



The Impact of Implementation Fidelity on Student Outcomes in the Life Skills Training Program

Enya B. Vroom¹ · Oliver T. Massey¹ · Svetlana Yampolskaya¹ · Bruce L. Levin^{1,2}

Published online: 20 June 2019

© Springer Science+Business Media, LLC, part of Springer Nature 2019

Abstract

Social and emotional learning (SEL) programs have become increasingly popular during the last 20 years. Although the benefits of SEL programs are significant, the issue of implementation fidelity often arises. The purpose of this study was to assess the impact of implementation fidelity in the Life Skills Training (LST) program implemented with middle school students in a large Florida school district. Three core elements of implementation fidelity were assessed including: (1) adherence; (2) participant responsiveness; and (3) quality of delivery. Student survey data were collected from 4812 students and 104 classrooms in 16 middle schools that participated in the LST program. Multilevel modeling was used to assess the effect of individual-level [gender, race/ethnicity, and socioeconomic status (SES)] and classroom-level characteristics (adherence, participant responsiveness, and quality of delivery) on students' SEL outcomes measured at posttest. At the individual level, results indicated that students' race/ethnicity and SES were significantly associated with predicting student SEL outcomes at posttest. At the classroom level, participant responsiveness was significantly associated with predicting student SEL skills at posttest. Findings are discussed in terms of implications for research and practice.

Keywords Social and emotional learning · Life skills training · Implementation fidelity · Multilevel modeling · School-based prevention

Introduction

Social and emotional learning (SEL) has become increasingly popular during the past two decades. Schools, families, researchers, and policy makers have come to the realization that a child's social well-being and emotional well-being are important (Weissberg, Durlak, Domitrovich, & Gullotta, 2015), and the skills gained from SEL curriculums can have a positive effect on outcomes later in life (e.g., mental health and substance use) (Klapp et al., 2017). SEL can be described as the process of youth gaining and properly employing the attitudes, knowledge, and skills that are necessary to comprehend and manage emotions, display and

feel empathy for others, build and maintain positive relationships, set and achieve goals, and make responsible decisions (Collaborative for Academic, Social, and Emotional Learning, n.d.; Weissberg et al., 2015). A child lacking the skills necessary to comprehend and manage his/her emotions may have less than optimal cognitive and social development. Youth with inadequate emotional skills may fail to feel empathy for others and have difficulties focusing on learning and controlling their behavior (Brackett, Elbertson, & Rivers, 2015). Social and emotional capacity can influence a youth's ability to meet the demands of the classroom, succeed from instruction (Zinsser & Dusenbury, 2015), and assist in preventing problem behaviors (e.g., substance use) (Durlak, Weissberg, Dymnicki, Taylor, & Schellinger, 2011).

Schools serve as a unique and effective outlet to provide prevention programs specifically focusing on SEL. This is due to schools becoming increasingly responsible for the social and emotional needs of youth (Meyers, Tobin, Huber, Conway, & Shelvin, 2015), and being a primary setting where problem behaviors can arise (e.g., substance use, relationship conflict, or violence) (Greenberg, 2010). The growth of evidence of empirically supported school-based

✉ Enya B. Vroom
evroom@usf.edu

¹ Department of Child and Family Studies, College of Behavioral and Community Sciences, University of South Florida, 13301 Bruce B. Downs Blvd., Tampa, FL 33612, USA

² College of Public Health, University of South Florida, 13201 Bruce B. Downs Blvd., Tampa, FL 33612, USA

prevention programs has caused a considerable change in both federal and state policy legislation (Greenberg, 2004). A major example of this is the *Every Child Succeeds Act* (2015), formerly known as the *No Child Left Behind Act* (2001), which places a heavy emphasis on SEL standards and the utilization of evidence-based practices in schools (Greenberg, 2004; Reyes, Brackett, Rivers, Elbertson, & Salovey, 2012). This shift in legislation and increased awareness of SEL and evidence-based programming, coupled with federal and state budget cuts, have caused an influx of evidence-based SEL programs into schools. This has subsequently created a need for greater accountability and efficiency within programs initiated with youth (Mihalic & Elliot, 2015) as well as more consideration of the barriers schools encounter when conducting large-scale implementation initiatives (Goncy, Sutherland, Farrell, Sullivan, & Doyle, 2015).

A major meta-analysis of universal school-based SEL programs has demonstrated that SEL programs can be effective in improving social and emotional skills, behaviors, attitudes, and academic performance as compared to control groups (Durlak et al., 2011). Although many of these programs showed positive results, they often did not monitor the implementation of the programs. Forty-three percent of the studies had to be excluded because they did not employ any technique to monitor the quality of implementation (Durlak et al., 2011). Durlak and DuPre's (2008) review of the literature addressing health and prevention programs for youth indicated that out of the 500 reviewed studies, only 59 studies assessed the relationship between implementation fidelity and program outcomes. Although 63% of the 59 studies examined adherence, only approximately 10% monitored participant responsiveness and competence (i.e., quality of delivery) (Durlak & DuPre, 2008; Goncy et al., 2015). The present study was designed to add to the knowledge base surrounding the core elements of implementation fidelity associated with program outcomes that are not frequently monitored within SEL programs.

Fidelity

Fidelity can be defined as the degree to which the key components of a program or practice that are essential for programmatic impact are maintained when a program is implemented (Allen, Linnan, & Emmons, 2012). Research has shown that high-quality implementation can be strongly associated with positive outcomes (Durlak, 2015). Fidelity may also be an important concern in determining why an intervention succeeds or fails (Dusenbury, Brannigan, Falco, & Hansen, 2003). If a program lacks quality implementation, the chances of producing significant and positive outcomes among its participants may be vastly reduced. As Fixsen and Blase emphasized (n.d.), it cannot be presumed

that consumers will achieve positive and expected outcomes when the program itself was not implemented with fidelity. Additionally, the worth of a program and the theoretical basis driving essential program components cannot be assessed unless it was implemented effectively (i.e., internal validity) (Allen, Shelton, Emmons, & Linnan, 2018; Dane & Schneider, 1998; Durlak & DuPre, 2008). This is why monitoring the implementation and collecting data can assist in determining what influenced the outcomes, whether they be positive or negative, and replicating the program (i.e., external validity) (Carroll et al., 2007; Durlak & DuPre, 2008).

Over time, the definition of fidelity has evolved to include multiple core elements, the most salient being: (1) adherence; (2) dosage; (3) quality of delivery; (4) participant responsiveness; and (5) program differentiation (Allen et al., 2012; Carroll et al., 2007; Dane & Schneider, 1998; Dusenbury et al., 2003). Although adherence and dosage are the most commonly assessed elements, high adherence achievement may be moderated by other elements such as participant responsiveness and quality of delivery (Carroll et al., 2007). Therefore, it would prove to be beneficial to examine such elements in order to establish internal and external validity of a program or intervention (Dane & Schneider, 1998). The current study focused on adherence, participant responsiveness, and quality of delivery, because these elements have proven to affect programmatic outcomes and influence the variability in program implementation (Berkel, Mauricio, Schoenfelder, & Sandler, 2011).

Adherence can be defined as whether a program is being implemented as it was originally developed, and the critical elements of the program are being presented or addressed (Carroll et al., 2007; Dusenbury et al., 2003). *Participant responsiveness* addresses how participants are engaged, involved, or respond to a program (Carroll et al., 2007). *Quality of delivery* can be defined as the manner and level of skill in which the teacher delivers a program (Sutherland, Conroy, McLeod, Algina, & Kunemund, 2018). This has less to do with how he/she follows the guidelines and reads from a script, and more to do with how he/she acts as a facilitator, coach, their level of program knowledge and understanding, attitude, preparedness, and enthusiasm (Carroll et al., 2007; Dusenbury et al., 2003; Mihalic, 2002).

Although the empirical results produced by researchers in a controlled setting suggest a program is generalizable, the need for adaptations may still be evident when administered in real-world settings (Wright, Lamont, Wandersman, Osher, & Gordon, 2015). Often, adaptations are needed in order to ensure the program is appropriate, and can serve to improve the impact and “fit” between the program and the specific population and setting (Allen et al., 2012; Durlak, 2015). For example, how fidelity interacts with certain demographic variables (e.g., race/ethnicity) may be particularly important to a minority group and their outcomes, and certain

modifications may need to be made in future efforts to obtain the best possible outcomes.

Life Skills Training (LST) Program

The LST program is an evidence-based SEL program that teaches social and emotional skills as well as drug resistance skills to elementary, middle, and high school students (Botvin & Griffin, 2004). The curriculum was originally designed to address development characteristics (e.g., psychological, cognitive, and attitudinal) associated with alcohol, tobacco, and illicit drug use. The goal is to reduce motivation to engage in drug use and decrease susceptibility to drug use associated with social influence (Botvin, Baker, Dusenbury, Botvin, & Diaz, 1995; Botvin, Griffin, Diaz, & Ifill-Williams, 2001a). The program emphasizes teaching competence enhancement approaches including decision-making skills, assertiveness skills, communication skills, and skills for coping with anxiety and anger (Botvin & Griffin, 2004).

LST is a classroom-based program that includes over 30 lessons that are taught over the course of 3 years. The first level of the program includes 15 core lessons, which are mandatory, and the second and third levels include 10 and 5 lessons, respectively (Blueprints for Healthy Youth Development [BHYD], n.d.). The first level is intended to be delivered to 6th and/or 7th grade students. Booster lessons are administered in the following two grades: 7th/8th–9th. The lessons range between 45 and 50 min, or the length of a class period, and are delivered one to two times per week. Trained classroom teachers, who are required to attend a 2-day workshop where intervention materials and guidelines are provided, deliver the program (Botvin & Griffin, 2004).

LST has frequently been evaluated for its effectiveness over the past 30 years (BHYD, n.d.), including a number of quasi-experimental studies and randomized control trials (Jagers, Harris, & Skoog, 2015). LST has proven to be effective at reducing substance use among participants (Botvin, et al., 2001a; Botvin, Griffin, Diaz, & Ifill-Williams, 2001b) as well as increasing self-efficacy, self-management, social skills, decision-making, and problem solving (Botvin & Griffin, 2004; Jagers et al., 2015). LST also has been successfully adapted for rural, urban, and economically disadvantaged minority youth (BHYD, n.d.) as well as internationally (Velasco, Griffin, Botvin, Celata, & Gruppo LST Lombardia, 2017).

Implementation fidelity has also been a concern in many of the evaluations conducted of the program. It has been found that adherence and dosage are important considerations concerning the effectiveness of the program. Disregarding the assessment of implementation fidelity may have detrimental effects on student outcomes (Botvin, Epstein, Baker, Diaz, & Ifill-Williams, 1997; Botvin & Griffin, 2004; Jagers

et al., 2015). Past research has indicated that higher levels of teacher adherence, dosage, and quality of delivery of LST can be related to better student outcomes and classroom behaviors (Botvin, Baker, et al., 1995; Botvin, Griffin, & Nichols, 2006; Mihalic, Fagan, & Argamaso, 2008). However, much of the literature tends to examine student classroom behavior as opposed to student programmatic outcomes, may not directly link certain elements of fidelity to outcomes (Mihalic et al., 2008), and may combine elements (i.e., adherence and dosage) (Botvin, Baker, et al., 1995). In addition, constructs like quality of delivery and participant responsiveness may be examined less frequently due to the nature of assessment (i.e., observations), which can be costly and time-consuming (Domitrovich, Gest, Jones, Gill, & Sanford DeRousie, 2010).

Current Study

The present study addressed the gap in the literature surrounding three elements of implementation fidelity, adherence, participant responsiveness, and quality of delivery, associated with student outcomes, which are core elements that are not frequently monitored within SEL programs. Although Life Skills Training (LST) programs have been extensively adopted in schools, research has shown that implementation fidelity varies widely by classroom teachers and the program may often be conducted with weak fidelity. This has also proven to be true during controlled research settings (Botvin & Griffin, 2004). Evaluations of prevention programs rarely examine more than one element, the elements' interactions, and their relationship with student outcomes (Berkel et al., 2011; Reyes et al., 2012). The objectives of this preliminary study were: (1) to assess the impact of implementation fidelity in the LST program on student outcomes; and (2) to identify if there is a differential effect of fidelity at the individual and classroom levels. Based on the core elements, it was hypothesized there would be significant associations between the core elements including adherence, participant responsiveness, and quality of delivery and an improvement in students' SEL skills. The two evaluation questions included: (1) Do any core elements of implementation fidelity (as defined by adherence, participant responsiveness, or quality of delivery) predict students' SEL skills differently? and (2) Does the overall quality of implementation affect students' SEL skills at the individual and classroom levels?

Methods

Participants

This preliminary cross-sectional study examined the relationship between implementation fidelity and student

outcomes in the first year and level (i.e., 6th grade) of LST program implementation within a large Florida school district. Student data were obtained from 16 middle schools in the school district, and teacher data (i.e., classroom data) were obtained from a not-for-profit organization (NFPO) that conducted classroom observations and fidelity checks during program implementation.

Participants included 4812 6th grade students attending the 16 middle schools as well as 104 teachers (i.e., representing 104 classrooms) who were observed by staff of the NFPO. Demographic data were collected by the school district. The demographics of the 16 schools that participated in the study included 47–52% male, 48–53% female, 8–96% on free or reduced lunch status, 25–95% minority, and had between 681 and 1405 students attending. The student sample consisted of 52.5% male and 47.5% female. With respect to race, 40% of students identified themselves as White, 34% as Hispanic, 16.5% as Black, 5% as Multiracial, and 4.5% as Asian-American. Slightly over half of the study population was on free and reduced lunch status (55%). Students' free and reduced lunch status was used as an indication of socioeconomic status (SES) of the students (Nicholson, Slater, Chriqui, & Chaloupka, 2014). The study was reviewed by the University's Institutional Review Board and the school district's review board. No incentives were given to participants for their participation.

Procedure

The staff of the NFPO trained elective teachers (e.g., art, band, and foreign language) to implement LST in the participating middle schools. The elective teachers were required to attend a 2-day workshop where they became familiar with the structure, content, and goals of the LST program. The elective teachers were given a manual that had detailed lesson plans, as well as the goals and objectives for each lesson. Lessons ranged from 45 to 50 min in length. Beginning in the fall semester, teachers delivered one LST lesson every school day for 3 to 4 weeks. The first level of LST was designed to be implemented over a two-month time span or longer (Botvin, Baker, et al., 1995; Botvin, Schinke, Epstein, Diaz, & Botvin, 1995). Sixth-grade students in the school district were only enrolled in one elective class in the fall semester; thus, the students experienced no overlap of LST.

The funding agency responsible for facilitating LST required the NFPO to conduct observations and fidelity checks of the program in order to assess the quality of implementation of the participating teachers and classrooms. The funding agency also was responsible for training the NFPO's staff and administered the training via a Skype conference call as well as provided technical assistance and conducted a site visit during the implementation of LST. In total, eight NFPO staff were trained on how to utilize the

fidelity checklist. The individuals who performed the observations and completed the fidelity checks were also required to attend an initial teacher training in the LST curriculum that was held at a participating school. The observers coordinated with the program champions (e.g., administrators or principals) of each school to establish the teachers' periods and schedules when they would be delivering LST. Although the teachers were aware the observers would be coming, they did not know the exact day they would be assessed.

Sixteen middle schools were randomly chosen from the 48 schools implementing LST in the school district. Each elective teacher conducting LST in the selected schools was observed one time. Between 2 and 12 observations and fidelity checks took place at each school based on the number of teachers delivering LST in that school. Due to random selection of one lesson per teacher by the observer, the types of lessons observed varied by teacher. The school district was responsible for ensuring consent for the project with teachers, parents, and students, disseminating and matching the students' pretest and posttest measures, and for de-identifying and matching fidelity checks to each classroom in order to make comparisons.

Measures

Level 1 Variables Seven individual-level student variables were included in the multilevel analysis: (1) gender (male = 0, female = 1); (2) race/ethnicity (1 = White, 2 = Black, 3 = Hispanic, 4 = Asian, 5 = Multiracial); and (3) whether students were or were not on free and reduced lunch status (i.e., SES) (0 = No, 1 = Yes).

Level 2 Variables The Botvin Life Skills Training Fidelity Checklist—Middle School Level 1—was used to assess fidelity of implementation at the classroom level. This checklist is specifically used to observe lessons delivered in 6th and 7th grades. The checklist assists with determining whether teachers are delivering the program content adequately and are utilizing the proper materials given to them in their training (Botvin LifeSkills® Training, n.d.). The checklist has been used in past evaluations and program replications of LST, with inter-rater reliability ranging from .80 to .90 (Botvin, Baker, et al., 1995; Botvin, Baker, Dusenbury, Tortu, & Botvin; 1990; Botvin et al., 1989; Hahn, Noland, Rayens, & Christie, 2002; Mihalic et al., 2008).

The checklist consists of 15 sections. Each section represents its own topic/lesson as well as the objectives and activities that should be present in the lesson (e.g., making decisions). For example, to assess *adherence*, the observers check yes/no on the multiple items that should be included in the lesson, indicating if the item was present or not. The numbers of items to assess adherence in each lesson ranged from 12 to 29 items. Attached to the checklist is a form that includes multiple items assessing participant responsiveness

and quality of delivery. To assess *participant responsiveness*, the observers rated how well the students understood (e.g., could answer questions about the lesson) and engaged (e.g., participated in discussion) in the lesson using a five-point Likert scale ranging from low (1) to high (5). Observers were trained to rate the classroom as a whole for participant responsiveness. To assess the *quality of delivery*, the observers rated the different attributes of the teacher's delivery of LST (e.g., how clear were the instructions given and to what extent did the presentation of materials seemed rushed or hurried). Observers were also required to rate the teachers' implementation qualities (e.g., level of enthusiasm, knowledge of program/lesson content, and effectively addressing questions or concerns) on a five-point Likert scale ranging from 1 (poor) to 5 (excellent). Higher scores on the checklists indicated better implementation fidelity.

The intraclass correlation coefficient (ICC) was used to examine the extent of measurement consistency by the same raters (i.e., intra-rater reliability) on adherence, participant responsiveness, and quality of delivery of the fidelity checklist. ICC estimates were based on a mean-rating ($k=8$), consistent, two-way mixed-effects model and were found to be highly reliable (Trevethan, 2017). The average value of the ICC for *adherence*, including all lessons, was .85; with individual lesson ICCs ranging from (lowest) .82, with a confidence interval from .55 to .96, to (highest) .95, with a confidence interval from .84 to .99. The ICC for *participant responsiveness* was .90, with a confidence interval from .90 to .91. For quality of delivery, the ICC was .94, with a confidence interval from .93 to .94.

Adherence was represented by adding the total number of Yes's and the total number of No's (recorded by the observer) and dividing that by the total number of items answered by the observer in the specific lesson of the fidelity checklist. *Participant responsiveness* was represented by the mean sums of questions that addressed student engagement and understanding, and *quality of delivery* was represented by the mean sums of questions that addressed overall quality of delivery (e.g., time management and use of relevant examples) on the fidelity checklist measure.

Outcome Variable Student SEL outcomes (i.e., social, emotional, and behavioral) can be defined as increases or decreases in certain behaviors related to the program curriculum (e.g., communication with others, conflict resolution, and self-management). To assess student SEL outcomes, the Life Skills Training Pretest/Posttest Measure (LSTM) was adapted from the original Life Skills Training Questionnaire—Middle School version (Botvin et al., 1997; Botvin, Schinke, Epstein, & Diaz, 1994). The LSTM is a 40-item self-report questionnaire that is divided into two sections that assessed knowledge and behavior and is completed by the students. For the current study, the 29 questions of the behavior section of the instrument were utilized for the

multilevel analyses. The first 17 questions of the behavior section were answered on a four-point Likert scale assessing how often or how likely students would engage in a behavior (e.g., "I am comfortable giving compliments to others"). The last 12 questions were answered on a four-point Likert scale assessing how much students agreed or disagreed with a statement (e.g., "It is easy for me to make friends").

The behavior section was found to be highly reliable (29 items; $\alpha = .86$). Higher student scores on the LSTM indicated better SEL skills. Student total outcomes were represented by the total mean sums of the students' posttest measures.

Analytic Approach

Two techniques were used to address the objectives of the study. First, descriptive statistics were used to examine frequency distribution of the variables (Table 1). Second, multilevel analysis, also known as multilevel modeling or hierarchical linear modeling (HLM), was utilized to assess the effect of individual (gender, race/ethnicity, SES) and classroom characteristics (adherence, participant responsiveness, quality of delivery) on student SEL outcomes measured at posttest. This method was chosen because the data used for this study had a cluster structure (i.e., students were nested within classrooms and classrooms were nested within schools) and it allowed for individual (i.e., within-persons) and contextual (i.e., between classrooms) variations. In order to address the data related to classrooms nested within schools (i.e., nested structure), a *school* clustering variable was included in the model. Standard errors were then computed in order to address the non-independence of observations due to the school clustering variable (Muthén & Muthén, 1998–2017).

Table 1 Frequency distributions of the independent variables

Individual level	<i>N</i>	%	<i>M</i>	<i>SD</i>
Gender				
Male	2527	52.5		
Female	2285	47.5		
Race/ethnicity				
White	1930	40.1		
Black	794	16.5		
Hispanic	1622	33.7		
Asian	217	4.5		
Multiracial	242	5		
Free and reduced lunch status	2662	55		
Classroom level				
Adherence			1.78	.22
Participant responsiveness			10.58	3.00
Quality of delivery			39.30	10.43

Data analyses were carried out in two steps. First, bivariate conditional models with one covariate at level 2 and all predictors at level 1 were examined. Secondly, a multivariate model with all predictors at level 2 was estimated. Mplus statistical software v.8.0 was used to carry out the multilevel analyses (Muthén & Muthén, 1998–2017).

Missing Data Little's Missing Completely at Random (MCAR) test was conducted and was significant, indicating that missing values of study variables are *not* missing completely at random, $\chi^2(30, N=4812) = 1955.47, p < .001$. Thus, multiple imputation was used to estimate missing values, as this approach is recommended in the extant literature (Graham, 2012). Specifically, missing data were handled by the Mplus estimation procedure (Muthén & Muthén, 1998–2017), which handles missing data through full information maximum likelihood (FIML) accounting for missing at random (MAR) assumptions (Arbuckle, 1996).

Results

Multilevel Analysis

Level 1 Among the individual variables, both race/ethnicity and SES were found to be statistically significant predictors of students' posttest outcomes. Specifically, students who identified as Black and Hispanic on average scored lower than students who identified as White on the LSTM (White students were used as the reference category). Asian and Multiracial race categories and gender were not significant predictors. SES (as represented by students free and reduced lunch status) was a significant predictor of students' posttest outcomes and on average, students who indicated they were

on free and reduced lunch status scored lower than students who indicated they were not on the LSTM (Table 2).

Level 2 When examining classroom variables individually, participant responsiveness and quality of delivery were found to be statistically significant predictors of students' posttest outcomes. Adherence was not found to be a statistically significant predictor. Results also indicated that a significant association between adherence, participant responsiveness, and quality of delivery with students who identified as Black and Hispanic and being on free and reduced lunch status was present. Table 2 depicts the estimate of the intercepts and standard errors at the between level for all classroom-level predictors.

In the last step, all covariates on the classroom level were included in the model. Results indicated that participant responsiveness was a statistically significant predictor of students' posttest outcomes ($\beta = .37, SE = .16, p < .05$). In other words, higher scores given by the observers on the participant responsiveness items of the fidelity checklist predicted higher rates of students' SEL skills at posttest. Adherence and quality of delivery were not significant predictors of students' posttest outcomes when all covariates were included in the model (Table 3). The standardized beta coefficient (β) was used to represent effect size. Results indicate what could be considered a large effect size for participant responsiveness ($\beta = .53$) (Table 3).

Discussion

This preliminary study examined the effects of individual-level (e.g., race/ethnicity) and classroom-level (e.g., adherence) predictors on student SEL skills after completing the

Table 2 Parameter estimates and standard errors for bivariate model

Level 1						
(Within level)	Adherence		Participant responsiveness		Quality of delivery	
	<i>B</i>	SE	<i>B</i>	SE	<i>B</i>	SE
Gender	0.37	0.38	0.37	0.38	0.36	0.38
Race/ethnicity**						
Black	−2.37*	0.53	−2.45*	0.52	−2.48*	0.53
Hispanic	−2.40*	0.56	−2.36*	0.56	−2.37*	0.56
Asian	1.30	0.72	1.32	0.71	1.28	0.72
Multiracial	−0.50	0.54	−0.48	0.55	−0.49	0.55
Free and reduced lunch status	−3.70*	0.32	−3.63*	0.32	−3.67*	0.33
Level 2						
(Between level)	<i>B</i>	SE	<i>B</i>	SE	<i>B</i>	SE
	0.98	1.22	0.26*	0.05	0.05*	0.02

* $p < 0.05$. **White was used as a reference category

Table 3 Parameter estimates and standard errors for final multivariate model

(Within level)		Level 1								
		<i>B</i>	SE	β						
Gender		0.39	0.38	0.04						
Race/ethnicity**										
Black		−2.36*	0.52	−0.22						
Hispanic		−2.40*	0.55	−0.22						
Asian		1.31	0.71	0.12						
Multiracial		−0.50	0.55	−0.05						
Free and reduced lunch status		−3.62*	0.32	−0.33						
(Between level)		Level 2								
		Adherence			Participant responsiveness			Quality of delivery		
		<i>B</i>	SE	β	<i>B</i>	SE	β	<i>B</i>	SE	β
		−0.96	1.26	−0.10	0.37*	0.16	0.53	−0.03	0.04	−0.17

* $p < 0.05$. **White was used as a reference category

LST program. To date, few studies have examined the less common elements of implementation fidelity (e.g., participant responsiveness and quality of delivery) and how they can affect student SEL outcomes (Berkel et al., 2011; Reyes et al., 2012). Although SEL approaches have proven to be effective in school-based settings, previous researchers have noted that there has been a lack of emphasis on evaluating implementation, even though it can be an essential component related to achieving positive outcomes (Low, Smolkowski, & Cook, 2016).

Past research on demographic characteristics affecting student SEL outcomes in SEL programs have been mixed. Several researchers have found that Black, Hispanic, and lower SES students are likely to be at higher risk for experiencing poorer social and emotional outcomes when compared to White students (Castro-Olivo, 2014; Garner, Mahatma, Brown, & Vesley, 2014). While in a study evaluating the effectiveness of a conflict resolution program, Black students' prosocial behavior increased at a slower rate than White students; however, no significant differences were found between Hispanic and White students (Aber, Brown, & Jones, 2003). Frequently, program effects are not disaggregated by demographic characteristics, which can make determining if a SEL program can yield equal gains among its participants problematic (Garner et al., 2014). The results of the current study indicated student race/ethnicity and SES (i.e., free and reduce lunch status) were significantly associated with student outcomes. Black, Hispanic, and lower SES students consistently scored lower when compared to White students when examining each core element separately, as well as when all elements were included in the multilevel model. This suggests adaptations may need to be made to LST components and how teachers are trained and

delivering lessons in order to address these students' social and emotional learning needs. This is consistent with the literature stating that tailoring LST interventions to a specific population and utilizing culturally adapted techniques were more beneficial for minority youth than a generic approach (Botvin, Schinke, et al., 1995).

When the core elements of fidelity were evaluated separately, participant responsiveness and quality of delivery were significantly associated with student SEL outcomes. However, when all individual- and classroom-level variables were included in the model, only participant responsiveness was significantly associated with student outcomes. This finding is unique, as the existing literature suggests that adherence is typically found to be the most common predictor of student outcomes (Durlak, 2016). The findings of adherence in the current study could be attributed to a ceiling effect where teachers were scored very positively on nearly all adherence items (Taylor, 2012).

Collectively, these findings may suggest that participant involvement and engagement may be an important and under-recognized element in assuring fidelity. While past research has established the importance of adherence to program protocol, the current research suggests that the engagement of students is also important. School professionals working with students may need to work actively to involve and engage students by using interactive techniques such as skill demonstration (e.g., giving a compliment to a peer) or group discussion (Domitrovich et al., 2010; Mihalic et al., 2008). It cannot be presumed that passive compliance will produce desired outcomes. One other study (Low et al., 2016) had findings consistent with the current study in that high or low "engagement" can be a significant component to the implementation of SEL programs, with low engagement

being significantly associated with poorer outcomes. However, it should be noted their study focused on an elementary SEL program and implementation fidelity was measured by teacher self-reports. Other studies have also shown that participant responsiveness can be a more reliable predictor of student and programmatic outcomes than adherence and suggest that this core element could be emphasized in trainings and ongoing technical assistance (Humphrey, Barlow, & Lendrum, 2018; Pettigrew et al., 2015). However, this may pose a challenge to programs that are more manualized and procedural in nature (Humphrey et al., 2018). In terms of effect size, the literature seems to be inconsistent when compared to the current study. For example, some studies found small-to-medium effect sizes for student engagement (Low, Cook, Smolkowski, Buntain-Ricklefs, 2015; Low et al., 2016) and another study found no significant results related to student participation (Mihalic et al., 2008). Additional research is needed to more fully examine the potential impact of participant responsiveness on student outcomes for the LST program.

Strengths and Limitations

In order to interpret results of the present study, strengths and limitations should be noted. This study was able to disaggregate student outcomes by demographic characteristics, which has been a noted barrier in determining gains among students with diverse demographic characteristics. This is critical due to the current study suggesting that modification of program materials and/or teacher delivery may be necessary in order to obtain significant outcomes from specific groups of students. In addition, direct observers were utilized to conduct the observations in order to assist with limiting rater bias and assist with accuracy (Allen et al., 2018).

The LSTM was highly reliable ($\alpha = .86$). However, this student outcome measure is based solely on student report. Therefore, the results may be biased in terms of how accurate and truthful the student responses were (Schwartz & Beaver, 2015). The anonymity of a self-report measure may lead to a reduction in accountability and diminish student motivation to respond to items conscientiously (Jia, Konold, Cornell, & Huang, 2018). Future research would benefit from also collecting data from teachers and/or parents in order to compare and potentially validate student outcomes (Schwartz & Beaver, 2015).

Although the fidelity checklist has been widely used in past studies, a large portion of the checklist is completed on a yes or no basis. This may have limited the ability of the measure to sufficiently capture variation in teachers' delivery of specific program components (i.e., adherence). Due to lack of resources, only one observation took place per classroom and may fail to be a complete representation of implementation fidelity. As a result of only conducting one

observation per classroom, inter-rater reliability was also impossible to assess, which could have threatened the validity of this study. To address this limitation, ICCs were computed as a measure of intra-rater reliability for adherence, participant responsiveness, and quality of delivery sections of the fidelity checklists. The observers were found to be highly reliable as a group. However, future research would potentially benefit from collecting a sufficient amount of data (i.e., teacher self-report and multiple external observations per classroom), providing a detailed description of the observers' training, and the extent of the reliability appraisal in order to provide a comprehensive assessment of inter-rater reliability.

Finally, dosage and program differentiation were not included in the analysis of the current study because there were no data available to accurately depict those specific elements. The study was also limited by the timeframe, which prevented the assessment of program dosage and other student outcomes. Future research could benefit from assessing length of the lessons, if all lessons were completed, and the extent to which each program component was covered. In addition, it is recommended that researchers employ follow-up measures to examine long-term social, emotional, and behavioral outcomes. This may allow researchers to identify if a program (e.g., LST) and its implementation (e.g., adherence, dosage, etc.) are related to certain behaviors (e.g., reduction in drug use) over time among a specific cohort (Botvin, Baker et al., 1995; Botvin & Griffin, 2004, 2015).

Implications for School Mental Health Research and Practice

Results from this study suggest that participant responsiveness may be a key element to implementation fidelity and can be linked to what fits the students' experiences. While delivering LST in schools, it may be important to ensure students are actively engaged, understand the material, and are participating in the lesson. This finding is consistent with social learning theory, which states that social interactions, such as verbal instruction, role modeling, support, and feedback, are germane in obtaining new behavior (Bandura, Adams, & Beyer, 1977; Brackett et al., 2015). To improve fidelity of LST implementation and enhance outcomes for students, teachers may need to modify their delivery methods and the examples used to explain a topic in order for it to be applicable to the students within their classrooms. Findings related to participant responsiveness may also help to account for lower gains among students with diverse demographic characteristics (i.e., race/ethnicity and SES). For example, if these students are less engaged, they may be less likely to benefit from an otherwise valuable learning experience.

Teacher demographics (e.g., age, race/ethnicity, or level of education) may also be an important consideration for implementation fidelity and student outcomes. The results of a study assessing individual factors (i.e., teacher demographics) and how they affect the implementation of the PAX Good Behavior Game indicated that “younger” teachers (i.e., under the age of 30) had significantly higher implementation quality when compared to “older” teachers (i.e., over the age of 30) (Domitrovich et al., 2015). In addition, a study by Williford, Wolcott, Whittaker, and Locasale-Crouch (2015) found that teachers who were of minority ethnicity engaged in the Banking Time program less often and with lower quality when compared to White, non-Hispanic teachers. It should be noted, however, that while both of these studies looked at aspects of implementation fidelity (i.e., dosage and quality of delivery), neither of them examined if teacher demographics are related to participant responsiveness. Future research is warranted to further examine teacher demographics and their potential relationship with participant responsiveness.

Further, Mihalic et al. (2008) examined quality of delivery by looking at teaching techniques (e.g., discussion, skill demonstration, and behavioral rehearsal). The results indicated teachers’ use of interactive techniques has been shown to be positively correlated with good student behavior, and therefore, greater knowledge and behavior acquisition. It is reasonable to expect that high-quality delivery of SEL programs such as LST can potentially mediate student behavior and lead to more active student engagement, a better understanding of the material, and increased responsiveness during a lesson. Additional research is needed to determine how to best navigate the challenges associated with attaining participant responsiveness and the relationship between participant responsiveness and quality of delivery, especially when working with minority and low-socioeconomic students. High-quality delivery within these populations may be particularly important (Berkel et al., 2011) due to program engagement being influenced by how receptive an intervention is to cultural characteristics (Dillman Carpentier et al., 2007; Prado, Pantin, Schwartz, Lupei, & Szapocznik, 2006).

Conclusions

The findings of this preliminary study may offer useful results regarding future directions for the implementation of school-based SEL programs. Future research would benefit from examining all feasible core elements of implementation fidelity and how they affect student outcomes in LST programs. It is suggested the five most common elements should be assessed in ordered to accurately capture a comprehensive picture of the implementation process and examine the relationship between

the different elements and their influence on student SEL outcomes (Carroll et al., 2007).

Adherence and dosage seem to be the most common focus of studies that examine implementation fidelity of SEL programs, but as this study suggests, other elements may also be critical for student outcomes. Participant responsiveness may be particularly important for student populations with diverse demographic characteristics. Similarly, program differentiation and quality of delivery are rarely measured. Assessing what makes a program unique and what delivery methods are essential to obtain classroom cohesion are not only essential for positive outcomes, but sustainability as well.

The *Every Child Succeeds Act* (2015) has created intense pressure for teachers and schools to ensure strong academic performance from their students, utilize evidence-based practices, and emphasize SEL standards (Reyes et al., 2012). However, it is important that ongoing education and coaching for teachers and schools and consideration for implementation fidelity are also included in federal and state legislation. The adoption of widespread initiatives and effective prevention programs will have minimal effect until the quality of implementation by schools and teachers can be established and positive student outcomes are achieved (Mihalic et al., 2008). In addition, it is important to also assess the reliability (i.e., inter-rater) of the measurement of fidelity to ensure accuracy and provide a comprehensive appraisal of the implementation process. Poor implementation can result in large economic losses for schools and can make it difficult to meaningfully and accurately interpret student outcomes. Monitoring program implementation can assist schools with obtaining better program outcomes, sustainability, and advocating for policies and guidelines that could bring effective programs into their districts (Damschroder et al., 2009; Durlak, 2016).

Compliance with Ethical Standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical Approval All procedures were in accordance with the ethical standards of the institutional review boards and with the 1964 Declaration of Helsinki and its later amendments or comparable ethical standards. For this type of study, formal consent was not required.

Informed Consent Informed consent was obtained from all individual participants included in the study.

References

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.
- Allen, J. D., Linnan, L. A., & Emmons, K. M. (2012). Fidelity and its relationship to implementation effectiveness, adaptation, and dissemination. In R. C. Brownson, G. A. Colditz, & E. K. Proctor (Eds.), *Dissemination and implementation research in health: Translating science to practice* (pp. 281–305). New York: Oxford University Press.
- Allen, J. D., Shelton, R. C., Emmons, K. M., & Linnan, L. A. (2018). Fidelity and its relationship to implementation effectiveness, adaptation, and dissemination. In R. C. Brownson, G. A. Colditz, & E. K. Proctor (Eds.), *Dissemination and implementation research in health: Translating science to practice* (2nd ed., pp. 267–284). New York: Oxford University Press.
- Arbuckle, J. (1996). Full information estimation in the presence of incomplete data. In G. Marcoulides & R. Schumacker (Eds.), *Advanced structural equation modeling: Issues and techniques* (pp. 243–277). New York: Psychology Press.
- Bandura, A., Adams, N. E., & Beyer, J. (1977). Cognitive processes mediating behavioral change. *Journal of Personality and Social Psychology*, 35(3), 125–139.
- Berkel, C., Mauricio, A. M., Schoenfelder, E., & Sandler, I. N. (2011). Putting the pieces together: An integrated model of program implementation. *Prevention Science*, 12(1), 23–33.
- Blueprints for Healthy Youth Development [BHYD]. (n.d.). *LifeSkills Training (LST)*. Retrieved November 8, 2018, from <http://www.blueprintsprograms.com/factsheet/lifeskills-training-lst>.
- 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.
- Botvin, G. J., Baker, E., Dusenbury, L., Tortu, S., & Botvin, E. M. (1990). 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., Batson, H. W., Witts-Vitale, S., Bess, V., Baker, E., & Dusenbury, D. (1989). A psychological approach to smoking prevention for urban black youth. *Public Health Reports*, 104(6), 573–582.
- Botvin, G. J., Epstein, J. A., Baker, E., Diaz, T., & Ifill-Williams, M. (1997). School-based drug abuse prevention with inner-city minority youth. *Journal of Child and Adolescent Substance Abuse*, 6(1), 5–19.
- Botvin, G. J., & Griffin, K. W. (2004). Life skills training: Empirical findings and future directions. *Journal of Primary Prevention*, 25(4), 211–232.
- Botvin, G. J., & Griffin, K. W. (2015). Life skills training: A competence enhancement approach to tobacco, alcohol, and drug abuse prevention. In L. M. Scheier (Ed.), *Handbook of adolescent drug use prevention: Research, intervention strategies, and practice* (pp. 177–196). Washington, DC: American Psychological Association.
- Botvin, G. J., Griffin, K. W., Diaz, T., & Ifill-Williams, M. (2001a). Drug abuse prevention among minority adolescents: Posttest and one-year follow-up of a school based preventive intervention. *Prevention Science*, 2(1), 1–13.
- Botvin, G. J., Griffin, K. W., Diaz, T., & Ifill-Williams, M. (2001b). Preventing binge drinking during early adolescence: One- and two-year follow-up of a school-based preventive intervention. *Psychology of Addictive Behaviors*, 15(4), 360–365.
- Botvin, G. J., Griffin, K. W., & Nichols, T. D. (2006). Preventing youth violence and delinquency through a universal school-based prevention approach. *Prevention Science*, 7(4), 403–408.
- Botvin, G. J., Schinke, S. P., Epstein, J. A., & Diaz, T. (1994). Effectiveness of culturally focused and generic skills training approaches to alcohol and drug abuse prevention among minority youths. *Psychology of Addictive Behaviors*, 8(2), 116–127.
- Botvin, G. J., Schinke, S. P., Epstein, J. A., Diaz, T., & Botvin, E. M. (1995). Effectiveness of culturally focused and generic skills training approaches to alcohol and drug abuse prevention among minority adolescents: Two-year follow-up results. *Psychology of Addictive Behaviors*, 9(3), 183–194.
- Botvin LifeSkills® Training. (n.d.). *LST fidelity checklists*. Retrieved October 11, 2018, from <https://www.lifeskillstraining.com/lst-fidelity-checklists/>.
- Brackett, M. A., Elbertson, N. A., & Rivers, S. E. (2015). Applying theory to the development of approaches to SEL. In J. A. Durlak, C. E. Domitrovich, R. P. Weissberg, & T. P. Gullotta (Eds.), *Handbook of social and emotional learning: Research and practice* (pp. 20–32). New York: The Guilford Press.
- Carroll, C., Patterson, M., Wood, S., Booth, A., Rick, J., & Balain, S. (2007). A conceptual framework for implementation fidelity. *Implementation Science*, 2(40), 1–9.
- Castro-Olivo, S. M. (2014). Promoting social-emotional learning in adolescent Latino ELLs: A study of the culturally adapted Strong Teens program. *School Psychology Quarterly*, 29(4), 567–577.
- Collaborative for Academic, Social, and Emotional Learning. (n.d.). *What is SEL?* Retrieved October 1, 2018, from <http://www.casel.org/what-is-sel/>.
- Damschroder, L. J., Aron, D. C., Keith, R. E., Kirsh, S. R., Alexander, J. A., & Lowery, J. C. (2009). Fostering implementation of health services research finding into practice: A consolidated framework for advancing implementation science. *Implementation Science*, 4(50), 1–15.
- 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(1), 23–45.
- Dillman Carpentier, F., Mauricio, A. M., Gonzales, N. A., Millsap, R. E., Meza, C. M., Dumka, L. E., et al. (2007). Engaging Mexican origin families in a school-based preventive intervention. *Journal of Primary Prevention*, 28(6), 521–546.
- Domitrovich, C. E., Gest, S. D., Jones, D., Gill, S., & Sanford DeRoussie, R. M. (2010). Implementation quality: Lessons learned in the context of the head start REDI trial. *Early Childhood Research Quarterly*, 25(3), 284–298.
- Domitrovich, C. E., Pas, E. T., Bradshaw, C. P., Becker, K. D., Keperling, J. P., Embry, D. D., et al. (2015). Individual and school organizational factors that influence implementation of the PAX good behavior game intervention. *Prevention Science*, 16(8), 1064–1074.
- Durlak, J. A. (2015). What everyone should know about implementation. In J. A. Durlak, C. E. Domitrovich, R. P. Weissberg, & T. P. Gullotta (Eds.), *Handbook of social and emotional learning: Research and practice* (pp. 395–405). New York: The Guilford Press.
- Durlak, J. A. (2016). Programme implementation in social and emotional learning: Basic issues and research findings. *Cambridge Journal of Education*, 46(3), 333–345.
- Durlak, J. A., & DuPre, E. P. (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(3–4), 327–350.
- Durlak, J. A., Weissberg, R. P., Dymnicki, A. B., Taylor, R. D., & Schellinger, K. B. (2011). The impact of enhancing students' social and emotional learning: A meta-analysis of school-based universal interventions. *Child Development*, 82(1), 405–432.
- 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.

- Fixsen, D. L., & Blase, K. A. (n.d.) *Implementation and system change* [PowerPoint Presentation]. Chapel Hill, NC: University of North Carolina, National Implementation Research Network, FPG Child Development Institute.
- Garner, P. W., Mahatmya, D., Brown, E. L., & Vesley, C. K. (2014). Promoting desirable outcomes among culturally and ethnically diverse children in social emotional learning programs: A multilevel heuristic model. *Educational Psychology Review*, *26*(1), 165–189.
- Goncy, E. A., Sutherland, K. S., Farrell, A. D., Sullivan, T. N., & Doyle, S. T. (2015). Measuring teacher implementation in delivery of a bullying prevention program: The impact of instructional and procedural adherence and competence on student responsiveness. *Prevention Science*, *16*(3), 440–450.
- Graham, J. W. (2012). *Missing data: Analysis and design*. New York: Springer.
- Greenberg, M. T. (2004). Current and future challenges in school-based prevention: The researcher perspective. *Prevention Science*, *5*(1), 5–13.
- Greenberg, M. T. (2010). School-based prevention: Current status and future challenges. *Effective Education*, *2*(1), 27–52.
- 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.
- Humphrey, N., Barlow, A., & Lendrum, A. (2018). Quality matters: Implementation moderates student outcomes in the PATHS curriculum. *Prevention Science*, *19*(2), 197–208.
- Jagers, R. J., Harris, A., & Skoog, A. (2015). A review of classroom-based SEL programs at the middle school level. In J. A. Durlak, C. E. Domitrovich, R. P. Weissberg, & T. P. Gullotta (Eds.), *Handbook of social and emotional learning: Research and practice* (pp. 167–180). New York: The Guilford Press.
- Jia, Y., Konold, T. R., Cornell, D., & Huang, F. (2018). The impact of validity screening on associations between self-reports of bullying victimization and student outcomes. *Educational and Psychological Measurement*, *78*(1), 80–102.
- Klapp, A., Belfield, C., Bowden, B., Levin, H., Shand, R., & Zander, S. (2017). A benefit-cost analysis of a long-term intervention on social and emotional learning in compulsory school. *The International Journal of Emotional Education*, *9*(1), 3–19.
- Low, S., Cook, C., Smolkowski, K., & Burtain-Ricklefs, J. (2015). Promoting social-emotional competence: An evaluation of the elementary version of Second Step®. *Journal of School Psychology*, *53*(6), 463–477.
- Low, S., Smolkowski, K., & Cook, C. (2016). What constitutes high-quality implementation of SEL programs? A latent class analysis of Second Step® implementation. *Prevention Science*, *17*(8), 981–991.
- Meyers, A. B., Tobin, R. M., Huber, B. J., Conway, D. E., & Shelvin, K. H. (2015). Interdisciplinary collaboration supporting social-emotional learning in rural school systems. *Journal of Educational and Psychological Consultation*, *25*(2–3), 109–128.
- Mihalic, S. (2002). *The importance of implementation fidelity*. Boulder, CO: University of Colorado, Blueprints Violence Prevention Initiative, Center for the Study and Prevention of Violence. Retrieved from <http://www.incredibleyears.com/wp-content/uploads/fidelity-importance.pdf>.
- Mihalic, S. F., & Elliot, D. S. (2015). Evidence-based programs registry: Blueprints for healthy youth development. *Evaluation and Program Planning*, *48*, 124–131.
- Mihalic, S. F., Fagan, A. A., & Argamaso, S. (2008). Implementing the life skills training drug prevention program: Factors related to implementation fidelity. *Implementation Science*, *3*(5), 1–16.
- Muthén, L. K., & Muthén, B. O. (1998–2017). *Mplus User's Guide*. (8th ed.). Los Angeles, CA: Muthén & Muthén.
- Nicholson, L. M., Slater, S. J., Chriqui, J. F., & Chaloupka, F. (2014). Validating adolescent socioeconomic status: Comparing school free or reduced price lunch with community measures. *Spatial Demography*, *2*(1), 55–65.
- Pettigrew, J., Graham, J. W., Miller-Day, M., Hecht, M. L., Krieger, J. L., & Shin, Y. J. (2015). Adherence and delivery: Implementation quality and program outcomes for the seventh-grade keepin' it REAL program. *Prevention Science*, *16*(1), 90–99.
- Prado, G., Pantin, H., Schwartz, S. J., Lupei, N. S., & Szapocznik, J. (2006). Predictors of engagement and retention into a parent-centered ecodevelopmental HIV preventive intervention for Hispanic adolescents and their families. *Journal of Pediatric Psychology*, *31*(9), 874–890.
- Reyes, M. R., Brackett, M. A., Rivers, S. E., Elbertson, N. A., & Salovey, P. (2012). The interaction effects of program training, dosage, and implementation quality on targeted student outcomes for the RULER approach to social and emotional learning. *School Psychology Review*, *41*(1), 82–99.
- Schwartz, J. A., & Beaver, K. M. (2015). Making (up) the grade? Estimating the genetic and environmental influences on discrepancies between self-reported grades and official GPA scores. *Journal of Youth and Adolescence*, *44*(5), 1125–1138.
- Sutherland, K. S., Conroy, M. A., McLeod, B. D., Algina, J., & Kune-mund, R. L. (2018). Factors associated with teacher delivery of a classroom-based tier 2 prevention program. *Prevention Science*, *19*(2), 186–196.
- Taylor, T. H. (2012). Ceiling effect. In N. J. Salkind (Ed.), *Encyclopedia of research design* (pp. 133–134). Thousand Oaks, CA: Sage Publications.
- Trevethan, R. (2017). Intraclass correlation coefficients: Clearing the air, extending some cautions, and making some requests. *Health Services and Outcomes Research Methodology*, *17*(2), 127–143.
- Velasco, V., Griffin, K. W., Botvin, G. J., Celata, C., & Lombardia, Gruppo L. S. T. (2017). Preventing adolescent substance use through an evidence-based program: Effects of the Italian adaptation of life skills training. *Prevention Science*, *18*(4), 394–405.
- Weissberg, R. P., Durlak, J. A., Domitrovich, C. E., & Gullotta, T. P. (2015). Social and emotional learning: Past, present, and future. In J. A. Durlak, C. E. Domitrovich, R. P. Weissberg, & T. P. Gullotta (Eds.), *Handbook of social and emotional learning: Research and practice* (pp. 3–19). New York: The Guilford Press.
- Williford, A. P., Wolcott, C. S., Whittaker, J. V., & Locasale-Crouch, J. (2015). Program and teacher characteristics predicting the implementation of Banking Time with preschoolers who display disruptive behaviors. *Prevention Science*, *16*(8), 1054–1063.
- Wright, A., Lamont, A., Wandersman, A., Osher, D., & Gordon, E. S. (2015). Accountability and SEL programs: The Getting To Outcomes® approach. In J. A. Durlak, C. E. Domitrovich, R. P. Weissberg, & T. P. Gullotta (Eds.), *Handbook of social and emotional learning: Research and practice* (pp. 500–515). New York: The Guilford Press.
- Zinsser, K. M., & Dusenbury, L. (2015). Recommendations for implementing the new Illinois early learning and development standards to affect classroom practices for social and emotional learning. *Early Childhood Research and Practice*, *17*(1), 1–8.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.