laura Ferrer-Wreder · Hans Saint-Eloi Cadely ·
Celene E. Domitrovich · Meg L. Small ·
Linda L. Caldwell · Michael J. Cleveland

Published online: 30 October 2010
© Springer Science+Business Media, LLC 2010

Abstract Two evidence-based interventions, Life Skills Training and TimeWise, were combined in an effectiveness trial. Participants were predominately African American youth ($N = 715$; $M_{\text{age}} = 12$). The study authors provide an empirical demonstration of the implications of incorporating dosage information in intervention outcome analyses. Study results showed no program-related benefits for drug use.

Results indicated intervention-related benefits for assertiveness and anxiety management skills and drug use intentions as well as a reduction in detrimental leisure motivations. High program exposure and lesson coverage tended to be connected to intervention benefits. Study findings also documented ways that dosage information provides insight into interventions and their effects.

Keywords Dosage · Intervention · Effectiveness · Life skills training · Time wise

First authorship is shared by Laura Ferrer-Wreder and Hans Saint-Eloi Cadely.

L. Ferrer-Wreder (✉)
Department of Psychology, Stockholm University,
Frescati Hagv. 14, Room 232, 106 91 Stockholm, Sweden
e-mail: laura.ferrer-wreder@psychology.su.se

H. S.-E. Cadely
Department of Human Development and Family Studies,
Auburn University, Auburn, AL, USA

C. E. Domitrovich · M. L. Small
Prevention Research Center for the Promotion of Human
Development, Pennsylvania State University, University
Park, PA, USA

L. L. Caldwell
College of Health and Human Development,
Pennsylvania State University, University Park, PA, USA

M. J. Cleveland
The Methodology Center, Pennsylvania State University,
University Park, PA, USA

Introduction

School-based prevention programs are an integral part of the broad spectrum of intervention efforts that are needed to support positive youth development (Greenberg 2004). The diversity, efficacy, and knowledge base surrounding school-based interventions, including school-based drug prevention programs, has grown over the last two decades (Faggiano et al. 2008; Gottfredson and Wilson 2003; Tobler et al. 2000). Historically, schools have served as a main staging ground for drug prevention activities. The present study was an effectiveness trial in which a cohort of adolescents (i.e., 6th-grade students) took part in Life Skills Training (LST) and/or TimeWise: Taking Charge of Leisure Time (Caldwell 2004). This study also illustrates the

insights gained from measuring dosage and linking it to outcomes (Durlak and DuPre 2008).

Successful drug prevention programs often share a common theory of intervention change (Bühler et al. 2008). For example, if an intervention can substantially deepen relevant knowledge and bolster life competence (e.g., including various forms of skills development), then it is posited that these activities will yield reduced drug use and promote positive forms of development. Though there may be commonalities, many useful drug prevention programs also diverge in a number of ways, such as the extent of tailoring done in order to be relevant to a particular group or situation. Programs also vary in focus (e.g., drug-specific versus or in addition to generic life skills development).

LST and TimeWise are universal, school-based drug prevention programs. The interventions share a focus on knowledge and competence promotion as well as have a common recognition that early to middle adolescence is a good point of intervention. Yet, the interventions also part ways on several counts including their overarching theoretical and problem/promotion orientations, level of situational specificity, and types of skills and knowledge targeted for change. These interventions were combined in the present trial because these were evidence-based interventions that were complementary yet still distinct, thereby yielding a potentially potent combination that would be novel for participating youth over the entire course of the intervention trial.

Teens encounter multiple co-occurring risks, and combining empirically supported interventions may be more effective in addressing those risks across contexts and problems (Greenberg 2004). The present study was designed to address the prevention of teen drug use inside and outside of leisure contexts (Caldwell et al. 2004; Ryan and Deci 2000). This point of focus is timely and warranted because leisure time is important to risk prevention and promoting positive development (Caldwell et al. 2004; Witt and Caldwell 2005) and has been less commonplace as a context for prevention than other settings. The combination of the particular empirically supported interventions tested in this trial, LST and TimeWise, may also allow for an improved understanding of the intersections between adolescent free time and other factors associated with teen drug use.

The Interventions

LST (Botvin and Griffin 2004) is a universal school-based drug prevention program that is designed for elementary and middle school students. LST is grounded in social learning, communication, and problem behavior theories. LST consists of activities designed to help youth learn how to say no when tempted to engage in substance use (i.e., drug resistance skills). LST aims to help youth learn to make good decisions, set goals for themselves, and realize the impact of their behaviors (i.e., personal self-management skills). Further, LST consists of activities aimed at improving general social skills. Other LST targets for change include identity development, problem solving or decision-making, interpersonal relationships, physical health maintenance, and correcting normative expectations about substance use. LST has been the subject of numerous large scale controlled intervention trials and has been found to be associated with reduced substance use several years after intervention and for youth with varying demographic characteristics (Botvin and Griffin 2005).

TimeWise: Taking Charge of Leisure Time (Caldwell 2004) is a universal school-based health promotion intervention that is designed to educate youth about how to use their leisure time in healthy ways. TimeWise's core lessons focus on helping students to (a) determine personally satisfying and meaningful leisure activities and interests, (b) understand the benefits of participating in healthy leisure, (c) understand how one's motivation affects one's experience and participation in healthy behaviors, (d) alleviate boredom and increase optimal experience in leisure time, (e) learn how to take responsible action to participate in desired activities, and (f) identify and overcome constraints that get in the way of participation in desired activities. TimeWise focuses largely on health promotion rather than risk reduction. It specifically targets the reduction of drug use in leisure time by helping youth better understand the connection between their leisure time and development. In TimeWise, youth learn that what they do in their leisure time and why they do it may lead to either positive or negative forms of development, including substance use and abuse.

The efficacy of TimeWise has been tested in one randomized control trial. In that study, four rural

schools in Pennsylvania served as treatment schools, and five schools served as no-treatment comparison schools (Caldwell et al. 2004). Results showed short-term benefits for the intervention on several leisure constructs (e.g., motivation, boredom). After three waves of posttest data (springs of 2001, 2002, and 2003), Caldwell et al. (2010) found differential long-term benefits by gender. Relative to comparison girls, TimeWise girls maintained an increased level of interest and intrinsic motivation across time. TimeWise boys showed reduced marijuana and inhalant use over time relative to comparison boys. This pattern of the data suggested that the trends found were promising and require longer-term follow-up.

The first aim of the present study was to evaluate the effectiveness of LST and TimeWise over the course of one academic school year. The same group of intervention teachers was trained in LST and TimeWise. Teachers had a two-day on-site LST training by National Health Promotion Associates and a one-day on-site TimeWise training by the intervention's developer, Dr. Linda Caldwell. Intervention training and programmatic lessons/activities were not integrated but rather occurred or were implemented sequentially (e.g., LST then TimeWise). It was hypothesized that the combined intervention would yield reduced teen drug use and beneficially change leisure constructs shown to be important for positive development and alcohol, tobacco, and other drug (ATOD) risk reduction (e.g., assertiveness skills, favorable motivations for free time). The second aim of the study was to empirically demonstrate the insights gained from the inclusion of implementation information in outcome analyses.

Is More Better?

Our understanding of the importance of implementation and its relation to intervention benefits has improved in recent years. Implementation is multifaceted and can be measured in a diversity of ways including quality, fidelity, program reach, and adaptation (Durlak and DuPre 2008). Another aspect of implementation is dosage, which can be conceptualized as quantity or “how much of the original program has been delivered” (Durlak and DuPre 2008, p. 329). Dosage can be measured, for example, by the time spent in an intervention session, number

of sessions, session lengths, and the duration of the program overall (Gottfredson and Wilson 2003; Nation et al. 2003). In the present trial, implementation was indexed by dosage (i.e., the quantity of program exposure). Dosage was operationalized as the amount of curriculum covered by intervention teachers and the number of prevention lessons attended by students.

As mentioned previously, LST is a well-studied intervention. In addition to traditional outcome evaluation studies, five intervention trials have linked implementation markers to LST program benefits (Botvin et al. 1989, 1990a, b, 1992, 1995). Across these trials, the predominant index of implementation has been observational measures of dosage, which were used as categorical variables in the outcome analyses of these intervention trials. The range of LST intervention outcomes linked to dosage varied from 50 to 100% (Durlak and DuPre 2008). As in much of the rest of the youth intervention research literature, these LST implementation trials demonstrated the positive connection between dosage and reductions in drug use. For example, students exposed to more of the LST program had less of a tendency to engage in substance use as opposed to students exposed to less of the program (e.g., Botvin et al. 1990). The present study has the potential to add to the LST implementation outcome evidence base as well as to build this kind of evidence base for a newer but promising TimeWise intervention.

Much of what we know about prevention in general and school-based drug prevention in particular is predicated on intervention trials that do not empirically link implementation to intervention outcomes. Despite this historical shortcoming, the field has shifted in recent years, and researchers have made the measurement of implementation a regular science activity (i.e., as a given or as the standard). Several hundred empirical studies and a growing number of meta-analyses have soundly connected increased implementation, measured in a variety of ways, to intervention benefits across a diversity of youth-serving promotion and prevention programs (e.g., Durlak and DuPre 2008). Therefore, there is a solid evidence base to support an association between increased implementation and intervention benefits. This article is designed to serve as an empirical demonstration of the now well-established connection between implementation and intervention

benefits as well as to reflect the need for the field to continue to progress and to make implementation outcome research even more commonplace.

In this article, dosage was linked to the outcomes of a combined universal, drug prevention intervention. Outcome analyses were conducted with and without regard to dosage information to demonstrate the contrast in knowledge gained from the inclusion of implementation markers in outcome evaluation studies. It was hypothesized that dosage levels (as indexed by program exposure and lesson coverage) that equaled or exceeded 60% would yield greater intervention benefits relative to lower exposure and coverage levels (less than 60%). The aims of this study were to document the outcomes of two combined, empirically supported interventions as well as reinforce the value of measuring implementation information and linking it to outcomes.

Method

Participants

Demographics presented here are for the final sample after pre-analyses (described in the results section) were conducted. Participants were 715 students from a northeastern American city ranging in age from 11–16 years old ($M = 12$). The sample consisted of 45% males, 54% females, and 1% missing. Participants identified themselves as follows: 64% African American/Black, 13% “Other” ethnic identifier, 10% Latino/Hispanic, 4% American Indian/Native American, 3% White/Non-Latino, 2% Asian, and 4% missing.

Procedure

The present study was embedded in a larger four-year, school-district-wide Safe Schools Healthy Students (SSHS) initiative. SSHS is a series of interventions designed to improve social emotional learning and academic achievement and reduce school dropout. Interventions were predominately implemented by grade level, and no other SSHS interventions were implemented with the participants in this trial.

This effectiveness trial used a quasi-experimental age-cohort design. This design was selected because

this was an effectiveness trial, and district officials opted for universal implementation by grade level. An age-cohort design permits a comparison between a group of individuals who are about the same chronological age, developmental level, and/or belong to a particular social organization (Olweus and Alsaker 1994). Participants in this design are surveyed at different time points and have differential exposure to the intervention. For example, intervention participants (Cohort 2) in the present study were exposed to LST and TimeWise as 6th graders and then surveyed as they entered the 7th grade ($n = 433$). LST and TimeWise were implemented universally with this cohort. Another non-intervention-exposed cohort (Cohort 1) of 7th graders in the same school district served as the comparison group and was surveyed the previous fall ($n = 482$). The survey response rate for Cohort 1 was 83% and 77% for Cohort 2 (see Fig. 1).

All participants completed a self-report survey on school grounds. This study was approved by an institutional review board and received a National Institutes of Health Certificate of Confidentiality. Passive parental consent and student assent were required for participation in the survey administration. Participants were given a participant number that linked their survey responses to their classroom attendance in LST and TimeWise lessons. Grade-level cohort was the unit of assignment in this study. All mainstream schools in this school district with a 6th grade implemented LST and TimeWise as part of standard school curriculum. Alternative schools did not take part in this intervention trial.

Fig. 1 Study design: age-cohort. Circles in this figure represent sampling times

Measures

Demographics

Participants were asked to give information regarding their gender, age, ethnicity, grade in school, and parents' or guardians' level of education.

Drug Use

Two scales were used to measure lifetime and recent drug use. These scales were derived from the Communities That Care Survey (CTC; Arthur et al. 2002) and from the Life Skills Training Questionnaire (LSTQ; National Health Promotion Associates 2001/2002). The Lifetime Drug Use Scale was made up of four items measured on a 7-point Likert scale ranging from 0 (*0 occasions*) to 6 (*40 or more occasions*). These items were averaged to create a lifetime alcohol, marijuana, and other drug use mean score with high numbers indicating an elevated level of drug use ($\alpha = .52$).

The Recent Drug Use Scale was made up of two items from the CTC survey. These items were rated on a 7-point Likert scale ranging from 0 (*0 occasions*) to 6 (*40 or more occasions*) and were averaged to create a recent alcohol and marijuana mean score with high numbers indicating an elevated level of drug use ($\alpha = .61$).

LST-related Scales

The following scales were derived from the LSTQ (National Health Promotion Associates 2001/2002). The Assertiveness Skills Scale was made up of three items that were averaged to create a mean score with low numbers indicating high assertiveness ($\alpha = .64$). Participants rated on a 5-point Likert scale ranging from 0 (*definitely would*) to 4 (*definitely would not*) how likely they would be to engage in different self-confident behaviors in various social situations.

The Anxiety Management Skills Scale was made up of two items averaged to create a mean score with high numbers indicating greater ability to positively handle stress ($\alpha = .64$). Participants were asked to rate their responses on a 5-point Likert scale ranging from 0 (*never*) to 4 (*always*).

Also from the LSTQ, the Refusal Skills Scale 1 ($\alpha = .98$) was made up of six items averaged to

create a mean score with low numbers indicating a high likelihood of refusing various drugs from someone when asked. Participants rated on a 5-point Likert scale ranging from 0 (*definitely would*) to 4 (*definitely would not*) how likely would they be able to handle certain drug-related situations in an anti-drug manner.

A second Refusal Skills Scale 2 ($\alpha = .86$) was made up of five items that were averaged to create a mean score with low numbers indicating a high likelihood of refusing drugs from someone when asked. Participants rated on a 5-point Likert scale ranging from 0 (*definitely would*) to 4 (*definitely would not*) how likely they were to use different drug refusal strategies.

A Drug Use Intentions Scale was made up of five items averaged to create a mean score with high numbers indicating a greater likelihood of engaging in substance use in the coming year ($\alpha = .68$). Participants were asked to rate their responses on a 5-point Likert scale ranging from 0 (*definitely not*) to 4 (*definitely would*).

A Normative Beliefs About Peer Drug Use Scale was made up of five items averaged to create a mean score with high numbers indicating a belief that a great number of one's peers are engaging in substance use ($\alpha = .87$). Participants indicated their responses on a 5-point Likert scale ranging from 0 (*none*) to 4 (*all or almost all*).

TimeWise-related Scales

The Free Time Motivation Scale for Adolescents (FTMS-A; Baldwin and Caldwell 2003) was designed to measure types of adolescent free time motivation and is grounded in self-determination theory (Ryan and Deci 2000).

The Amotivation Scale was made up of four items averaged to create a mean score with low numbers indicating a potentially greater self-determination level (e.g., I don't know, nothing much interests me, $\alpha = .83$). Participants rated their responses on a 5-point Likert scale ranging from 0 (*strongly disagree*) to 4 (*strongly agree*).

The External Motivation Scale was composed of four items averaged to create a mean score with low numbers indicating that youths do not take part in free time activities to gain rewards external to an activity (e.g., I do what I do in my free time because I

am supposed to, $\alpha = .86$). Participants rated their responses on a 5-point Likert scale ranging from 0 (*strongly disagree*) to 4 (*strongly agree*).

The Identified Motivation Scale consisted of three items averaged to create a mean score with high numbers indicating greater goal-oriented behavior (e.g., I do what I do in my free time because I develop skills that I can use later in life, $\alpha = .83$). Participants rated their responses on a 5-point Likert scale ranging from 0 (*strongly disagree*) to 4 (*strongly agree*).

The Introjected Motivation Scale was made up of four items averaged to create a mean score with low numbers indicating a little inclination to engage in activities to gain recognition or avoid guilt from others (e.g., I do what I do in my free time because I want people to like me, $\alpha = .80$). Participants were asked to rate their responses on a 5-point Likert scale ranging from 0 (*strongly disagree*) to 4 (*strongly agree*).

The Intrinsic Motivation Scale was made up of four items averaged to create a mean score with high numbers indicating greater internally regulated behavior (e.g., I do what I do in my free time because I want to have fun, $\alpha = .92$). Participants rated their responses on a 5-point Likert scale ranging from 0 (*strongly disagree*) to 4 (*strongly agree*).

Dosage Instrument

Dosage was measured by the amount of exposure (i.e., quantity) students had to LST and TimeWise. Items for the dosage survey were derived from the TimeWise Classroom Observation Form (Caldwell 2004), the TOP Guide for Program Sites (Philliber Research Associates, n.d.), and the Life Skills Training Fidelity Instrument (National Health Promotion Associates 2001/2002). Intervention teachers ($n = 15$) completed this survey online and logged student attendance and teacher's lesson coverage after the completion of each prevention class. Process data were merged with student survey responses by a participant ID number. The loss of intervention participants when process data was linked to outcome data was 37% (i.e., 433 sixth-grade students should have been exposed to the interventions and dosage information was linked to 271 students; 63% of the sample was retained).

Dosage was operationalized as involving a composite score made up of two variables: students'

classroom attendance during prevention classes and the number of prevention curricula lessons covered by teachers. A composite score has explanatory advantages. A participant could have exemplary classroom attendance during the intervention but his or her teacher may have not given many LST or TimeWise lessons, and vice versa. Thus, a single dosage indicator considered in isolation may mask important characteristics of study participation.

Due to variations in implementation, some students took part only in LST and some participated in both LST and TimeWise. The following dosage groups were created to capture this variation by program and dosage level: high implementation groups (a) LST Hi Group and (b) LST_TW Hi Group, and less than ideal implementation groups (c) LST Mixed Group and (d) LST_TW Mixed Group. The final number of participants after pre-analyses (e.g., inconsistent responders and outliers removed; see Results section) were Intervention Group ($n = 250$) and Comparison Group ($n = 465$), with the Intervention Group consisting of the four dosage groups: LST Hi Group ($n = 97$), LST_TW Hi Group ($n = 51$), LST Mixed Group ($n = 32$), and LST_TW Mixed Group ($n = 70$).

In the LST Hi Group, participants attended 60% or more of LST lessons given by their teacher, and their teacher covered 60% or more of the curriculum. For the LST_TW Hi Group, participants attended 60% or more of LST and TimeWise lessons given by their teacher, and their teacher covered 60% or more of both curricula. For the LST Mixed Group, participants either attended less than 60% of the LST lessons given by their teachers, and/or their teachers covered less than 60% of the curriculum. For the LST_TW Mixed Group, participants either attended less than 60% of the LST and/or TimeWise lessons given by their teachers, and/or their teachers covered less than 60% of either curriculum.

The 60% criterion for the dosage groups was partly based on past LST implementation studies (e.g., Botvin et al. 1989, 1990, 1992, 1995). LST implementation has been measured in different ways both in terms of implementation quality and dosage (e.g., Botvin et al. 1989). A commonality across studies is that most approaches have been based on random observations of LST teachers. Trained observers have made surprise visits to LST classrooms. Dosage-related implementation ratings in LST trials have

often involved observer ratings of how many of the targeted LST lesson objectives are covered by a particular teacher during an observed lesson. LST teachers are observed on more than one occasion and an average implementation score has been created based on the total number of lessons observed for a particular teacher. Participating youth then have received a weighted cumulative implementation score based on the observers' evaluation of teachers' ability to cover LST lesson objectives over time (e.g., Botvin et al. 1992, 1995). Botvin et al. (1995) have conducted outcome analyses for a "high fidelity" subset of intervention participants in a prior LST trial and defined this subset as having been in LST classrooms in which teachers, on average, covered 60% or more of the observed LST lesson objectives. Measurement of students' actual attendance in LST classrooms is not commonly reflected in this dosage index.

The measure of dosage used in this study was a composite score that included individual students' classroom attendance along with number of intervention lessons covered by teachers. The dosage index used in this study has some parallels and points of departure from prior LST trials. Regarding differences, Botvin and colleagues' strategy to measure lesson coverage has tended to focus on observers recording if particular objectives within randomly selected lessons were covered or not (e.g., Botvin et al. 1992, 1995). In this study, intervention teachers self reported on all lessons they implemented and the dosage index is not based on an average across a random sample of observed intervention lessons. Student attendance in intervention classrooms figured prominently in the present study and this aspect of dosage has been implied rather than explicit in several past LST trials. Although there are differences, commonalities include measurement of dosage, particularly in terms of content coverage (although measured differently), as well as the treatment of dosage as a categorical variable in outcome analyses and analysis of a high-dosage group with a 60% cutoff based on level of content coverage (e.g., Botvin et al. 1990, 1992, 1995).

The 60% criterion for the dosage groups was also partly based on Durlak and DuPre's (2008) meta-analysis of youth promotion and prevention studies. These researchers concluded based on part of their analysis that "...near-perfect implementation is unrealistic. Positive results have often been obtained with levels around 60%; few studies have attained levels

greater than 80%. No study has documented 100% implementation by all providers" (p. 331). In this quote, Durlak and DuPre (2008) are referring to implementation in a broad sense as they conceptualized implementation as including dosage but also going well beyond it in their analysis.

Results

Preliminary Analyses

An examination of univariate indices of skewness and kurtosis for the outcome variables revealed skewness and kurtosis values well above an absolute value of 2.0 for the Lifetime and Recent ATOD Use Scales as well as for the Drug Use Intentions Scale. Skewness was also evident on the Intrinsic Motivation Scale. To address this non-normality, a log transformation was used for all non-normal outcome variables.

Because ATOD use was a key programmatic outcome, student responses to the lifetime and recent ATOD use questions were checked for logical consistencies. Students ($n = 20$) who reported an inconsistent response were eliminated from further analyses. Two Pearson chi-square tests (Intervention vs. Comparison Group on gender and ethnicity) and an independent samples t test (Intervention vs. Comparison on age) showed no statistically significant differences between the Intervention ($n = 250$) and Comparison Groups ($n = 465$) on key demographic variables.

When process data were merged with student survey responses by ID number, there was a 37% loss of intervention participants with 63% of the sample was retained. A series of t tests were conducted on dosage instrument items to determine if there was a selection bias in the intervention group ($n = 433$). Two independent samples t tests showed that intervention youth with dosage information plus an ID link to the student survey attended more LST, $t(426) = 3.81$, $p < .001$, and TimeWise classes, $t(200) = 2.72$, $p = .007$, in comparison to intervention youth who did not have an ID link to a student survey (LST student attendance: Intervention Youth _{With Link}: $M = 11.17$, $SD = 5.5$, Intervention Youth _{No Link}: $M = 9.04$, $SD = 5.7$; TimeWise student attendance: Intervention Youth _{With Link}: $M = 4.13$, $SD = 1.3$, Intervention Youth _{No Link}: $M = 3.56$, $SD = 1.6$). Two additional

t tests showed that teachers of intervention youth with and without a link to a survey did not significantly differ in terms of the number of LST and TimeWise classes that were taught to their students. The selection effect seemed to involve students, with a bias towards retaining better attending students in the outcome analysis relative to students who were more likely to be absent for intervention classes.

Other descriptive exploratory analyses showed that relative to youth who only took part in LST (i.e., the stand alone intervention), youth that took part in LST and TimeWise (i.e., the paired intervention) attended significantly more intervention classes $t(378) = 10.04, p = .000$ (LST Only Youth number of classes attended: $M = 10, SD = 5.2$; LST_TW Youth classes attended: $M = 15, SD = 6.7$).

Analysis Strategy for Outcome Analyses

Traditional statistical methods assume that data from each individual are independent—an assumption often violated when students are clustered within schools as in the present study. The use of ordinary least squares regression with clustered data results in underestimated standard error estimates, overestimated test statistics, and inflated Type I error rates (Hox 2002). Thus, we employed multilevel modeling using SAS Proc Mixed to examine the effect of the interventions while accounting for the clustering of individual students within schools as well as within classrooms/teachers.

Outcome analyses were conducted with and without regard to dosage to highlight the level of variability found across analyses as well as to demonstrate what may be gained from dosage information. The dosage variables were entered in the models as a series of four dummy codes to capture contrasts between each of the levels of dosage (i.e., LST Hi, LST_TW Hi, LST Mixed, and LST_TW Mixed) and the comparison group. All models were estimated using full information maximum likelihood, which calculates parameter estimates using the observed information from all cases, rather than deleting those with missing data.

In the tested models, experiment-wise controls were invoked by grouping outcome variables within families of theoretically/empirically meaningful group contrasts (Family 1: primary study outcomes [lifetime and recent ATOD use]; Family 2:

intermediary LST-related outcomes [assertiveness, anxiety management, and refusal skills; drug use intentions, normative beliefs about peer drug use]; Family 3: intermediary TimeWise-related outcomes [amotivation and external, identified, introjected, and intrinsic motivations]). All group contrasts on outcome variables were conducted with a Bonferroni correction to control for Type I error within families.

For each family of outcomes, two sets of analyses were conducted. The first set of analyses conducted did not account for dosage (i.e., an “intention to treat”-type analysis approach; Pagoto et al. 2009). In these models, Group was specified as the fixed factor (Intervention, Comparison). In the second set of analyses, Dosage was specified as a series of four dummy coded variables (Comparison vs. LST Hi; Comparison vs. LST_TW Hi; Comparison vs. LST Mixed; Comparison vs. LST_TW Mixed). School and classroom/teacher were used as the clustering variables in both sets of models. All statistically significant group contrasts reported in this article are described in Table 1 (e.g., mean differences and other key statistics).

No Dosage versus Dosage Outcome Analyses— Family 1: Drug Use

In the first series of tested multilevel models, the outcome variables were the Lifetime and Recent Drug Use Scales. Analyses indicated that the effects of the intervention were statistically non-significant when comparing the Intervention versus Comparison Groups, with no consideration of dosage. The next set of analyses contrasted the four dosage levels versus the Comparison group and yielded non-significant results on all but one group contrast, the Comparison versus the LST Mixed Group ($\beta = 0.09, t = 2.26, p < .05$). The mean difference indicated that the Mixed Group was elevated on Lifetime Drug Use relative to the Comparison Group (see Table 1).

No Dosage versus Dosage Outcome Analyses— Family 2: LST-Related Outcomes

For Family 2, the first set of analyses conducted did not account for dosage. In these analyses, Group was specified as the fixed factor (Intervention, Comparison) in a series of multilevel models using the Family 2 outcome variables (i.e., assertiveness,

Table 1 Key statistics for statistically significant group contrasts

Scale	Groups	Parameter	SEdiff	<i>t</i> value	<i>p</i> value	Low	Hi	<i>df</i>
Lifetime ATOD	LST Mix/C	−.09	.040	2.26	.047	−.18	−.01	10
Assertiveness skills	LST Hi/C	.42	.126	3.30	.011	.13	.71	8
Anxiety mangement	I/C	−.24	.102	2.35	.051	−.48	.00	7
Skills	LST Hi/C	−.34	.147	2.29	.052	−.68	.00	8
Refusal skills scale 1	I/C	−.36	.156	2.33	.053	−.73	.01	7
Drug use intentions	LST_TW Hi/C	.10	.038	2.57	.037	.01	.19	7
External motivation	I/C	.38	.093	4.05	.005	.16	.60	7
	LST Hi/C	.35	.137	2.53	.035	.03	.66	8
	LST Mix/C	.58	.202	2.87	.017	.13	1.03	7
	LST_TWMix/C	.45	.160	2.83	.020	.09	.82	9
Introjected motivation	I/C	.28	.092	2.97	.021	.06	.50	7
	LST_TW Hi/C	.46	.174	2.64	.033	.05	.87	7

Notes. ATOD Alcohol, Tobacco, and Other Drug Use, I/C Intervention versus Comparison Group, LSTHi/C LST Hi versus Comparison Group, LST Mix/C LST Mixed versus Comparison Group, LST_TW Hi/C LST_TW Hi versus Comparison Group, LST_TW Mix/C LST_TW Mixed versus Comparison Group, *Parameter* difference of predicted least square means with Bonferroni adjustment, *SEdiff* standard error difference, *Low* lower 95% confidence interval boundary, *Hi* upper 95% confidence interval boundary, *df* degrees of freedom

anxiety management, and refusal skills [Scales 1, 2], drug use intentions, and normative beliefs about peers drug use). Statistically significant differences were found when comparing the Intervention Group relative to the Comparison Group on Anxiety Management Skills ($\beta = 0.24$, $t = 2.35$, $p < .05$) and Drug Refusal Skills 1 ($\beta = 0.36$, $t = 2.33$, $p < .05$).

The second set of analyses compared the effects of the four dosage groups on the Family 2 outcomes. These results are also displayed in Table 1 and show that dosage was significantly associated with three outcomes: LST Hi versus the Comparison Group on Assertiveness Skills ($\beta = -.42$, $t = 3.30$, $p < .05$) and Anxiety Management Skills ($\beta = 0.34$, $t = 2.29$, $p < .05$) as well as the LST_TW Hi Group versus the Comparison Group on Drug Use Intentions ($\beta = -0.10$, $t = 2.57$, $p < .05$).

Family 2 group contrasts (with and without dosage information) are of primary theoretical and practical interest. Findings described below are reported by Family 2 outcome variables to better highlight the pattern of results with and without dosage information.

Assertiveness Skills

A contrast with dosage information included in the analysis showed that the LST Hi Group reported better Assertiveness than the Comparison Group.

Anxiety Management Skills

Without the inclusion of dosage information, a contrast indicated that the Intervention Group reported better Anxiety Management Skills than the Comparison Group. A contrast with dosage information showed that the LST Hi Group reported better Anxiety Management Skills than the Comparison Group.

Drug Refusal Skills

Counter to the researchers' expectations, analyses without dosage information showed that the Comparison Group reported better drug refusal skills than the Intervention Group.

Drug Use Intentions

Analyses without a consideration of dosage (i.e., Intervention versus Comparison Groups) for this scale were non-significant. Dosage analyses showed one significant group contrast with the LST_TW Hi Group less likely to report intentions to use drugs in the future relative to the Comparison Group (see Table 1).

No Dosage versus Dosage Outcome Analyses— Family 3: TimeWise-Related Outcomes

For Family 3, the first set of analyses conducted did not account for dosage. In these analyses, Group was specified as the fixed factor (Intervention, Comparison) in a series of multilevel models using the Family 3 outcome variables (i.e., amotivation and external, identified, introjected, and intrinsic motivations). Statistically significant differences were found when comparing the Intervention Group relative to the Comparison Group on External Motivation ($\beta = -0.38$, $t = 4.05$, $p < .05$) and Introjected Motivation ($\beta = -0.28$, $t = 2.97$, $p < .05$).

The second set of analyses compared the effects of the four dosage groups on the Family 3 outcomes. These results are displayed in Table 1 and show that dosage was significantly associated with two outcomes: External and Introjected Motivations. There was a significant group contrast between the LST Hi Group versus the Comparison Group on External Motivation ($\beta = -.35$, $t = 2.53$, $p < .05$). The LST Mixed ($\beta = -0.58$, $t = 2.87$, $p < .05$) and LST_TW Mixed Groups ($\beta = -0.45$, $t = 2.83$, $p < .05$) also differed versus the Comparison Group on External Motivation. There was also significant group contrast between the LST_TW Hi Group versus the Comparison Group ($\beta = -.46$, $t = 2.64$, $p < .05$) on Introjected Motivation.

Family 3 group contrasts are of main interest. The findings described below are only for significant group contrasts on Family 3 outcome variables (with and without dosage information by outcome).

External Motivation

A group contrast without dosage information showed that the Intervention Group reported less inclination to engage in free time activities to meet the expectations of others. A group contrast that included dosage information showed that the LST Hi and Mixed Groups (i.e., LST Mixed and LST_TW Mixed) reported lower motivation to engage in free time activities to gain rewards external to their leisure activities relative to the Comparison Group.

Introjected Motivation

A group contrast that did not consider dosage indicated that the Intervention Group reported less inclination to engage in free time activities to gain recognition or avoid guilt from others than the Comparison Group. The LST_TW Hi Group also reported lower levels of Introjected Motivation relative to the Comparison Group.

Discussion

The purpose of the initiative described in this article was first and foremost to prevent teen drug use and beneficially change important constructs related to positive development and reduced ATOD risk. Two empirically supported interventions, LST and TimeWise, were implemented in a school-based effectiveness trial context. Across the analyses, with and without dosage information, intervention-related benefits were found on anxiety management skills as well as external and introjected motivations relative to a Comparison Group. When considering dosage information only, the findings expand to include program-related benefits on assertiveness skills and drug use intentions.

The noted intervention benefits are consistent with past research on LST (Botvin and Griffin 2004, 2005) and TimeWise (Caldwell et al. 2004; Caldwell et al. 2010). However, the primary goal of the initiative was only partially met in that there were no demonstrated program-related benefits for ATOD use. One potential reason for this lack of findings is that a longer term follow-up of participants may be needed to capture changes in drug taking behavior. Many of the LST intervention studies are designed to examine program effects among middle school students up to three years after program exposure (Botvin et al. 1990). In addition, ATOD intervention-related benefits emerged in a TimeWise efficacy trial at the 3-year follow-up (for boys; Caldwell et al. 2010). LST and TimeWise trials both point toward the need for long term follow-up with youth whose patterns of drug use normatively change over the course of adolescence. Additional data were collected and follow-up analyses are planned for this study sample.

There was an observed iatrogenic finding for drug refusal skills. In this case, the Comparison Group reported better drug refusal skills than the Intervention Group. Contrasts between the dosage and comparison groups yielded non-significant results on this scale. Main patterns of results in this study included the detection of intention-to-treat-type group differences (or Intervention versus Comparison Group) plus differences by dosage group relative to the Comparison Group. Another pattern of results was to only detect group differences in a dosage group (or groups) versus the comparison group. The iatrogenic finding for refusal skills is at odds with these two patterns of findings as well as contradicts the overall direction of this study's results which were predominately showing intervention-related benefits. For example, 10 out of 12 statistically significant group contrasts found in this study were in the direction of an intervention benefit. This type of contraindicated finding on drug refusal skills is also not commonly reported in the published LST or TimeWise intervention research literature. Thus, this result generally departs from other findings in the study and the most closely relevant intervention research literature. The careful documentation of intervention-related benefits and harm are of critical importance to the prevention field (Flay et al. 2005). Though this finding is indeed important to record, it should be noted that it awaits replication.

Is More Better?

As many researchers and practitioners have recognized, an empirically sound intervention is only the first step in the effort to promote positive change on a large scale and in real world settings. More scientific knowledge is needed about how programs perform under a variety of circumstances, ranging from the ideal to the less than ideal. Better understanding the limitations and values of dosage information in intervention trials is one of the many steps to be taken that will move the intervention field forward.

Along these lines, it was hypothesized that moderate-to-high levels of program exposure and lesson coverage (60% or more) would yield greater intervention benefits relative to lower exposure and coverage levels. High dosage groups evidenced hypothesized intervention benefits on five out of seven (i.e., 71%) dosage-linked statistically significant and

beneficial group contrasts (i.e., three to LST Hi and two to LST_TW Hi). In these contrasts, the LST Hi Group showed better assertiveness and anxiety management skills and lower levels of external motivation relative to the Comparison Group. The LST_TW Hi Group was less likely to engage in free time activities in order to gain recognition from others (i.e., Introjected Motivation Scale) and was lower in their intention to use drugs in the future relative to the Comparison Group.

The Mixed Dosage Groups were involved in two significant group contrasts that showed intervention benefits on the External Motivation Scale (i.e., 29% of the dosage-linked beneficial, statistically significant group contrasts) but were also involved in one iatrogenic finding. Specifically, the LST Mixed Group was elevated on Lifetime ATOD use relative to the Comparison Group. An iatrogenic finding on drug use is particularly important in consideration of the harm associated with drug use and because drug use was a primary target for intervention change. A potential implication of this finding is that if empirically supported interventions are not implemented well, at least with respect to dosage as one index of implementation, then the expected benefits may not be evidenced in the variety and abundance that one would expect, and an iatrogenic effect may even be found.

An alternative interpretation for this iatrogenic finding on lifetime drug use would be that the Comparison Group differed from the Intervention Groups in some important, unexamined way. If this were the case, it would be unlikely that the iatrogenic effect would only be demonstrated for one dosage group (i.e., the LST Mixed Group) and not in any other dosage group. Though possible, it is considered unlikely that the Comparison Group differed markedly from the Intervention Groups in that the life circumstances of the students in this school district were not markedly different from one school year to the next. Comparisons of the Intervention and Comparison Groups on demographic variables also yielded no statistically significant differences.

Focusing on the question of “Is more better,” the answer to this question based on these study results is a tentative yes. The overall pattern of results provides support for the dosage hypothesis put forward in this study (i.e., moderate-to-high exposure and lesson coverage tends to be connected to intervention

benefits with some caveats) and is in line with the LST implementation outcome evidence base (e.g., Botvin et al. 1989, 1990; 1992, 1995). Less than ideal implementation—either students' not being present during intervention lessons and/or poor coverage of the intervention curriculum—tended to not yield the array of benefits that would be expected from what are otherwise empirically supported interventions (e.g., Botvin and Griffin 2004; Caldwell et al. 2004). It is important to note that other relevant studies are needed in order to have greater confidence in these conclusions. However, the present study results support the existing implementation outcome intervention literature (e.g., Durlak and DuPre 2008; Gottfredson and Wilson 2003; Nation et al. 2003) and specifically highlight the importance of adequate coverage and the presence of students in prevention classrooms.

What Else Is Gained from Dosage Information?

The value of measuring dosage is borne out in a comparison of the analyses conducted with and without dosage information. In several of the contrasts, the no-dosage information (or intention-to-treat-type) analyses pointed to an intervention-related effect, and the dosage analyses were able to point to the group of participants in which the effect was most evident. For example, regarding benefits, there was a significant group contrast in which the Intervention Group showed better anxiety management skills relative to the Comparison Group. Contrasts by dosage group indicated that the LST Hi Group was the group that reported the improvements relative to the Comparison Group, also illustrating the broader point that dosage may be important to intervention outcomes. A parallel pattern of findings were also demonstrated in Family 3: intermediary TimeWise-related outcomes for the External and Introjected Motivation scales.

Dosage analyses were also able to detect two group differences that would have not been identified if one relied solely on intention-to-treat-type analyses (i.e., Intervention versus Comparison Group). These significant group contrasts dealt with both a beneficial and iatrogenic finding. On the positive side, the LST_TW Hi Group had lower reported drug use intentions than the Comparison Group. On the problematic end, the LST Mixed Group was elevated

on lifetime ATOD relative to the Comparison Group. In both of these cases, the Intervention versus Comparison Group showed no significant group differences. Therefore, the analysis of dosage and outcome information together lent unique insight into the tested interventions and their effects.

Programmatic Findings

Shifting to the subject of what the paired intervention (i.e., LST plus TimeWise) yielded versus LST on its own, it can be concluded from the study results that intervention-related benefits tended to be found when the paired (i.e., LST_TW) and stand alone (i.e., LST Only) interventions were implemented with a moderate-to-high level of dosage. Relative to youth in the stand alone intervention, youth in the paired intervention attended significantly more intervention classes. Yet, the number of intervention-related benefits was relatively similar for the paired and stand alone programs (e.g., LST Hi, three, to LST_TW Hi, two, or, collapsing across dosage groups, LST Hi + Mixed, four, to LST_TW Hi + Mixed, three). The test of the effectiveness of the paired intervention (i.e., LST_TW) at a moderate-to-high level of dosage may have not been ideal in that there was a concern for Type 2 error in this study. Whereas the number of participants per intervention group was closer to adequate for the LST Hi and LST_TW Mixed Groups, the number of participants and the resulting power to detect statistically significant group differences in the LST_TW Hi and LST Mixed Groups may have been compromised by the number of participants in these groups. Additional intervention waves were implemented in this school district, and analyses with more cohorts of LST and TimeWise participants may yield improved insight into the relative benefits of paired versus stand alone interventions as well as the researchers' concern about Type 2 error in this study.

Another noteworthy finding from a programmatic view of the study results was that the LST Groups (Hi and Mixed) showed a program-related benefit on a TimeWise-related outcome (the External Motivation scale). There may be some unexamined commonalities between these two interventions. Future studies that combine what are otherwise free-standing empirically supported interventions would benefit from a careful curriculum and logic model study as well as

factor analytic studies of intervention outcome variables (prior to intervention) that informs an a priori strategy for how to best implement the two interventions in tandem. An alternative explanation for the aforementioned finding is that intervention teachers were trained in both curricula and this may have lent itself to combining of ideas and intervention strategies.

Study Limitations

This was a unique effectiveness trial. Study findings may not generalize to other samples. Participants were either only exposed to LST or were exposed to LST first and then TimeWise. An ideal trial would have been if the two interventions were counterbalanced among different schools in the school district (with and without the program pairing). In this effectiveness study, it is only possible to make programmatic comparisons between the combination of LST plus TimeWise and LST on its own, with the added dimension of variability in intervention-related effects according to dosage.

Another limitation involved the less-than-ideal internal consistency of selected scales. For example, the Lifetime and Recent Drug Use Scales as well as the Assertiveness and Anxiety Management Skills Scales had alphas ranging from .52 to .64. These scales were created based on prior psychometric studies by the scale developers as well as item analysis for this sample. Future research would benefit from additional factor analytic and psychometric studies of these scales. Intervention-related results were found for assertiveness and anxiety management skills as well as lifetime drug use. In light of the less than ideal internal consistencies of these scales with this sample, these findings should be interpreted with caution and await confirmation by future research.

A dosage hypothesis was put forward in this study, and supporting evidence was provided that pointed towards a connection between moderate-to-high dosage levels and intervention-related benefits. An alternative explanation for the dosage hypothesis was that a selection effect was at work. The most difficult-to-reach students may not be attending prevention classes held in schools or even may not be in school regularly or at all. There was support for a student-level selection effect in that a number of participants

were lost to the outcome analysis when process data were linked to outcome data. Analysis of the lost versus retained intervention youth indicated that participants who were more likely to attend intervention classes, both LST and TimeWise, were also more likely to have completed the student survey and as a consequence have been included in the outcome evaluation analyses. Teachers' lesson coverage was not shown to be a significant factor in whether intervention students were retained for the outcome analysis or not. Thus, a student-level selection effect serves as a qualification to the evidence presented in this study in support the dosage hypothesis.

Beyond the aforementioned selection bias in which student attendance in intervention classes was connected to retention of those youth in the outcome analyses, many other factors may have contributed to the implementation variability that was shown in this trial. The variability in dosage in the present study was probably non-random. As Stuart et al. (2008) noted, "Characteristics of these participants determine how much they comply with the demands of the treatment protocol" (p. 289). Analysis strategies designed to better understand factors that contribute to variability in program participation include complier average causal effects (CACE), instrumental variables analyses, and propensity score matching (e.g., Foster 2003; Jo 2002; Stuart et al. 2008). The spirit of the argument implied by many of these analyses approaches is that additional insight into intervention effects can be gained by better understanding "compliers" (i.e., those who complete the intervention as intended by program developers). An important innovation in these analysis strategies is that steps are taken to compare compliers in the intervention group and a similar subgroup within the comparison group: those youth in the comparison group who are most like compliers in the intervention group. Not having the most appropriate comparison group in outcome evaluation analysis may lead to an underestimation of intervention effects (Stuart et al. 2008).

Although the spirit of these approaches is certainly relevant to the present study, a literal application of some of these approaches, particularly CACE analysis, may not be ideally suited to the design properties and intentions of this study. For example, as an effectiveness trial, in this study there was no random assignment to condition, and participants

with less than ideal intervention participation (i.e., those in the mixed groups) evidenced program-related change, thereby violating the exclusion–restriction assumption (Stuart et al. 2008). The idea of compliance to a study protocol was also complicated in the present study. The intervention teachers' and program participants' compliance was of importance in this effectiveness trial, not just the compliance of the young person taking part in the intervention. It is likely that those intervention teachers with more resources at hand may have had an implementation advantage relative to those teachers who had less to draw upon. The dosage index used in this study was a composite score that reflected participants' and teachers' compliance with the intervention protocols. A more nuanced approach such as those discussed by Stuart et al. (2008) such as latent class analysis to identify clusters of "participation classes" (p. 296) would provide a stronger test of the dosage hypothesis. Developing appropriate strategies to incorporate an improved understanding of intervention compliance in students and teachers (or implementers of interventions) in outcome evaluation analyses is a promising strategy to explore in future effectiveness trials. Measuring the wide range of factors that contributed to variability in dosage level or compliance to the intervention protocols was not done in this study but would have had the potential to be very enlightening.

Even in the context of these important caveats, it may be worthwhile to consider this study's dosage-related findings in light of the increasingly well-developed evidence base that supports an association between "good implementation" and intervention benefits (Durlak and DuPre 2008). A graduated understanding of what good implementation ultimately consists of will certainly improve interventions that aim to benefit youth. Dosage is but one marker of good implementation and linking this apparently simple index to outcome data is a practically complicated matter. Overcoming the challenges associated with efficiently collecting linked implementation and outcome data at the individual level is a promising direction for future research. Such strategies are of importance to advancement in the intervention field. Other study limitations included the need for even finer grained implementation data that spoke to the quality of implementation as well as youth perceptions of the interventions.

Study Contributions

The present study differs from the previous research literature in that LST was paired with TimeWise. Few researchers have reported the results of the combination of two empirically supported interventions. As noted, there was an indication that the two interventions may have more in common than would be expected. For example, participants exposed to LST at a moderate-to-high level showed an intervention-related benefit on the leisure-related External Motivation Scale relative to the Comparison Group. This was unexpected, and additional research is needed to replicate and better understand this finding. Further, the present study was also an effectiveness trial, meaning that it was conducted in a real-life setting by service providers. More effectiveness trials are needed in the research literature. Another unique contribution of this study is that it adds to the LST implementation outcome evidence base.

In summary, study results indicated that there were no program-related benefits for drug use. Yet, promising intervention-related benefits were found for assertiveness and anxiety management skills and leisure-related external and introjected motivations as well as drug use intentions. In this study, moderate-to-high exposure and lesson coverage tended to be connected to intervention benefits. Within the noted limitations of this study, it appeared that less-than-ideal implementation was linked to lackluster benefits and one iatrogenic finding. Beyond the potentially important connection to outcome, dosage information also provided insight to better understand the tested interventions and their effects.

Acknowledgments We gratefully acknowledge the assistance and guidance of the faculty, staff, and students at the school district in which this study was conducted. We also acknowledge the long time support of this project by Mark T. Greenberg, Bennett Chair of Prevention Research, Director of the Prevention Research Center and Associate Director, Child Youth and Families Consortium, The Pennsylvania State University. Financial support for this project was provided through funding from SSHS, NIDA F31 DA024535, and NIDA T32 DA017629.

References

- Arthur, M. W., Hawkins, J. D., Pollard, J. A., Catalano, R. F., & Baglioni, A. J. (2002). Measuring risk and protective factors for substance use, delinquency, and other

- adolescent problem behaviors. *Evaluation Review*, 26(6), 575–601.
- Baldwin, C. K., & Caldwell, L. L. (2003). Development of the free time motivation scale for adolescents. *Journal of Leisure Research*, 35(2), 129–151.
- Botvin, G. J., & Griffin, K. W. (2004). Life skills training: Empirical findings and future directions. *Journal of Primary Prevention*, 25(2), 211–232.
- Botvin, G. J., & Griffin, K. W. (2005). Prevention science, drug abuse prevention, and life skills training: Comments on the state of the science. *Journal of Experimental Criminology*, 1(1), 63–78.
- Botvin, G. J., Dusenbury, L., Baker, E., James-Ortiz, S., & Kerner, J. (1989). A skills training approach to smoking prevention among Hispanic youth. *Journal of Behavioral Medicine*, 12(3), 279–296.
- Botvin, G. J., Baker, E., Dusenbury, L., Tortu, S., & Botvin, E. M. (1990a). Preventing adolescent drug abuse through a multimodal cognitive-behavioral approach: Results of a 3 year study. *Journal of Consulting and Clinical Psychology*, 58(4), 437–446.
- Botvin, G. J., Baker, E., Filazzola, A. D., & Botvin, E. M. (1990b). A cognitive-behavioral approach to substance abuse prevention: 1 year follow-up. *Addictive Behaviors*, 15(1), 47–63.
- Botvin, G. J., Dusenbury, L., Baker, E., James-Ortiz, S., Botvin, E. M., & Kerner, J. (1992). Smoking prevention among urban minority youth: Assessing effects on outcome and mediating variables. *Health Psychology*, 11(5), 290–299.
- Botvin, G. J., Baker, E., Dusenbury, L., Botvin, E. M., & Diaz, T. (1995). Long-term follow up results of a randomized drug abuse prevention trial in a white middle-class population. *Journal of the American Medical Association*, 273(14), 1106–1112.
- Bühler, A., Schröder, E., & Silbereisen, R. K. (2008). The role of life skills promotion in substance abuse prevention: A mediation analysis. *Health Education Research*, 23(4), 621–632.
- Caldwell, L. L. (2004). *Time wise: Taking charge of leisure time*. Scotts Valley, CA: ETR Associates.
- Caldwell, L. L., Baldwin, C. K., & Walls, T. (2004). Preliminary effects of a leisure education program to promote healthy use of free time among middle school adolescents. *Journal of Leisure Research*, 36, 310–335.
- Caldwell, L. L., Ridenour, T. A., Smith, E. A., & Maldonado-Molina, M. (2010). *Results from a three-year efficacy trial on the effects of a leisure-based intervention on adolescent substance use*. The Pennsylvania State University: Unpublished manuscript.
- Durlak, J., & DuPre, E. (2008). Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology*, 41, 327–350.
- Faggiano, F., Vigna-Taglianti, F. D., Versino, E., Zambon, A., Borraccino, A., & Lemma, P. (2008). School-based prevention for illicit drugs use: A systematic review. *Preventive Medicine: An International Journal Devoted to Practice and Theory*, 46(5), 385–396.
- Flay, B. R., Biglan, A., Boruch, R. F., Castro, F. G., Gottfredson, D., Kellam, S., et al. (2005). Standards of evidence: Criteria for efficacy, effectiveness and dissemination. *Prevention Science*, 6(3), 151–175.
- Foster, E. M. (2003). Propensity score matching: An illustrative analysis of dose response. *Medical Care*, 41(10), 1183–1192.
- Gottfredson, D. C., & Wilson, D. B. (2003). Characteristics of effective school-based substance abuse prevention. *Prevention Science*, 4(1), 27–37.
- Greenberg, M. T. (2004). Current and future challenges in school-based prevention: The researcher perspective. *Prevention Science*, 5(1), 5–13.
- Hox, J. (2002). *Multilevel analysis techniques and applications*. Mahwah, NJ: Lawrence Erlbaum.
- Jo, B. (2002). Estimation of intervention effects with non-compliance: Alternative model specifications. *Journal of Educational and Behavioral Statistics*, 27(4), 385–409.
- Nation, M., Crusto, C., Wandersman, A., Kumpfer, K. L., Seybolt, D., Morrissey-Kane, E., et al. (2003). What works in prevention: Principles of effective prevention programs. *American Psychologist*, 58(6/7), 449–456.
- National Health Promotion Associates. (2001/2002). *Life Skills Training Questionnaire*. White Plains, NY: Author.
- Olweus, D., & Alsaker, F. D. (1994). Assessing change in a cohort-longitudinal study with hierarchical data. In D. Magnusson, L. R. Bergman, G. Rudinger, & B. Toerestad (Eds.), *Problems and methods in longitudinal research: Stability and change* (pp. 107–132). New York, NY: Cambridge University Press.
- Pagoto, S. L., Kozak, A. T., John, P., Bodenlos, J., Hedeker, D., Spring, B., et al. (2009). Intention-to-treat analyses in behavioral medicine randomized clinical trials. *International Journal of Behavioral Medicine*, 16(4), 316–322.
- Philliber Research Associates (n.d.). *The evaluation of the teen outreach program: A guide for program sites*. Accord, NY: Author.
- Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, 55(1), 68–78.
- Stuart, E. A., Perry, D. F., Le, H.-N., & Ialongo, N. S. (2008). Estimating intervention effects of prevention programs: Accounting for noncompliance. *Prevention Science*, 9(4), 288–298.
- Tobler, N. S., Roona, M. R., Ochsborn, P., Marshall, D. G., Streke, A. V., & Stackpole, K. M. (2000). School-based adolescent drug prevention programs: 1998 meta-analysis. *The Journal of Primary Prevention*, 20(4), 275–336.
- Witt, P. A., & Caldwell, L. L. (2005). *Recreation and youth development*. State College, PA: Venture Publishing.