School Mental Health
https://doi.org/10.1007/s12310-020-09392-9

ORIGINAL PAPER

Twice as Nice? Sustained Exposure to a Universal Social–Emotional Learning Program Across Multiple Grades

Leah J. Hunter[1] · James C. DiPerna[1] · Weiyi Cheng[1] · Puiwa Lei[1] · Susan Crandall Hart[1]

Accepted: 1 September 2020

Abstract

Although universal social–emotional learning programs are increasingly implemented across the USA, few studies have examined the effects of sustained exposure to the same program across multiple grade levels. As such, the goal of this study was to isolate the effects of sustained exposure to a universal social–emotional learning program (Elliott & Gresham, Social skills improvement system: classwide intervention program [SSIS-CIP]. NCS Pearson, Bloomington, MN, 2007), on elementary students' social, behavioral, and academic outcomes. Compared to students who experienced the SSIS-CIP in second grade only ($N=218$), students exposed to the SSIS-CIP in first and second grade ($N=181$) showed further gains in several social skill domain areas as well as academic engagement at the end of second grade. Interactions, however, indicated that some effects were potentially moderated by student or class-level skills at the beginning of second grade. Future research considerations and practical implications for universal social–emotional learning programming are discussed.

Keywords Social–emotional learning · Intervention · Dosage · Longitudinal research

Introduction

Research on school-based social–emotional learning (SEL) has shown that early acquisition of prosocial skills improves later learning and adjustment outcomes (Durlak, Weissberg, Dymnicki, Taylor, & Schellinger, 2011; Taylor, Oberle, Durlak, & Weissberg, 2017). Moreover, the explicit teaching of social skills results in positive long-term gains over and above the absence of negative behavior alone (Caprara, Barbaranelli, Pastorelli, Bandura, & Zimbardo, 2000; Jones, Greenberg, & Crowley, 2015). In response, schools across the USA and internationally are adopting universal SEL programs to provide a more holistic approach to education (Elbertson, Brackett, & Weissberg, 2009). The Social Skills Improvement System—Classwide Intervention Program (SSIS-CIP; Elliott & Gresham, 2007) is a universal SEL program intended to assist the development of social–emotional and learning-related behaviors.

The SSIS-CIP was developed for Grades K-8 and is intended to be implemented universally by general education teachers within the context of their daily instruction. The lower elementary version (Grades K-2) was the focus of an efficacy trial with results indicating positive effects for increasing young students' social and learning-related behaviors (DiPerna, Lei, Bellinger, & Cheng, 2015, 2016; DiPerna, Lei, Cheng, Hart, & Bellinger, 2018) and cost-effectiveness, particularly in second grade (Hunter, DiPerna, Hart, & Crowley, 2018). Because the instructional content, approach, and supporting materials are identical across grade levels within each version of the SSIS-CIP (i.e., Grades Pre-K, K-2, 3–8), an important practical question from a universal implementation standpoint is whether or not there is benefit to students being exposed to the program across successive grade levels. As such, the goal of the current study was to understand the effects of sustained exposure to the K-2 version of the SSIS-CIP in Grades 1 and 2.

✉ Leah J. Hunter
ljh222@psu.edu

James C. DiPerna
jdiperna@psu.edu

Weiyi Cheng
wuc131@gmail.com

Puiwa Lei
puiwa@psu.edu

Susan Crandall Hart
susan.hart@psu.edu

[1] Department of Educational Psychology, Counseling, and Special Education, The Pennsylvania State University, University Park, PA 16802, USA

Published online: 18 September 2020

Universal SEL Program Content

By design, school-based universal SEL interventions are intended to promote all students' prosocial skills while also reducing risk factors and future mental health difficulty (Belfield et al., 2015; Greenberg & Abenavoli, 2017). The classroom is a popular setting for universal SEL given the potential for generalization and maintenance of social skills instruction alongside academic instruction (Durlak et al., 2011). The Collaborative for Academic, Social, and Emotional Learning (CASEL) periodically reviews SEL programs for their effectiveness using rigorous criteria that defines "well-designed" programs as "classroom-based programs that systematically promote students' social and emotional competence, provide opportunities for practice, and offer multi-year programming" (CASEL, 2013, p. 4). Although explicit SEL skills instruction across multiple years is not the only component of effective SEL, it is a hallmark in programs that have shown effectiveness in controlled studies (January, Casey, & Paulson, 2011).

Despite the high level of interest in universal SEL in schools today, many available programs have not been evaluated to determine their effectiveness. Table 1 provides an overview of 11 programs that (a) focus on explicit teaching of SEL skills, (b) are intended for multi-year delivery, and (c) have been subjected to randomized controlled trials. Although they target similar outcomes, these programs vary greatly in their intensity, content, and cost. For example, the number of lessons students receive ranges from 11 to 140 per year, and the length of these lessons varies widely. Several programs provide manuals and/or curricular materials specific to each grade level offered, while the rest (designated in italics in the table) provide curricular materials in "levels" that may span several grades. Of the programs that provide content specific to each grade, detailed information concerning developmental sequencing (i.e., what makes the content at one grade level different from the next) is not readily available on the program websites, and we could find no empirical studies that review the grade-level sequencing of such curricula. Given the likelihood that schools will purchase one program for delivery school-wide, students will likely be exposed to similar (if not repeated) content at multiple grade levels. As such, schools are charged with making a difficult decision concerning selecting a program that is not only effective but also relevant to their preferred period(s) of delivery and budgetary constraints.

Implementation Issues in Universal SEL Program Delivery

The impetus to focus on issues of content and dosage stems from the field of implementation science (Durlak & DuPre, 2008). Success of any intervention largely hinges on the quality of its implementation, including factors such as training, fidelity to the original protocol, and dosage (Durlak & DuPre, 2008; Greenberg, Domitrovich, Graczyk, & Zins, 2005; Proctor et al., 2011), and universal SEL interventions are no exception. For example, large meta-analyses of SEL programs have reported small to moderate effect sizes for the promotion of prosocial and academic skills, with student outcomes moderated by implementation quality (Durlak et al., 2011; Taylor et al., 2017). Further, the positive impact appears to diminish over time, as effect sizes are much smaller at follow-up than when measured immediately following interventions (Durlak et al., 2011). It is possible that implementation issues are partly responsible for this "fade-out" effect (Rimm-Kaufman et al., 2014).

Dosage, or program intensity, is of key importance to universal SEL, given recommendations that students are exposed to programs across multiple years (CASEL, 2013; Nation et al., 2003). Indeed, dosage is a strong moderator of treatment effect in prevention programs (January et al., 2011). For example, Nelson, Westhues, and McLeod (2003) examined the length and intensity of preschool prevention programs and found that the average weighted effect sizes for longer and more intensive programs were substantially higher than for programs that were shorter and less intense. Specifically, children's social–emotional and cognitive outcomes were significantly higher when exposed to a program across multiple years (Cohen's $d = .27–.53$) relative to single-year exposure ($d = .06–.09$). Similarly, Beets et al. (2009) demonstrated that students who received 3 or more years of the Positive Action program in first through fifth grade reported 41–73% fewer experiences with substance use and violent behaviors and an 89% lower rate of engaging in voluntary sexual activity in fifth grade than did students who received less exposure to the program. January et al.'s (2011) meta-analysis specific to school-based social skills programs reported a mean dosage of 25.13 h of exposure to programs (range 3.3–90 h), and hours of exposure significantly predicted program results.

SSIS-CIP Dosage

The SSIS-CIP (Elliott & Gresham, 2007) consists of 10 units focused on skills identified by teachers as important

Table 1 Universal social-emotional learning program content and grade-level sequencing

Program name	Grade range	Target areas	Average number sessions/year	Description of "grade-by-grade sequence"
4Rs (Reading, Writing, Respect, & Resolution)	PK-8	Literacy; conflict resolution; intergroup understanding	35 period-long class sessions	Standardized teaching manual specific to each grade PK-8
Caring School Community (2nd Ed.)	K-8	Core values: prosocial behavior; school-wide feeling of community	30–35 class meetings	Teacher's Manual, booklets, & literature specific to each grade; "Caring School Discipline" teacher's guide sequenced by level (e.g., K-1; 2–5, 6–8)
I Can Problem Solve	PK-5	Problem-solving; academic skills	59–83 lessons	Teachers Manuals available for 3 levels: PreK (59 lessons), K & primary grades (83 lessons), & intermediate elementary grades (77 lessons)
Incredible Years Classroom Dinosaur Curriculum	PK-2	Social skills; problem behavior	64 lessons	Curriculum packages including lessons/activities for 3 levels: ages 3–4, 5–6, and 7–8
Promoting Alternative Thinking Strategies (PATHS)	PK-6	Social-emotional competence; learning skills; problem behavior	40–52 lessons plus extension activities	Six sets of lessons/activities: PK/K; grade 1; grade 2; grade 3; grade 4; grade 5/6
Positive Action	PK-12	Character; social-emotional skills; problem behavior	140 lessons	Separate kits for each grade PK-12 following "common units"
Resolving Conflict Creatively Program	PK-5	SEL; character development; conflict resolution; relationships	16 lessons	Separate curriculum manual for 2 levels: K-2; 3–5
Second Step	PK-8	Problem behavior; social–emotional competence; self-regulation	22–28 weekly topics across 5 days/week	Separate lessons and curricular materials for PK-8
SSIS-CIP	PK-8	Social skills; relationships, & enable academic achievement	30 lessons	Manual with 23 skill units for all grades PK-8 + student digital licenses
Steps to Respect	K-5	Bullying prevention	11 lessons + 2 literature units (7–10 lessons in each)	Available for purchase as "Second Step Bullying Prevention Units" with teacher manuals and student materials at each grade K-5
Too Good For Violence	K-8	Character values, SEL; healthy beliefs	7 30–60 min lessons plus infusion activities	Manual and program materials for each grade level K-8; separate high school kit

Note: Information summarized from CASEL (2013), reviews by the What Works Clearinghouse (n.d.), and program websites. Names in italics designate programs that provide curricular materials in "levels" that may span several grades

to students' learning and behavior: (1) listening to others, (2) following directions, (3) following classroom rules, (4) ignoring peer distractions, (5) asking for help, (6) taking turns in conversations, (7) cooperating with others, (8) controlling temper in conflict situations, (9) acting responsibly with others, and (10) showing kindness to others. Each of the 10 units includes 3 lessons (30 lessons total), with each lesson requiring approximately 20–25 min of instructional time. In the original (Elliott & Gresham, 2007) teacher's manual, the 10 units are available across three levels: PK-K, early elementary (Grades 1–2), and upper elementary/middle (Grades 3–6). CASEL (2013) described the SSIS-CIP as featuring a "grade-by-grade" sequence (CASEL, 2013); however, the same lesson plans, student booklets, and videos are used across all grades within each of these "levels".

Overall, research has underscored the importance of dosage in SEL interventions, showing that in general, more intervention exposure tends to improve student outcomes (January et al., 2011). A universal approach to schoolwide SEL implementation also has been encouraged by SEL program reviewers and prevention researchers (CASEL, 2013; Greenberg & Abenavoli, 2017). As such, schools may be interested in adopting programs similar to the SSIS-CIP in a universal fashion, with students being exposed to consistent SEL programming each year. However, few studies to date have addressed the effects of providing similar content to students across grade levels. Although teachers may modify curricular content based on assessments (formal or informal) of their students' needs, general education teachers receive limited training in evidence-based methods for program modification and, in one study, expressed confusion around these issues (Conderman, Liberty, & DeSpain, 2017).

In some ways, the idea of yearly exposure to a universal SEL program suggests that key concepts and skills should be re-introduced and built upon at an appropriate level for a child's age. This "spiral curricular approach" of revisiting foundational topics at each grade level has been a basis for the development many US academic curricula and textbooks (Bruner, 1960; Snider, 2004). However, spiral curricula have been criticized as re-introducing topics across grade levels without enough depth to promote mastery; rather, students simply receive "more of the same" (Schmidt, McKnight, & Raizen, 1997). A similar practical question concerns whether learning highly similar or the same SEL content across multiple years provides additional benefit for students.

Current Study

The purpose of the current study was to determine if students exposed to SSIS-CIP program across successive grade levels (first and second) benefitted beyond those exposed to the program only in a single year (second grade). The first hypothesis was that students with sustained exposure to the SSIS-CIP over time would demonstrate greater gains in social skills compared to their peers exposed to the program only once. Similarly, the second hypothesis was that sustained exposure to SSIS-CIP would increase students' approaches to learning and academic skills. The third and final hypothesis was that students experiencing the SSIS-CIP across multiple grade levels would demonstrate fewer problem behaviors than their peers who experienced the program in 1 year only. In addition to addressing an emerging question from the empirical literature regarding the benefit (or lack thereof) of sustained exposure to a universal SEL program over time, the study addresses a practical question with significant resource implications for schools considering adoption of the SSIS-CIP as a universal program within their primary grades.

Table 2 Student demographic characteristics by amount of exposure to SSIS-CIP

	Sustained ($N=181$)	Single ($N=218$)
Female	47.51	45.87
Male	52.49	54.13
Racial minority	19.55	31.31
White[a]	80.45	68.69
Special education services	8.67	10.42
Supplementary services	25.43	21.35

Table entries are mean (SD) for age and % for other variables

[a]Chi-square result significant at .05 level

Methods

Participants

Data for the current study were drawn from a longitudinal efficacy trial evaluating the SSIS-CIP. Participating students (Table 2) were from six elementary schools in the Mid-Atlantic region of the USA. "Single exposure" students ($N=218$) received the SSIS-CIP only during their second-grade school year. "Sustained exposure" students ($N=181$) initially completed the SSIS-CIP in their first-grade classroom and then were exposed to the program again during their second-grade year (Fig. 1). In both groups, approximately half of the students were female and the majority of students were White. There were, however, significantly fewer students of racial minority status in the sustained-exposure group (20%) compared to the single-exposure group (31%; $\chi^2 < .05$). The sustained exposure group also included slightly more students requiring special education or supplementary services,

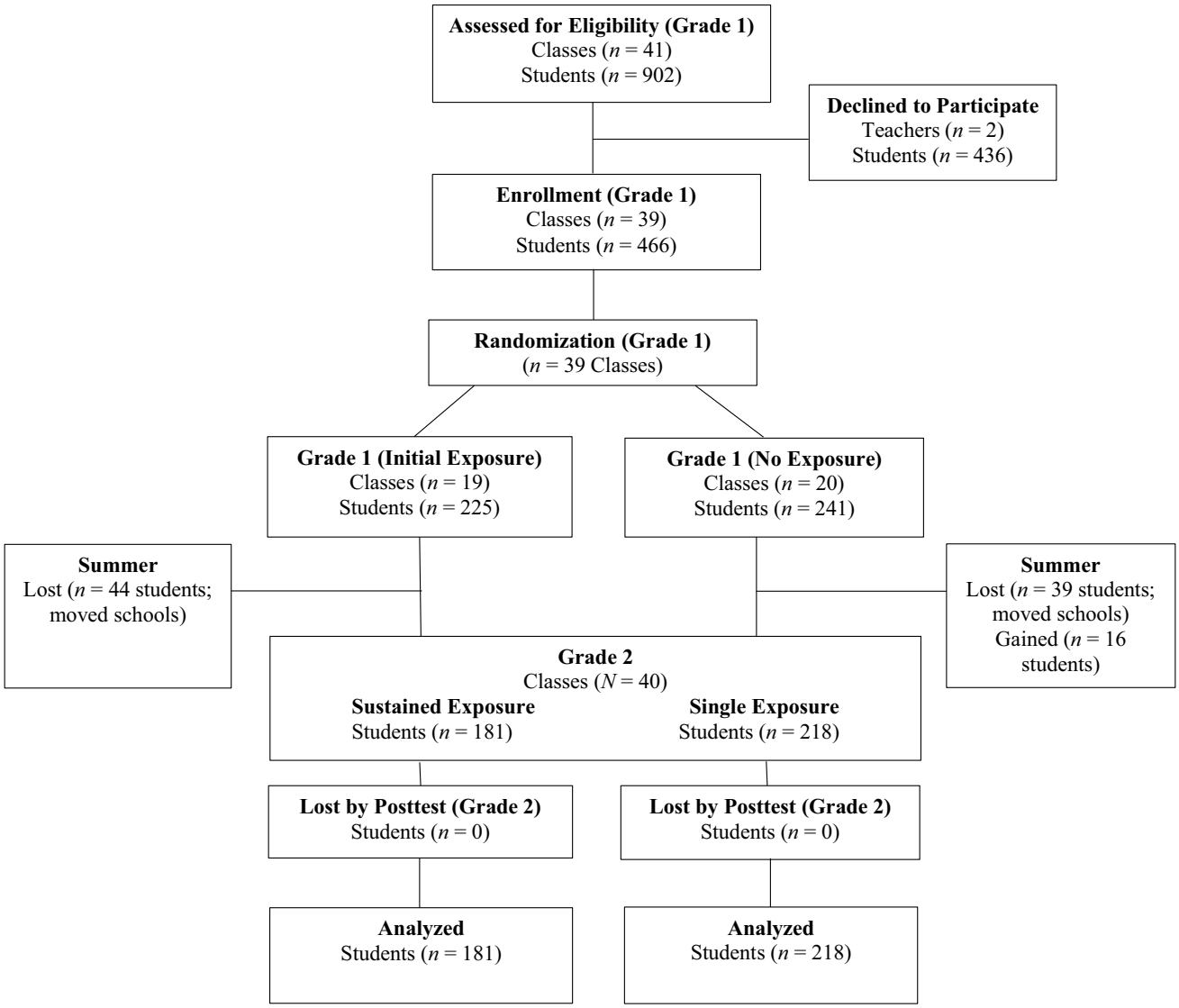

Fig. 1 Flow of participants through study conditions. Note: Because first-grade students were assigned to second-grade classrooms based on the school's typical practices, "sustained exposure" students were mixed with "initial exposure" students in second-grade classrooms. In some second-grade classes, however, there only were students with consent to participate in data collection from one group (sustained-only = 1 class, single only = 5 classes) due to the assignment of students to classrooms by the school

although this difference was not statistically significant (Table 2). All participating teachers were White, and 79% were female. In addition, teachers reported significant classroom experience ($M = 14.4$ years of experience, SD = 9 years).

Measures

Several measures were used to assess the primary student outcomes of interest (social skills, problem behaviors, academic engagement, and academic skills).

Social Skills and Problem Behavior

The Social Skills Improvement System Rating Scales–Teacher Form (SSIS-RST; Gresham & Elliott, 2008) were used to assess students' social behaviors in the classroom setting. The Social Skills scale includes 46 items across seven subscales (Communication, Cooperation, Assertion, Responsibility, Empathy, Engagement, and Self-control). The Problem Behavior scale includes 24 items across five subscales (externalizing, bullying, hyperactive-inattentive, internalizing, and autistic behavior). Teachers rate each item on a 4-point scale ranging from *never* to

almost always. The SSIS-RST has strong psychometric properties for the elementary grades and has been used widely to assess social skills and problem behaviors in research and practice (Gresham, Elliott, Vance, & Cook, 2011).

Academic Engagement and Motivation

Two subscales from the teacher version of the Academic Competence Evaluation Scales (ACES; DiPerna & Elliott, 2000) were used to measure students' approaches to learning. The Academic Motivation subscale contains 11 items that measure a student's approach, persistence, and level of interest regarding academic learning. The Academic Engagement subscale includes 8 items that reflect attention and active participation in classroom activities rated by teachers on a scale of 1 (*never*) to 5 (*almost always*). Scores from the ACES Motivation and Engagement subscales have demonstrated high reliability and been shown to be strongly related to academic achievement (DiPerna & Elliott, 2000).

In addition to teacher ratings via the ACES, the Cooperative Learning Observation Code for Kids (CLOCK; Volpe & DiPerna, 2010) was used to directly measure students' engagement during academic instruction. The CLOCK is a 12-min structured observation protocol with each minute divided into four 15-s intervals. As defined in the CLOCK, engaged time includes both active (e.g., raising hand, asking teacher a relevant question) and passive engagement (e.g., listening to a teacher talk, looking at the whiteboard or a worksheet) in classroom instruction. Engaged time is observed using momentary time sampling with engagement recorded at the beginning of each observation interval.

Academic Skills

STAR Math (Renaissance Learning, 2009) and Reading (Renaissance Learning, 2010) were used to assess students' academic skills. The STAR assessments are online adaptive tests that administer specific items based on a student's performance during the assessment. Although STAR measures are completed individually, they were administered in small groups, with each student having their own computer (laptop). Each STAR assessment (reading and math) required approximately 10–12 min to complete, and they were administered on different days to minimize fatigue. Scaled scores were used as indicators of academic achievement in the current study.

Procedure

Data Collection

Data were collected as part of a multi-year project examining efficacy of the SSIS-CIP. Both single- and sustained-exposure conditions followed the same data collection procedure. Within the original efficacy trial, all first- and second-grade teachers in the six participating schools were invited to participate in the project and subsequently randomized to condition. (Teachers were not informed of their condition until after baseline data collection was complete). Participating classrooms enrolled 20–25 students, and all students were invited to participate in the data collection to evaluate the efficacy of the SSIS-CIP. Approximately 12 students were randomly selected to participate in the data collection protocol from the students within each class with parental consent (stratified by sex). Child-level data were collected in all participating classrooms during 4-week periods before (November–December; baseline) and after (March–April; post-test) SSIS-CIP implementation in the classrooms randomly assigned to the treatment condition. Specifically, teachers completed the SSIS-RST for all participating children from their classroom, and they were paid for the time required to complete questionnaires.

In addition, trained observers completed CLOCK observations for a randomly selected (stratified by sex) subsample of 6 students ($n = 3$ boys and 3 girls) drawn from the 12 students participating in data collection from each classroom. Each selected student was observed during mathematics instruction on three separate occasions within each of the pre- and post-data collection periods. Observers ($N = 39$) had at least a bachelor's degree in psychology, education, or a related discipline. In addition, they completed formal training regarding the CLOCK (approximately 12 h of didactic instruction, practice observations, and individualized feedback) and had to meet a mastery criterion (80% accuracy when observing a video of students in an elementary classroom) before they could conduct observations for the project. Observations were distributed approximately evenly across observers, and each observation lasted for 12 min. One-third of the CLOCK observations were completed by pairs of observers, and agreement was high (*kappa* = .88–.94) across all target behavior domains and paired observations.

Sustained Exposure Trial

The sustained exposure trial was completed during the latter years of the larger project with two cohorts of students who entered the original efficacy trial during their first-grade year. Figure 1 shows the flow of participants from their enrollment in the larger project through collection of outcome data for the sustained exposure trial. All students entered the study in first grade, and their classrooms ($N = 39$) were randomly assigned to either the business as usual (control) condition ("no exposure;" classroom $N = 20$; student $N = 241$) or SSIS-CIP implementation (treatment) condition ("initial exposure;" classroom $N = 19$; student $N = 225$) for the original single-exposure efficacy trial. Students then

transitioned to second grade with a new set of teachers (who were all previously enrolled within the larger study and had the same amount of prior experience teaching the SSIS-CIP).

In second grade, the majority of students received the SSIS-CIP such that the "sustained exposure" group received the program a second time (student $N = 181$); and the "single exposure" group received it for the first time (student $N = 218$).[1] Given that first-grade students were assigned into second-grade classrooms based on each school's regular classroom assignment practices and all second-grade teachers ($N = 40$) implemented the SSIS-CIP program class-wide, sustained exposure students were mixed with single exposure students within second-grade classrooms. As noted in Fig. 1, student attrition occurred across both conditions during the summer months (between the end of first grade and beginning of second grade) in the first year of the sustained exposure trial due to closure and consolidation of elementary schools within one participating district. Finally, absence rates were similar across conditions and grade levels (single exposure means = 11 days in first grade & 6 days in second grade; sustained exposure means = 7 days in first grade and 6 days in second grade). It is important to note that mean absence rates are for the entire school year. As such, absences during SSIS-CIP instructional days likely were even lower given teachers typically taught one lesson per day and three lessons per week for approximately 10 weeks.

SSIS-CIP Implementation

The SSIS-CIP includes 10 instructional units focused on key classroom social behaviors identified by teachers as important for classroom success. Specifically, Units 1–3 target receptive skills (i.e., listening to others, following the steps, following the rules), Unit 4 teaches selective input (i.e., paying attention to your work), Unit 5 focuses on productive skills (i.e., asking a question), and Units 6–10 target interactive skills (i.e., communicating, cooperating, reading or managing emotions, and showing an understanding of rules). Each unit includes three scripted lessons, brief video vignettes (30–90 s), and practice exercises (student booklets). Each lesson requires approximately 20–25 min to complete and relies on six instructional strategies (describe, model, role-play, do, practice, monitor progress, and generalize) to help children learn the target skill for that unit. Additional information regarding the SSIS-CIP is available in the Instructor's Handbook (Elliott & Gresham, 2007).

Teachers were formally trained in advance of curriculum implementation. Specifically, the project director conducted a 1-day workshop with teachers from the implementation condition. During the first half of the workshop the facilitator provided a detailed overview of the SSIS-CIP curricular materials, including lesson plans, student booklets, and video vignettes. During the second half, teachers practiced teaching each lesson from the first SSIS-CIP unit in small groups. As teachers practiced, the workshop facilitator provided structured feedback regarding fidelity of their role-play lessons. In addition, teachers had the opportunity to ask questions regarding curricular implementation. After completion of the formal training, implementing teachers were expected to teach one SSIS-CIP unit (3 lessons) per week.

Two methods were used to evaluate and ensure fidelity of implementation of the SSIS-CIP lessons. First, implementing teachers completed weekly standardized checklists indicating their level of implementation (using a 4-point scale ranging from *Not Implemented* to *Full Implementation*) for the five core components (introduce, define, discuss, identify & practice steps, model/role-play) of each lesson within the unit. In addition, research staff completed direct observations for approximately 20% of the SSIS-CIP lessons taught by each teacher. Specifically, staff observed the entire lesson and then completed a structured report form that included 20 specific instructional actions/activities. Observers recorded if each activity was completed (or not) during the observed lesson and provided a summative judgment regarding the overall implementation of the five core lesson components using a 4-point scale ranging from *Not Implemented* (1) to *Full Implementation* (4). Observers completed a 2-h formal training prior to beginning the lesson observations with periodic monitoring from the lead research team to ensure they were completing lesson observations in accord with the observation protocol.

During the implementation period, the research team monitored implementation (via both self-report and independent observations) to ensure that teachers demonstrated at least 90% fidelity when teaching the SSIS-CIP lessons. Teachers were instructed to follow the lesson scripts regardless of students' initial or sustained exposure. If a teacher's implementation fell below the criterion threshold for a unit, a member of the research team contacted the teacher to discuss the area(s) of difficulty, reasons for the difficulty, and what needed to be done differently to achieve the curricular implementation standard. In addition, the research team periodically checked with all teachers (approximately every other week) to see if they had any implementation questions, make sure no unexpected barriers/difficulties had arisen that would adversely impact their ability to implement the SSIS-CIP lessons, and thank them for their efforts. As a result of the scripted format of the SSIS-CIP lessons and

[1] Sixteen second-grade students initially received the program in first grade; however, their second-grade teachers participated in data collection only (i.e., did not implement the SSIS-CIP), so these students were included in the single exposure condition.

Table 3 Reliability indices and ICC for student outcomes

	Reliability index		ICC (School)		ICC (Class)	
	Pretest	Posttest	Pretest	Posttest	Pretest	Posttest
Social skills composite	.98	.98	.03	.01	.24	.25
Communication	.91	.91	.06	.02	.25	.25
Cooperation	.94	.94	.003	.03	.11	.16
Assertion	.88	.88	< .0001	< .0001	.40	.37
Responsibility	.92	.91	.05	.04	.12	.18
Empathy	.93	.95	.04	.02	.15	.17
Engagement	.93	.94	.05	.001	.25	.22
Self-control	.93	.94	.05	.03	.27	.20
Problem behavior composite	.95	.95	.11	.11	.15	.23
Externalizing	.94	.94	.10	.13	.10	.18
Bullying	.90	.92	.12	.22	.10	.15
Hyperactive/inattentive	.90	.91	.06	.07	.14	.15
Internalizing	.89	.88	.11	.10	.20	.26
Approaches to learning						
Academic motivation	.98	.98	.01	< .0001	.09	.16
Academic engagement	.96	.95	< .0001	< .0001	.20	.24
Active engaged time[a]	.94	.92	< .0001	< .0001	.19	.11
Passive engaged time[a]	.90	.90	.26	.24	.21	.06
Academic skills						
Math scaled score	–	–	.17	.15	.10	.05
Reading scaled score	–	–	.14	.16	.13	.04

Reliability indices are Cronbach's α unless noted otherwise

ICC Intra-class correlation

"–" Not available

[a]Kappa agreement index

these monitoring efforts, implementation fidelity was high (> 95%) across all lessons, units, and implementing classrooms based on summative ratings by teachers and independent observers.

Design and Data Analysis

Multilevel modeling was used to account for the data structure of students nested within classes. Unconditional modes were first applied to report intraclass correlation (ICC). Table 3 presents ICCs at both school- and class-levels for all outcomes. Class-level ICCs for posttest outcome measures ranged from small (.04 for Reading scaled score) to large (.37 for Assertion). These levels of ICCs suggested that standard errors might be underestimated if the nested data structure was not taken into account. School-level variances of all posttest outcome measures were small and statistically nonsignificant based on z tests (2-tailed *p*s < .05). Therefore, we analyzed a two-level model for each outcome for parsimony and school indicators were included in the model as covariates.

To address the primary research questions regarding SSIS-CIP outcomes, both student-level and class-level variables were included to adjust for their effects. Student-level predictors included pretest scores of respective outcome measures (group-mean centered), gender (1 = male, 0 = female), race (1 = white, 0 = racial minority), receipt of supplemental services (1 = yes, 0 = no), and receipt of special education services (1 = yes, 0 = no). The dummy predictors were grand-mean centered. Class-level predictors included grand-mean centered class average of pretest scores of the respective outcome measures. Treatment condition was tested at the class-level using dummy codes (1 = sustained exposure, 0 = single exposure).

In addition to testing for main effect, interaction effects between condition and pretest scores (both student- and class-levels), as well as student demographic variables, were examined by adding product terms to the main effect model. If a product term was statistically significant after applying the Benjamini–Hochberg correction (1995) to control for false discovery rate, the pattern of interaction was further examined by plotting the adjusted means.

We estimated multilevel models using the Mixed procedure of SAS (version 9.4) for teacher ratings of social skills, approaches to learning, and academic skills as well as for two direct observation outcomes (active and passive

Table 4 Student-level means (SD) for pretest and posttest measures by condition

	Pretest		Posttest	
	Sustained	Single	Sustained	Single
Social skills composite	2.24 (.53)	2.29 (.53)	2.36 (.54)	2.35 (.53)
Communication	2.33 (.57)	2.37 (.58)	2.45 (.56)	2.45 (.55)
Cooperation	2.10 (.68)	2.19 (.63)	2.26 (.54)	2.23 (.65)
Assertion	2.03 (.61)	2.16 (.60)	2.25 (.65)	2.30 (.61)
Responsibility	2.30 (.61)	2.33 (.60)	2.42 (.60)	2.37 (.58)
Empathy	2.25 (.61)	2.29 (.62)	2.36 (.64)	2.35 (.60)
Engagement	2.35 (.59)	2.35 (.59)	2.42 (.64)	2.40 (.60)
Self-control	2.28 (.62)	2.32 (.62)	2.36 (.66)	2.36 (.66)
Problem behavior composite	.42 (.43)	.42 (.47)	.40 (.43)	.39 (.46)
Externalizing	.40 (.50)	.41 (.51)	.37 (.49)	.38 (.50)
Bullying	.22 (.40)	.24 (.52)	.21 (.39)	.21 (.41)
Hyperactivity-Inattention	.68 (.63)	.64 (.60)	.59 (.64)	.60 (.63)
Internalizing	.36 (.42)	.35 (.48)	.37 (.42)	.35 (.48)
Approaches to learning				
Academic motivation	3.57 (1.08)	3.57 (1.08)	3.81 (1.04)	3.69 (1.10)
Academic engagement	3.79 (.94)	3.88 (1.01)	4.01 (.97)	4.03 (.91)
Active engaged time[a]	.44 (.13)	.41 (.12)	.46 (.11)	.41 (.14)
Passive engaged time[a]	.36 (.14)	.35 (.13)	.35 (.12)	.38 (.15)
Academic skills				
Math scaled score	434.33 (100.44)	419.04 (102.73)	485.35 (107.94)	463.84 (113.29)
Reading scaled score	227.54 (125.72)	203. 55 (120.69)	287.75 (137.82)	256.23 (132.52)

Sustained-exposure $N = 181$; Single-exposure $N = 218$

[a]Direct observation data (Sustained-exposure $N = 99$; Single-exposure N = 109)

engagement). We used SAS Glimmix procedure for teacher ratings of problem behaviors. In addition, we estimated effect sizes of sustained-exposure compared with single-exposure condition. Specifically, the effect size was computed as a standardized mean difference by dividing the adjusted (for pretest scores and other student- and class-level covariates) group mean difference by the unadjusted pooled within-group student-level deviation of the pretest outcome measure, and 95% confidence intervals (CIs) were calculated for each effect size to indicate the precision of the estimate and range of possible effects. What Works Clearinghouse (2017) guidelines also recommend reporting effects as an "improvement index" representing expected percentile rank improvement for a comparison group participant (i.e., single exposure) had they received the intervention (i.e., sustained exposure). We present results using the improvement index to help readers better understand the practical impact of sustained exposure to the SSIS-CIP intervention.

Results

Our first hypothesis was that students exposed to the SSIS-CIP in first- and second-grade experience more improvement in social skills compared to their peers exposed to the intervention in second grade only. Tables 4 and 5 report student- and class-level means by condition. As shown in Table 6, sustained exposure to SSIS-CIP intervention yielded positive impact across all social skills outcomes. After controlling for pretest scores and demographic variables, sustained exposure to intervention achieved a statistically significant difference on posttest teacher ratings of overall social skills, assertion, and responsibility. After applying the Benjamini–Hochberg (1995) correction to control for false discovery rate, observed differences on assertion still met the adjusted threshold criterion for statistical significance. As expected, both student- and class-level pretest scores were statistically significant predictors for all the corresponding posttest outcome scores. Effect sizes were calculated at the mean of their respective pretest scores and controlling for demographic variables, and their 95% confidence intervals, and improvement indices were reported as well (Table 7). The magnitudes of the effect sizes were small–medium according to Cohen's (1988) criterion with responsibility having the largest effect size and improvement index. The improvement index for the overall social skills composite was 6.75, meaning that students in the single exposure group would have showed an approximate 7-point percentile rank increase in social skills had they been in the sustained exposure group. Scores on the assertion (7.53%)

Table 5 Class-level means (SD) for pretest and posttest measures by condition

	Pretest		Posttest	
	Sustained	Single	Sustained	Single
Social skills				
Social Skills composite	2.16 (.34)	2.28 (.34)	2.31 (.39)	2.36 (.32)
Communication	2.27 (.38)	2.37 (.37)	2.40 (.40)	2.47 (.33)
Cooperation	2.04 (.40)	2.20 (.34)	2.22 (.44)	2.26 (.32)
Assertion	1.97 (.46)	2.13 (.42)	2.21 (.49)	2.28 (.41)
Responsibility	2.22 (.37)	2.33 (.35)	2.36 (.42)	2.38 (.33)
Empathy	2.16 (.38)	2.29 (.37)	2.32 (.42)	2.36 (.34)
Engagement	2.27 (.39)	2.35 (.38)	2.36 (.43)	2.41 (.35)
Self-control	2.19 (.48)	2.32 (.40)	2.28 (.48)	2.34 (.40)
Problem behaviors				
Problem behaviors composite	.45 (.34)	.44 (.27)	.44 (.34)	.43 (.31)
Externalizing	.44 (.37)	.43 (.28)	.42 (.38)	.41 (.32)
Bullying	.23 (.30)	.26 (.27)	.23 (.30)	.25 (.27)
Hyperactivity-Inattention	.73 (.47)	.66 (.33)	.66 (.48)	.62 (.38)
Internalizing	.39 (.35)	.38 (.28)	.40 (.33)	.38 (.31)
Approaches to learning				
Academic Motivation	3.47 (.65)	3.58 (.54)	3.74 (.68)	3.73 (.54)
Academic Engagement	3.72 (.61)	3.82 (.58)	3.96 (.67)	4.05 (.51)
Active Engaged Time	.44 (.10)	.39 (.08)	.45 (.08)	.42 (.09)
Passive Engaged Time	.36 (.13)	.36 (.10)	.35 (.09)	.37 (.10)
Academic skills				
Math scaled score	425.98 (66.08)	417.05 (68.43)	472.88 (57.79)	457.76 (69.19)
Reading scaled score	219.42 (90.08)	203.11 (70.14)	275.15 (86.20)	253.52 (78.22)

Second-grade teacher ratings ($N = 40$)

Table 6 Mixed model estimates (standard errors) on social skills outcomes

	Teacher rating							
	Social Skill	Communication	Cooperation	Assertion	Responsibility	Empathy	Engagement	Self-control
Intercept	2.39** (.19)	2.53** (.23)	2.48** (.23)	1.97** (.22)	2.47** (.21)	2.35** (.24)	2.45** (.24)	2.47** (.24)
Covariates								
Student pretest	.72** (.04)	.63** (.04)	.72** (.04)	.70** (.05)	.68** (.04)	.65** (.04)	.69** (.05)	.82** (.05)
Class-level pretest	.84** (.07)	.69** (.08)	.73** (.09)	.87** (.05)	.86** (.08)	.79** (.08)	.79** (.08)	.86** (.07)
Male	−.05 (.03)	−.05 (.04)	−.07 (.04)	−.03 (.04)	−.09* (.04)	−.10* (.04)	−.004 (.04)	−.06 (.04)
White	.01(.05)	.04 (.06)	.03 (.06)	.03 (.06)	−.03 (.06)	.04 (.07)	−.03 (.07)	.01 (.06)
Supp. services	−.10* (.04)	−.05 (.05)	−.09 (.05)	−.08 (.05)	−.14** (.05)	−.17** (.05)	−.11* (.05)	−.10 (.05)
Special education	−.14* (.06)	−.15* (.07)	−.10 (.07)	−.19** (.07)	−.11 (.07)	−.19* (.08)	−.20* (.08)	−.09 (.08)
Treatment effects								
Sustained exp.	.09* (.04)	.06 (.06)	.11 (.06)	.12** (.04)	.11* (.05)	.09 (.05)	.05 (.05)	.08 (.05)
	$p = .04$	$p = .32$	$p = .08$	$p = .01$	$p = .02$	$p = .11$	$p = .32$	$p = .14$
Random effects								
Intercept variance	.01 (.01)	.03** (.01)	.02** (.01)	<.0001	.01 (.01)	.01 (.01)	.01 (.01)	.01 (.01)
Residual variance	.09** (.01)	.13** (.01)	.13** (.01)	.14** (.01)	.12** (.01)	.16** (.010	.15** (.01)	.15** (.01)

Analytical sample size = 364. Cohort and school indicators are included in the model but not reported. Only treatment effects for Assertion remained statistically significant after applying the Benjamini–Hochberg correction

*$p < .05$; **$p < .01$

Table 7 Standardized group differences, 95% confidence intervals, and improvement indices

Variables	Effect size	95% Confidence interval	Improvement index (%)
Social skills			
Social skills composite	.17	[.01, .32]	6.75
Communication	.10	[− .10, .31]	3.98
Cooperation	.17	[− .01, .34]	6.75
Assertion	.19	[.06, .32]	7.53
Responsibility	.18	[.02, .34]	7.14
Empathy	.14	[− .03, .31	5.57
Engagement	.09	[− .08 .26]	3.59
Self-control	.11	[− .04, .27]	4.38
Problem behaviors			
Problem behaviors composite	− .01	[− .09, .39]	− .40
Externalizing	− .01	[− .07, .26]	− .40
Hyperactive-Inattentive	− .05	[− .05, .19]	− 1.99
Internalizing	.04	[− .09, .44]	1.60
Approaches to learning			
Academic motivation	.18	[.02, .34]	7.14
Academic engagement	.11	[− .06, .28]	4.38
Active engaged time	.37	[.01, .69]	14.43
Passive engaged time	− .24	[− .53, .06]	− 9.48
Academic skills			
Math	.13	[− .005, .27]	5.17
Reading	.09	[− .02, .21]	3.59

Standardized differences adjusted for pretest and other student- and class-level covariates

Table 8 Mixed model estimates (standard errors) for sustained exposure and pretest interaction effect on social skills outcomes

	Social Skills composite	Communication	Cooperation	Assertion	Responsibility	Empathy	Engagement	Self-control
Intercept	2.41** (.19)	2.54** (.23)	2.49** (.23)	1.98** (.22)	2.51** (.21)	2.37** (.25)	2.47** (.24)	2.47** (.24)
Covariates								
Student pretest	.68** (.05)	.58** (.06)	.79** (.05)	.61** (.06)	.63** (.05)	.62** (.06)	.60** (.06)	.81** (.06)
Class pretest	.09 (.08)	.64** (.11)	.65** (.12)	.84** (.07)	.73** (.09)	.73** (.10)	.74** (.10)	.81** (.09)
Male	− .05 (.03)	− .05 (.04)	− .06 (.04)	− .03 (.04)	− .10** (.04)	− .10* (.05)	− .01 (.04)	− .06 (.04)
White	− .003 (.05)	.02 (.06)	.05 (.06)	.02 (.06)	− .05 (.06)	.03 (.07)	− .04 (.07)	.01 (.07)
Supp. Services	− .10* (.04)	− .06 (.05)	− .09 (.05)	− .08 (.05)	− .14** (.05)	− .17** (.05)	− .11 (.05)	− .10 (.05)
Special Educ.	− .13* (.06)	− .15* (.07)	− .10 (.07)	− .18* (.07)	− .11 (.07)	− .18* (.08)	− .18* (.08)	− .09 (.08)
Treatment effects								
Sustained exp.	.08* (.04)	.06 (.06)	.10 (.06)	.11* (.04)	.11* (.04)	.08 (.05)	.05 (.05)	.07 (.05)
Interaction effect								
Sustained exp.*	.09 (.08)	.12 (.09)	− .15* (.07)	.18 (.09)	.09 (.07)	.08 (.08)	.19* (.09)	.03 (.09)
Student pretest	$p = .23$	$p = .16$	$p = .04$	$p = .05$	$p = .20$	$p = .34$	$p = .041$	$p = .77$
Sustained exp.*	.25 (.13)	.12 (.16)	.16 (.187)	.07 (.10)	.33* (.13)	.16 (.15)	.14 (.15)	.09 (.13)
Class pretest	$p = .06$	$p = .44$	$p = .35$	$p = .50$	$p = .02$	$p = .31$	$p = .34$	$p = .48$
Random effects								
Intercept var.	.01 (.01)	.03** (.01)	.02** (.01)	.001 (.01)	.01 (.01)	.01 (.01)	.01 (.01)	.01 (.01)
Residual var.	.09** (.01)	.13** (.01)	.13** (.01)	.14** (.01)	.12** (.01)	.16** (.01)	.15** (.01)	.15** (.01)

Cohort and school indicators are included in the model but not reported. Treatment differences were not statistically significant after applying the Benjamini–Hochberg correction

*$p < .05$; **$p < .01$

Table 9 Mixed model estimates (standard errors) for problem behaviors outcomes

	Problem behaviors	Externalizing	Hyperactive-inattentive	Internalizing
Intercept	− .89 (1.05)	− 1.77 (1.57)	− .98 (.96)	− .30 (.75)
Covariates				
Student-level pretest	.92** (.17)	.97** (.15)	1.01** (.12)	.91** (.19)
Class-level pretest	1.79** (.38)	1.53** (.34)	1.13** (.21)	1.97** (.36)
Male	.24 (.19)	.34 (.20)	.32* (.16)	.05 (.20)
White	.05 (.25)	.03 (.25)	.12 (.21)	.25 (.27)
Supp. services	.20 (.19)	.18 (.19)	.21 (.16)	.28 (.21)
Special education	.09 (.29)	− .01 (.31)	.03 (.23)	.19 (.30)
Treatment effect				
Sustained-exposure	− .01 (.18)	− .03 (.18)	− .07 (.14)	.05 (.19)
	$p = .96$	$p = .89$	$p = .64$	$p = .77$
Random effect				
Intercept variance	< .0001	< .0001	< .0001	< .0001

Analysis sample $N = 364$. Cohort and school indicators are included in the model but not reported. Scores transformed to a log scale for analysis

*$p < .05$; **$p < .01$

and responsibility (7.14%) subscales showed similar rates of expected improvement.

There was a statistically significant interaction between amount of exposure and student-level pretest on cooperation and social engagement (Table 8). Compared to their peers in single-exposure group, students in the sustained-exposure group had a slightly higher adjusted posttest mean on cooperation, and that difference was larger for students who had lower pretest scores. Students in the sustained-exposure group also had a slightly higher adjusted posttest score on social engagement, but the difference was larger for those who had higher pretest scores. Finally, there was an interaction between amount of exposure and class-level pretest on responsibility, wherein the adjusted differences between sustained- and single-exposure groups were larger for classes that had higher average pretest scores. After applying the Benjamini–Hochberg correction to control for false discovery rate, none of these interactions met the adjusted threshold for statistical significance.

The second hypothesis was that students exposed to the SSIS-CIP across multiple grade levels demonstrate fewer problem behaviors than their peers exposed to the program in second grade only. Parameter estimates for the main effect model for problem behavior outcomes are presented in Table 9. Sustained exposure to SSIS-CIP intervention yielded a negative effect on most of the problem behavior outcomes except for Internalizing. However, there were no statistically significant differences between sustained- and single exposure on any of the problem behavior outcomes. Both student- and class-level pretest scores significantly predicted posttest outcomes. Effect size estimates for the problem behaviors subscales were close to 0 with improvement indexes < 2 (Table 7). There was an observed interaction showing higher rates of externalizing behavior for students in the sustained exposure group with higher initial levels of externalizing behaviors (Table 10); however this interaction was not statistically significant after applying the Benjamini–Hochberg correction.

Our third hypothesis was that sustained exposure to SSIS-CIP intervention across grade levels increases students' approaches to learning and academic skills. The parameter estimates for the main effects models are reported in Table 11. Teacher rating of academic motivation and direct observations of active engaged time were consistent with this hypothesis, but neither of these differences satisfied an adjusted threshold for statistical significance after the Benjamini–Hochberg correction. There was a statistically significant interaction between amount of exposure and student-level pretest on observed active engaged time, showing higher adjusted posttest scores for sustained exposure students who had lower pretest scores (Table 12), but again, this difference was not statistically significant after applying the Benjamini–Hochberg correction. Finally, main effects for students' math and reading performance were not statistically significant between single- and sustained-exposure groups (Table 11).

Discussion

The goal of this study was to examine the effects of elementary students experiencing a universal SEL curriculum across two consecutive school years—first and second grade. Results showed that sustained exposure to the SSIS-CIP yielded small positive effects for assertion, responsibility, and social skills overall, with improvement rates

Table 10 Mixed model estimates (standard errors) for sustained exposure and pretest interaction effect on problem behavior outcomes

	Problem Behaviors	Externalizing	Hyperactive-inattentive	Internalizing
Intercept	.75 (.39)	− 1.67 (1.57)	− .96 (.9574)	− .26 (.75)
Covariates				
Student pretest	− .30* (.14)	.77** (.18)	.89** (.15)	.84** (.22)
Class pretest	− .34 (.22)	1.54** (.49)	1.22** (.33)	1.87** (.45)
Male	− .03 (.07)	.43* (.20)	.36* (.16)	.05 (.20)
White	.002 (.11)	.1002 (.26)	.15 (.21)	.26 (.28)
Supp. services	− .04 (.09)	.22 (.20)	.22 (.16)	.29 (.21)
Special education	− .15 (.13)	− .05 (.31)	.02 (.23)	.12 (.31)
Treatment effect				
Sustained-exp.	.03 (.07)	− .23 (.22)	− .14 (.17)	− .03 (.21)
Interaction effect				
Sustained-exp.* student pretest	− .14 (.22) $p = .54$	.77* (.34) $p = .03$	.28 (.24) $p = .25$	.40 (.43) $p = .36$
Sustained-exp.* class pretest	.01 (.27) $p = .98$	.02 (.52) $p = .97$	− .14 (.38) $p = .71$	.23 (.51) $p = .66$
Random effect				
Intercept variance	< .0001	< .0001	< .0001	< .0001

Cohort and school indicators are included in the model but not reported. Scores transformed to a log scale for analysis. Differences were not statistically significant after applying the Benjamini–Hochberg correction

*$p < .05$;; **$p < .01$

Table 11 Mixed model estimates (standard errors) for approaches to learning and academic skills outcomes

	Teacher rating		Direct observation		Academic skills	
	Academic motivation	Academic engagement	Active engaged time[a]	Passive engaged time[a]	Math	Reading
Intercept	3.61** (.33)	4.14** (.31)	.47** (.09)	.45** (.09)	433.20** (40.92)	261.89** (42.35)
Covariates						
Student-level pretest	.79** (.03)	.74** (.03)	.14 (.10)	.09 (.10)	.76** (.05)	.83** (.04)
Class-level pretest	.82** (.08)	.72** (.07)	.14 (.13)	.15 (.11)	.72** (.09)	.78** (.07)
Male	− .19** (.06)	− .06 (.05)	− .03 (.02)	− .02 (.02)	2.50 (7.34)	− 6.23 (7.75)
White	− .02 (.09)	.003 (.08)	− .02 (.03)	− .02 (.03)	7.40 (11.16)	11.09 (11.59)
Supp. Services	− .21** (.09)	− .21** (.07)	− .04 (.02)	− .01 (.02)	− 25.75	− 30.56** (9.71)
Special Education	− .08 (.10)	− .25** (.09)	.01 (.03)	− .05 (.03)	− 58.47** (13.10)	− 41.40** (14.06)
Treatment effect						
Sustained-exposure	.19* (.08)	.10 (.08)	.05* (.02)	− .03 (.02)	14.25 (7.50)	12.45 (7.92)
	$p = .03$	$p = .21$	$p = .04$	$p = .28$	$p = .06$	$p = .12$
Random effect						
Intercept variance	.06** (.02)	.06** (.020	.002 (.001)	.001 (.001)	<.0001	<.0001
Residual variance	.24** (.02)	.21** (.02)	.01** (.002)	.01 (.002)	4693.89** (356.87)	5025.26** (385.99)

Analytical sample sizes: Teacher Rating $N = 364$; Direct Observation $N = 184$; Math $N = 360$; Reading $N = 353$. Cohort and school indicators are included in the model but not reported. Differences were not statistically significant after applying the Benjamini–Hochberg correction for false discovery rate

[a]Direct observation data; scores transformed to a log scale for data analysis

*$p < .05$; ** $p < .01$

Table 12 Mixed model estimates (standard errors) for sustained exposure and pretest interaction effect on approaches to learning and academic skills outcomes

	Academic motivation	Academic engagement	Active engaged time[a]	Passive engaged time[a]	Math	Reading
Intercept	3.61** (.33)	4.19** (.31)	.50** (.09)	.45** (.09)	429.79** (41.21)	261.56** (42.48)
Covariates						
Student pretest	.84** (.04)	.70** (.04)	.24 (.13)	.09 (.13)	.82** (.06)	.84** (.05)
Class pretest	.71** (.11)	.65** (.09)	.50* (.19)	.15 (.16)	.73** (.09)	.80** (.09)
Male	−.19** (.06)	−.06 (.05)	−.03 (.02)	−.02 (.02)	1.50 (7.42)	−6.41 (7.79)
White	−.01 (.09)	−.002 (.08)	−.02 (.03)	−.02 (.03)	8.29 (11.18)	10.71 (11.66)
Supp. services	−.21** (.07)	−.20** (.07)	−.04 (.02)	−.01 (.02)	−26.66** (9.27)	−30.86** (9.77)
Special education	−.08 (.10)	−.24* (.09)	.02 (.03)	−.05 (.03)	−56.80** (13.16)	−41.03** (14.17)
Treatment effect						
Sustained-exp.	.19* (.08)	.10 (.08)	.05* (.02)	−.03 (.02)	14.08 (7.52)	12.09 (8.01)
Interaction effect						
Sustained-exp* student pretest	−.11 (.06) $p=.06$	.11 (.06) $p=.08$	−.20 (.20) $p=.31$	.002 (.21) $p=.99$	−.13 (.09) $p=.16$	−.02 (.07) $p=.83$
Sustained-exp* class pretest	.22 (.16) $p=.16$	.18 (.15) $p=.23$	−.61* (.26) $p=.02$	.01 (.20) $p=.98$	−.05 (.13) $p=.70$	−.05 (.12) $p=.70$
Random effect						
Intercept variance	.06** (.02)	.06** (.02)	.001 (.001)	.001 (.001)	<.0001	<.0001
Residual variance	.24** (.02)	.21** (.02)	.01** (.002)	.01** (.002)	4692.34** (357.79)	5052.14** (389.20)

Cohort and school indicators are included in the model but not reported. Differences were not statistically significant after applying the Benjamini–Hochberg correction for false discovery rate

[a]Direct observation data; scores transformed to a log scale for data analysis

*$p<.05$; ** $p<.01$

of approximately 7% for each. Only effects for assertion, though, met the threshold for statistical significance after controlling for false discovery rate (What Works Clearinghouse, 2017). Although several interactions were observed between initial skill level at the student or class level and sustained exposure, none of these interactions remained statistically significant after controlling for false discovery rate.

Specifically, we found that sustained exposure to the SSIS-CIP curriculum benefitted second-grade students who had participated in the curriculum previously in first grade ($d=.17$ for overall social skills). Differences between single- and no-exposure in second grade as measured in the initial efficacy trial were larger than those found here ($d=.36$; DiPerna et al., 2015); however, the effects reported in this study represent benefit *above and beyond* exposure to the SSIS-CIP in one grade. Among social skills subscales, second-grade students in the sustained exposure condition showed the most improvement in responsibility ($d=.17$) and assertion ($d=.19$). These results generally are consistent with previous research showing that higher dosage of SEL curricula leads to more gains in social and emotional skills (Beets et al., 2009; Nelson et al., 2003). Effects also were similar in magnitude to those cited in previous randomized-controlled trials (e.g., January et al., 2011).

We also examined the effects of sustained exposure on academic engagement, motivation, problem behaviors, and academic skills. We observed positive effects for sustained exposure on second graders' academic motivation and social engagement, although these findings did not remain statistically significant after adjustment for false discovery. Statistically significant improvements were observed in these domains for second grade-only exposure to the SSIS-CIP with motivation being moderated by pre-intervention scores (DiPerna et al., 2015). Findings from the current study suggest that primary students' academic motivation and social engagement may be *further* improved to a small degree with subsequent exposure to the SSIS-CIP. We did not find a significant benefit of sustained exposure for problem behaviors or academic skills, and this could be reflective of the relative difficulty of changing these skills with a program focused on exclusively teaching prosocial skills, as indicated in studies of other SEL programs (Taylor et al., 2017). In addition, findings are consistent with those of Hart, DiPerna, Lei, and Cheng (2020), where single exposure to the SSIS-CIP in second grade did not yield many statistically significant

differences in students' state test scores in the intermediate grades.

Our results suggest some potential benefit to sustained exposure to SEL program implementation, consistent with the idea that foundational topics should be constantly revisited over time (Bruner, 1960). The original intent of a "spiral" approach to curricula was to teach fundamental concepts according to students' developmental level, with students "building on them, making them more complex, and understanding them more fully" (Kazdin, 2005). In the SSIS-CIP program manual specifically, teachers are encouraged to adapt role play and other practice activities to situations relevant to their classrooms. In this study, teachers in both grades maintained a high level of adherence to the program lessons (over 90%); however, it is possible that teachers made developmental adjustments even while attaining high fidelity given the program affords opportunity for customizing components such as role-plays and practice situations. Although studying developmental adaptations were not focus of the efficacy trial, the nature and effectiveness of teachers' approach to adaptations remain an important area for future research.

Practical Implications

This study has several implications for school teams making decisions about implementing social-emotional learning programs on a multi-year basis. First, teams should carefully define their goals for SEL prior to program implementation. For example, teams may wish to conduct a school-based needs assessment or other systematized evaluation in order to clarify whether (and in what areas) a multi-year intervention is necessary (see CASEL, 2019). If social-emotional improvement, especially in prosocial skills, is a school-wide early intervention goal, then a "boost" from subsequent exposure to programs like the SSIS-CIP in primary grades may assist with achieving this goal. Conversely, if the primary goal is to reduce externalizing problem behavior, then implementing universal SEL programs like the SSIS-CIP across repeated years is unlikely to meet this need. Other important practical considerations include the class time required for implementation and the cost of purchasing materials for several grade levels.

Several universal SEL programs like the SSIS-CIP have similar (or matching) content across grade levels (see Table 1 for examples). Thus, student engagement during lessons containing the same material across several grades becomes a consideration. Although skilled teachers may be able to adjust role plays, examples, or practice activities based on their observation of student skills and areas for growth, school teams should systematically address how students will receive *sustained* exposure to key SEL concepts without simply receiving "more of the same" (e.g., NASP, 2011). SEL program developers and researchers should also address issues of social validity and student engagement for programs that recommend sustained, year-after-year implementation, given the importance of user satisfaction for successful program implementation (Durlak & DuPre, 2008).

Finally, the observed instances of moderation in our results suggest that some students or classrooms may have benefitted more than others from sustained exposure to the SSIS-CIP. In general, classes showed more improvement on the observational measure of engagement and teacher-rated cooperation if their pre-test engagement scores were lower. This result may support a more targeted approach to implementation (sustained exposure) for classes experiencing difficulty at the beginning of the year. In contrast, we found that other skill areas such as teacher-rated social engagement and responsibility showed an interaction in the opposite direction, with students performing higher at post-test if their pre-test scores were higher. Thus, the program potentially offered an extra "boost" to students who showed high initial prosocial behavior. For teacher-rated externalizing behavior, students with higher initial rates of externalizing behavior showed more externalizing behavior at post-test compared to peers in the single exposure group. These interactions of sub-skill areas, however, were not statistically significant after adjusting for multiple tests using the Benjamini–Hochberg correction (What Works Clearinghouse, 2017). Thus, replication of this study is necessary to further examine if effects are moderated by pretest scores at the student and/or class levels.

Conclusions and Future Directions

Although this study features a unique controlled design that allows for causal inferences about the relative benefit of second-year universal SEL program exposure compared to 1 year of exposure, there are several limitations to consider when interpreting findings. First, data for the current study were collected during the latter phase of a larger efficacy trial studying 1-year implementation across two grade levels, and this approach did not allow for additional control groups (for the single exposure group) or re-randomization to condition in second grade. In addition, due to the differences in raters (teachers) across grade levels, we were not able to examine longer-term effects on teacher ratings across both grade levels. Thus, the current study represents an initial attempt to understand the potential benefits of sustained implementation but does not represent the most rigorous, controlled evaluation of this question that would be possible with additional resources.

Third, given the diverse range of SEL programs available (CASEL, 2013), it is unknown if these results generalize to other available programs. Indeed, the effects of sustained

exposure may be different for programs that more explicitly differentiate content/target skills across grade levels. While a number of SEL programs have components similar to those of the SSIS-CIP (e.g., teacher-instructed lessons, role play, practice opportunities), future research should explicitly test the effects of sustained exposure for varied SEL programs using rigorously controlled designs. Fourth, effect sizes should be interpreted relative to those from previous studies featuring similar methodology and interventions (Durlak, 2009; Ferguson, 2009). Unfortunately, we could not locate similar previous "sustained exposure" trials to contextualize the current findings. Completion of such studies is imperative to help interpret efficacy of sustained exposure approaches and ultimately inform policy and practice decisions. Finally, this study did not isolate the specific "active ingredients" that produced student skill change (e.g., opportunities for practice, teacher modeling), so examining the relative contributions of SSIS-CIP program components remains an important area for future research (Jones & Bouffard, 2012).

With universal SEL programs increasingly being implemented across schools in the USA, it is necessary to empirically examine the effects of students' exposure to the programs across several grades. Results of the current study show that second-grade exposure to the SSIS-CIP further promoted first-grade students' social skills and showed promising trends for other areas of academic motivation and engagement to a small, but meaningful degree based on reviews of similar programs (January et al., 2011). Problem behavior and academic skills, however, were not affected after 2 years of exposure to the SSIS-CIP. As such, schools interested in using the SSIS-CIP may benefit from re-introducing key SEL concepts each year at a level appropriate for students' developmental needs.

Funding The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305A090438 to The Pennsylvania State University. The opinions expressed are those of the authors and do not represent the views of the Institute or the U.S. Department of Education.

Availability of Data and Materials N/A.

Code Availability N/A.

Compliance with Ethical Standards

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

References

Beets, M. W., Flay, B. R., Vuchinich, S., Snyder, F. J., Acock, A., Li, K.-K., et al. (2009). Use of a social and character development program to prevent substance use, violent behaviors, and sexual activity among elementary-school students in Hawaii. *American Journal of Public Health, 99*(8), 1438–1445. https://doi.org/10.2105/AJPH.2008.142919.

Belfield, C., Bowden, A. B., Klapp, A., Levin, H., Shand, R., & Zander, S. (2015). The economic value of social and emotional learning. *Journal of Benefit-Cost Analysis, 6*(3), 508–544.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B. Methodological, 57,* 289–300.

Bruner, J. (1960). *The process of education*. Cambridge: Harvard University Press.

Caprara, G. V., Barbaranelli, C., Pastorelli, C., Bandura, A., & Zimbardo, P. G. (2000). Prosocial foundations of children's academic achievement. *Psychological Science, 11*(4), 302–306.

CASEL. (2019). District Theory of Action: Needs Assessment. Retrieved from https://casel.org/in-the-district/needs/

CASEL (Producer). (2013). Effective social and emotional learning programs: Preschool and elementary school. Retrieved from http://casel.org/wp-content/uploads/2016/01/2013-casel-guide-1.pdf

Cohen, J. (1988). *Statistical power analysis for the beahvioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Conderman, G., Liberty, L., & DeSpain, S. (2017). Understanding accommodations, modifications, and interventions. *Kappa Delta Pi Record, 53*(2), 70–75. https://doi.org/10.1080/00228958.2017.1299545.

DiPerna, J. C., & Elliott, S. N. (2000). *Academic competence evaluation scales*. San Antonio, TX: Psychological Corporation.

DiPerna, J. C., Lei, P., Bellinger, J., & Cheng, W. (2015). Efficacy of the social skills improvement system classwide intervention program (SSIS-CIP) primary version. *School Psychology Quarterly, 30*(1), 123–141. https://doi.org/10.1037/spq0000079.

DiPerna, J. C., Lei, P., Bellinger, J., & Cheng, W. (2016). Effects of a universal positive classroom behavior program on student learning. *Psychology in the Schools, 53*(2), 189–203. https://doi.org/10.1002/pits.21891.

DiPerna, J. C., Lei, P., Cheng, W., Hart, S. C., & Bellinger, J. (2018). A cluster randomized trial of the Social Skills Improvement System-Classwide Intervention Program (SSIS-CIP) in first grade. *Journal of Educational Psychology, 110*(1), 1.

Durlak, J. A. (2009). How to select, calculate, and interpret effect sizes. *Journal of Pediatric Psychology, 34*(9), 917–928.

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. https://doi.org/10.1111/j.1467-8624.2010.01564.x.

Elbertson, N. A., Brackett, M. A., & Weissberg, R. P. (2009). School-based social and emotional learning (SEL) programming: Current perspectives. In A. Hargreaves, A. Lieberman, M. Fullan, & D. Hopkins (Eds.), *Second international handbook of educational change* (pp. 1017–1032). Dordrecht: Springer.

Elliott, S. N., & Gresham, F. M. (2007). *Social skills improvement system: Classwide intervention program*. Bloomington, MN: NCS Pearson.

Ferguson, C. J. (2009). An effect size primer: A guide for clinicians and researchers. *Professional Psychology, Research and Practice, 40,* 532–538. https://doi.org/10.1037/a0015808.

Greenberg, M. T., & Abenavoli, R. (2017). Universal interventions: Fully exploring their impacts and potential to produce population-level impacts. *Journal of Research on Educational Effectiveness, 10*(1), 40–67. https://doi.org/10.1080/19345747.2016.1246632.

Greenberg, M. T., Domitrovich, C. E., Graczyk, P. A., & Zins, J. E. (2005). The study of implementation in school-based preventive interventions: Theory, research, and practice. *Promotion of Mental Health and Prevention of Mental and Behavioral Disorders, 3,* 21.

Gresham, F. M., & Elliott, S. N. (2008). *Social skills improvement system: Rating scales*. Bloomington, MN: Pearson Assessments.

Gresham, F. M., Elliott, S. N., Vance, M. J., & Cook, C. R. (2011). Comparability of the social skills rating system to the social skills improvement system: Content and psychometric comparisons across elementary and secondary age levels. *School Psychology Quarterly, 26*(1), 27.

Hart, S. C., DiPerna, J. C., Lei, P.-W., & Cheng, W. (2020). Nothing lost, something gained? Impact of a universal social–emotional learning program on future state test performance. *Educational Researcher, 49*(1), 5–19. https://doi.org/10.3102/0013189X19898721.

Hunter, L. J., DiPerna, J. C., Hart, S. C., & Crowley, M. (2018). At what cost? Examining the cost effectiveness of a universal social–emotional learning program. *School Psychology Quarterly, 33*(1), 147.

January, A. M., Casey, R. J., & Paulson, D. (2011). A meta-analysis of classroom-wide interventions to build social skills: Do they work? *School Psychology Review, 40*(2), 242–256.

Jones, S. M., & Bouffard, S. M. (2012). Social and Emotional Learning in Schools: From Programs to Strategies and commentaries. *Social Policy Report, 26*(4), 1–33.

Jones, D. E., Greenberg, M. T., & Crowley, M. (2015). Early social-emotional functioning and public health: The relationship between kindergarten social competence and future wellness. *American Journal of Public Health, 105*(11), 2283–2290. https://doi.org/10.2105/AJPH.2015.302630.

Kazdin, A. E. (2005). Social validity. In B. S. Everitt & D. C. Howell (Eds.), *Encyclopedia of statistics in behavioral science* (pp. 1875–1876). New York: Wiley. https://doi.org/10.1002/0470013192.bsa632.

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.

National Association of School Psychologists. (2011). Position statement: Grade retention and social promotion. Retrieved from https://www.nasponline.org/x26820.xml

Nelson, G., Westhues, A., & MacLeod, J. (2003). A Meta-Analysis of Longitudinal Research on Preschool Prevention Programs for Children. *Prevention & Treatment, 6*(1), No Pagination Specified-No Pagination Specified. https://doi.org/10.1037/1522-3736.6.1.631a

Proctor, E., Silmere, H., Raghavan, R., Hovmand, P., Aarons, G., Bunger, A., et al. (2011). Outcomes for implementation research: Conceptual distinctions, measurement challenges, and research agenda. *Administration and Policy in Mental Health and Mental Health Services Research, 38*(2), 65–76.

Renaissance Learning. (2009). *STAR Math technical manual*. Wisconsin Rapids, WI: Renaissance Learning.

Renaissance Learning. (2010). *STAR Reading technical manual*. Wisconsin Rapids, WI: Renaissance Learning.

Rimm-Kaufman, S. E., Larsen, R. A., Baroody, A. E., Curby, T. W., Ko, M., Thomas, J. B., et al. (2014). Efficacy of the responsive classroom approach: Results from a 3-year, longitudinal randomized controlled trial. *American Educational Research Journal, 51*(3), 567–603. https://doi.org/10.3102/0002831214523821.

Schmidt, W. H., McKnight, C. C., & Raizen, S. A. (1997). *A splintered vision: an analysis of US mathematics and science curricula*. Paper presented at the US National Research Center for the Third International Mathematics and Science Study, Michigan State University, Lansing, MI.

Snider, V. E. (2004). A comparison of spiral versus strand curriculum. *Journal of Direct Instruction, 4*(1), 29–39.

Taylor, R. D., Oberle, E., Durlak, J. A., & Weissberg, R. P. (2017). Promoting positive youth development through school-based social and emotional learning interventions: A meta-analysis of follow-up effects. *Child Development, 88*(4), 1156–1171. https://doi.org/10.1111/cdev.12864.

Volpe, R. J., & DiPerna, J. C. (2010). Cooperative learning observation code for kids. Unpublished observation code.

What Works Clearinghouse. (n.d.). Find what works. https://ies.ed.gov/ncee/wwc/FWW.

What Works Clearinghouse. (2017). Procedures Handbook Version 4.0. Retrieved from https://ies.ed.gov/ncee/wwc/Docs/ReferenceResources/wwc_procedures_handbook_v4_draft.pdf.