Journal of Educational Psychology

A Cluster Randomized Trial of the Social Skills Improvement System-Classwide Intervention Program (SSIS-CIP) in First Grade

James Clyde DiPerna, Puiwa Lei, Weiyi Cheng, Susan Crandall Hart, and Jillian Bellinger

Online First Publication, March 23, 2017. http://dx.doi.org/10.1037/edu0000191

Journal of Educational Psychology

0022-0663/17/$12.00 http://dx.doi.org/10.1037/edu0000191

A Cluster Randomized Trial of the Social Skills Improvement System-Classwide Intervention Program (SSIS-CIP) in First Grade

James Clyde DiPerna, Puiwa Lei, Weiyi Cheng, Susan Crandall Hart, and Jillian Bellinger
The Pennsylvania State University

The purpose of this study was to evaluate the efficacy of a universal social skills program, the Social Skills Improvement System Classwide Intervention Program (SSIS-CIP; Elliott & Gresham, 2007), for students in first grade. Classrooms from 6 elementary schools were randomly assigned to treatment or business-as-usual control conditions. Teachers assigned to the treatment condition implemented the SSIS-CIP over a 12-week period. Students' social skills, problem behaviors, and approaches to learning were assessed via teacher ratings and direct observations of classroom behavior. In addition, their early literacy and numeracy skills were measured via computer-adaptive standardized tests. SSIS-CIP participation yielded small positive effects in students' social skills (particularly empathy and social engagement) and approaches to learning (academic motivation and engagement). Students' problem behaviors and academic skills, however, were unaffected by SSIS-CIP exposure.

Educational Impact and Implications Statement

The purpose of this study was to evaluate student outcomes associated with a classroom social skills program, the Social Skills Improvement System Classwide Intervention Program (SSIS-CIP; Elliott & Gresham, 2007). Participation in the SSIS-CIP yielded small positive effects in first grade students' empathy, social engagement, academic motivation, and academic engagement. Students' problem behaviors and academic skills, however, were unaffected by SSIS-CIP exposure. Although some outcomes were similar to an earlier study of the SSIS-CIP in second grade classrooms, the first grade findings were consistently smaller in magnitude. If these findings are replicated in future studies, educators and administrators contemplating adoption of the SSIS-CIP should consider prioritizing second grade for implementation of the program within the primary grades.

Keywords: SSIS-CIP, social skills, approaches to learning, social emotional learning, cluster randomized trial

Supplemental materials: http://dx.doi.org/10.1037/edu0000191.supp

The development of social-emotional competence is critical for young children's later success and well-being. Researchers have identified a number of key social-emotional skills within the school setting such as maintaining positive relationships with peers and adults, social problem solving, effectively communicating emotions, listening, and being attentive (Ashdown & Bernard, 2012; DiPerna, Volpe, & Elliott, 2005; Shonkoff & Philips, 2000). The National Academy of Sciences reported that 60% of children enter school with the cognitive skills needed to be successful, but only 40% have the social-emotional skills to succeed (Ashdown & Bernard, 2012). Evidence-based universal interventions focused on social, emotional, and academic competence represent a promising approach to promoting positive youth development (Bradshaw, Zmuda, Kellam, & Ialongo, 2009). Fostering students' social-emotional learning (SEL) using such approaches has been shown to improve both social and academic outcomes (Ashdown & Bernard, 2012).

There is a substantive body of research linking the development of social—emotional competence with school success (Blair & Raver, 2015; Denham et al., 2003; Miles & Stipek, 2006; Wentzel & Asher, 1995). Within the classroom, positive social-emotional skills enable children to develop constructive relationships with teachers and peers, accompanied by foundational learning-related attitudes and behaviors that allow them to become engaged in the many new tasks put before them (Denham, Way, Kalb, Warren-Khot, & Bassett, 2013). Children who demonstrate behavior in a manner consistent with classroom expectations, engage with in-

James Clyde DiPerna, Puiwa Lei, Weiyi Cheng, Susan Crandall Hart, and Jillian Bellinger, Department of Educational Psychology, Counseling, and Special Education, The Pennsylvania State University.

Jillian Bellinger is now a licensed psychologist in Austin, Texas.

James Clyde DiPerna is the lead author of the Academic Competence Evaluation Scales and receives a royalty from the publisher (Pearson).

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.

Correspondence concerning this article should be addressed to James Clyde DiPerna, Department of Educational Psychology, 105 CEDAR Building, The Pennsylvania State University, University Park, PA 16802. E-mail: jdiperna@psu.edu

struction, and persist with learning tasks exhibit higher levels of achievement in school (McClelland, Acock, & Morrison, 2006). Conversely, attention problems undermine effective learning and contribute to off-task behavior and reduced achievement (Hughes & Kwok, 2006). As such, social and behavioral competencies may be as important for children's later success as their early cognitive skills (Bub, 2009; Heckman, 2006).

Universal SEL and Students' Social, Emotional, and Academic Development

Universal intervention programs implemented within a classroom can complement academic instruction to promote children's social, emotional, cognitive, and academic skill development (Blair & Raver, 2015; Greenberg, Domitrovich, & Bumbarger, 2001; Linares et al., 2005). For example, Bierman et al. (2008) implemented a social-emotional and literacy intervention in Head Start classrooms that yielded moderate effects on students' literacy skills and social cognitions. Similarly, the Positive Action program (Flay, Allred, & Ordway, 2001) has demonstrated positive effects on students' behavior, school involvement, and achievement (Flay & Allred, 2003). With regard to long-term positive outcomes, implementation of the Good Behavior Game (Barrish, Saunders, & Wolf, 1969) and an enhanced academic curriculum in first grade resulted in higher scores on standardized achievement tests in twelfth grade, higher rates of high school graduation, and higher rates of college attendance (Bradshaw et al., 2009).

Beyond these studies of individual programs, there have been several meta-analyses of outcomes associated with SEL programs during the past decade. Nelson, Westhues, and MacLeod (2003) completed a meta-analysis of 34 universal SEL programs and reported small-to-moderate positive effects on K-8 students' cognitive development, social-emotional behavior, and parent-family wellness. More recent meta-analyses have provided support for the effectiveness of universal SEL programs for students as well. For example, January, Casey, and Paulson (2011) reported a small mean overall effect of classwide social skills interventions with a wide range of effects reported across studies. In a broader meta-analysis of 213 school-based, universal SEL programs, Durlak, Weissberg, Dymnicki, Taylor, and Schellinger (2011) reported small-to-moderate effects on students' positive social behavior in daily situations, as rated by students, teachers, parents, or independent observers. The study also reported larger mean effects for students' social-emotional skill performance, which included skills assessed via hypothetical scenarios, test situations, or structured tasks but did not include teacher ratings of students' behaviors in daily classroom settings. Finally, Durlak et al. (2011) also reported positive effects relative to students' academic performance, emotional distress, and conduct problems.

SEL Programs and the Primary Grades

As children move through the primary grades, they are establishing key social-emotional (e.g., self-regulation, peer relationships) and academic (e.g., letter naming, vocabulary skills) competencies necessary for later school success (Blair & Raver, 2015; Early, Pianta, & Cox, 1999). As such, universal programs incorporating evidence-based instruction and learning activities characterized by systemic, direct, intentional instruction are critical for supporting children's social, emotional, and educational outcomes during this critical developmental period (Bub, 2009; Durlak, Weissberg, & Pachan, 2010; Nelson et al., 2003). The Social Skills Improvement System-Classwide Intervention Program (SSIS-CIP; Elliott & Gresham, 2007) is a universal program developed for use within the general education classroom, and the program utilizes instructional strategies (e.g., reinforcement, modeling, role-playing, and problem-solving) grounded in several established theories of student learning (e.g., operant, social learning). The curriculum includes 10 instructional units targeting social-emotional skills. Each unit targets one specific skill (e.g., listening to others, asking for help, and getting along with others) and includes three brief lessons. Across the 10 skill units, five focus on cooperation skills, two feature self-control skills, and the remaining three units focus on assertion, responsibility, and empathy. (Additional details regarding the SSIS-CIP are provided in the Method section.)

Because the SSIS-CIP targets self-regulatory behaviors that have been shown to complement and enhance learning in classroom settings (e.g., Blair & Raver, 2015), its underlying theory of change postulates proximal, medial, and distal student outcomes resulting from exposure to the program. Specifically, proximal outcomes include improvement in the social and emotional skills explicitly taught within the SSIS-CIP curriculum. Medial outcomes consist of improvements in students' approaches to learning (academic motivation and engagement) and reductions in problematic classroom behaviors (e.g., acting out, inattention). Distal outcomes are positive changes in academic skills. The SSIS-CIP theory of change is not only grounded in theoretical models linking student behavior, approaches to learning, and academic achievement (e.g., DiPerna et al., 2005) but also consistent with the SEL outcomes framework specified by the Collaborative for Academic, Social, and Emotional Learning (2016).

Two other popular universal SEL programs that, like SSIS-CIP, focus on the promotion of social skills and academic readiness behaviors in young children are The Incredible Years Classroom Dinosaur Social Skills and Problem-Solving Curriculum (IYCD; Webster-Stratton & Reid, 2004) and Second Step (Committee for Children, 1992). An adaptation of a popular small-group clinic-based intervention, IYCD is designed for students ages 3–8; Second Step and SSIS-CIP are available in versions from preschool to early adolescence. Though all three curricula have secondary aims of reducing problem behaviors, they approach SEL through different curricular foci. IYCDs approach is broad, including lessons in anger management, problem solving, emotion language, and friendship skills, while Second Step emphasizes the importance of self-regulation, empathy, and problem-solving. In contrast, SSIS-CIP's curriculum focuses on specific discrete social skills identified by a nationally representative sample of teachers as critical for classroom success (Elliott & Gresham, 2007). All three programs are designed to be taught by general education teachers using a variety of instructional strategies including direct instruction, modeling, and role play. IYCD requires the most instructional time to complete (60 lessons of approximately 35–40 min each), while the early elementary versions of Second Step (22 lessons, 25–40 min each) and SSIS-CIP (30 lessons, 20–25 min each) are less time-intensive.

Several randomized trials have been published to date examining the efficacy of these three early elementary programs. With

regard to IYCD, Webster-Stratton, Jamila Reid, and Stoolmiller (2008) reported significant improvements in students' (preschool – Grade 1) emotional self-regulation skills, social competence, and conduct problems based on ratings by independent observers. Results demonstrated a pattern of differential effectiveness; specifically, students from classrooms with the most initial risk (i.e., lower school readiness skills and higher conduct problems at baseline) benefited most from the intervention (Webster-Stratton et al., 2008). Furthermore, IYCD was associated with statistically significant increases in preschool students' appropriate behavior, interest, and enthusiasm based on postobservation ratings of whole-class behavior (Baker-Henningham, Walker, Powell, & Gardner, 2009). It is important to note that in both of these trials IYCD was implemented in conjunction with the Incredible Years teacher training program, so the observed outcomes reflect the cumulative effects of both approaches.

Results from two randomized controlled trials indicate mixed support for Second Step's efficacy in early elementary classrooms, with outcomes varying across measurement methods (i.e., behavior ratings vs. direct observation). Grossman et al. (1997) found no significant differences for parent and teacher behavior ratings for second- and third-grade students assigned to intervention and control conditions, but reported significant decreases in observed physical aggression and increases in neutral/prosocial behavior for the Second Step group. In a large trial involving students in kindergarten through second grade (Low, Cook, Smolkowski, & Buntain-Ricklefs, 2015), Second Step yielded marginally significant small main effects in reducing behavior problems, increasing social-emotional skills, and improving skills for learning. However, moderation analyses indicated that effects varied by pretest scores. Specifically, students exhibiting more severe problem behaviors or lower levels of social-emotional skills at pretest demonstrated larger positive gains from Second Step participation than their peers with moderate or higher levels of initial skills (Low et al., 2015).

Efficacy of the SSIS-CIP has been tested in one randomized trial to date (DiPerna, Lei, Bellinger, & Cheng, 2015, 2016). This trial included 432 second-grade students, 38 classrooms, and multiple outcome measures (including teacher report and direct observation) consistent with the SSIS-CIP theory of change. Results indicated that second-grade students demonstrated greater gains in teacher ratings of overall social skills, communication, cooperation, responsibility, and empathy upon completion of the SSIS-CIP (Early Elementary Level) than students who were not exposed to the program (DiPerna et al., 2015). Although students in SSIS-CIP classrooms demonstrated fewer withdrawn behaviors, other internalizing and externalizing behaviors did not demonstrate statistically significant differences. In addition, students exposed to the SSIS-CIP demonstrated improvement in their academic motivation and engagement relative to their peers in nonimplementing (control) classrooms (DiPerna et al., 2016). Improvement in academic skills relative to peers, however, was only observed in mathematics for students receiving supplemental (Title 1) services. Similar to the Webster-Stratton et al. (2008) study of IYCD and Low et al. (2015) study of Second Step, the impact of the SSIS-CIP was moderated by pretest skill level with students demonstrating lower skills at pretest (social skills, academic motivation, and engagement) benefitting most from the program.

Purpose, Rationale, and Hypotheses

Although the SSIS-CIP Early Elementary Level curriculum was developed for Grades 1–3 and is being implemented in elementary schools across the United States, there has only been one study of its efficacy to date. In addition, this study focused solely on students in second grade, and outcomes may not generalize to younger students as a result of age-related differences in social and cognitive functioning. Given the importance (as well as paucity) of systematic replication in educational and psychological science (e.g., Makel & Plucker, 2014; Makel, Plucker, & Hegarty, 2012), the purpose of this study was to examine the efficacy of the SSIS-CIP in first grade classrooms. Specifically, we tested hypotheses informed by the SSIS-CIP theory of change, initial efficacy trial in second grade, and results of studies of other universal programs in the primary grades (Webster-Stratton et al., 2008; Low et al., 2015). The first hypothesis was that children in classrooms implementing the SSIS-CIP demonstrate improved social skills compared to children in nonimplementing (business-as-usual) control classrooms. Second, children in the SSIS-CIP condition were expected to demonstrate fewer problem behaviors than their peers in control classrooms. The third hypothesis was that children exposed to the SSIS-CIP demonstrate improved approaches to learning, and the fourth hypothesis was that SSIS-CIP students demonstrate improved academic skills relative to their peers. Finally, given the findings of the Grade 2 trial as well as Low et al.'s (2015) Second Step efficacy trial, we examined if SSIS-CIP effects with first grade students are moderated by their initial skill level in the target outcome variable (individually or at the class level.).

Method

Participants

Participating classrooms were drawn from six elementary schools in the Mid-Atlantic region of the United States. All first grade teachers in these six schools ($N = 61$) were invited to participate in the study; however, two teachers were unable to participate because of extended absences resulting from medical or family leave. As such, the total number of participating classrooms at the beginning of data collection was 59 (see Figure 1). Approximately half of these classrooms were from four schools in a small urban school district, and the remaining classrooms were from two elementary schools in a small rural district. Most of the classroom teachers reported extensive classroom experience ($M = 21.81$ years, $SD = 9.50$), and all were White and female.

All students from each participating classroom were invited to participate in the data collection associated with the efficacy trial, and approximately 56% ($N = 766$) received parental consent (see Figure 1). Thirty of the participating students (3.9%) moved before posttest data were collected, and 17 of these students were in the SSIS-CIP condition. An additional 40 students (5.2% of the initial participating student sample) were excluded from the final analyses because of missing data, and 32 of these students were from two classrooms where participating teachers were unable to com-

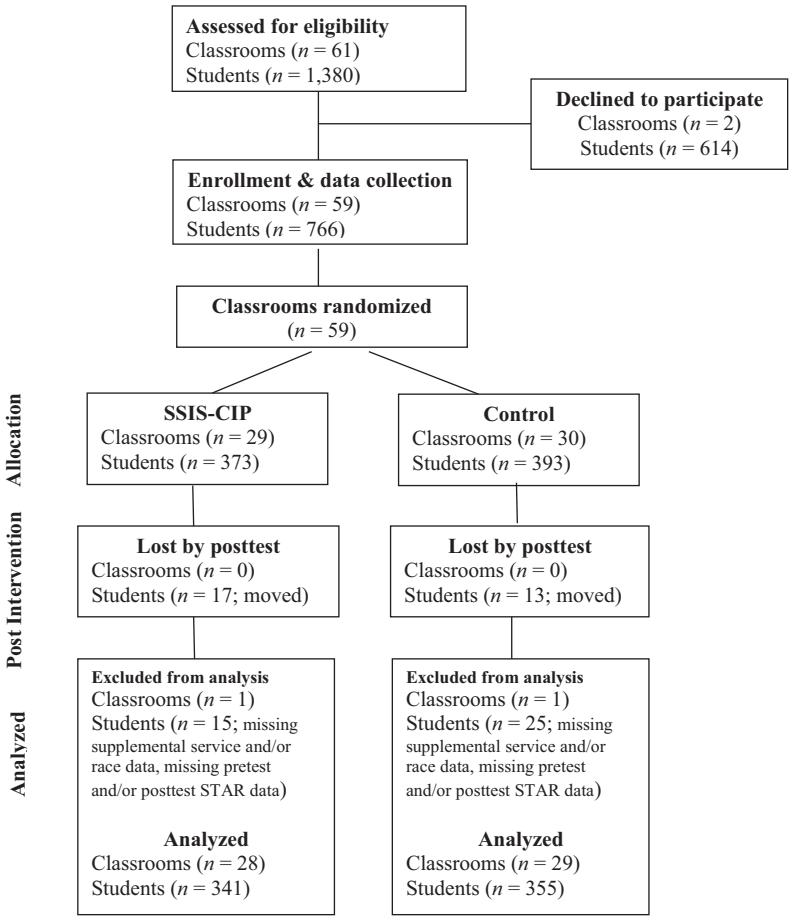

Figure 1. Participant flowchart.

plete the final round of data measures at the end of the academic year.

Demographic characteristics of the analyzed participant sample (see Table 1) were consistent with the first grade student population across the participating elementary schools. The mean age of the overall student sample was 6.29 years, and about half of students were male (53.3%). The racial composition of the student sample included White (70.1%) Black/African American (24%), Hawaiian-Pacific (0.4%), and Asian (4.9%), and approximately 9.4% of students also identified as Hispanic or Latino. A majority of students (94.1%) spoke English as their primary language. Students receiving special education services (via an Individualized Education Program) comprised 6.2% of the total sample, while 27.7% of students received academic support through supplemental services (i.e., Title 1). A small number of students (1.9%) were repeating first grade while enrolled in the study. None of the slight observed differences across demographic characteristics of the SSIS-CIP and control subsamples were statistically significant. All student and teacher participants were treated in accord with the ethical principles of the American Psychological Association.

Table 1
Student Demographic Characteristics by Treatment Condition

Variable	SSIS-CIP $N = 341$	Control $N = 355$
Age (in years)	6.29 (.42)	6.30 (.43)
Male	51.61	54.93
White	72.43	67.89
Black/African American	21.54	26.07
Hawaiian-Pacific	.90	0
Asian	4.42	5.28
Hispanic or Latino	9.21	9.51
Other race	5.36	3.09
English as primary language	94.72	93.52
Special education	4.40	7.89
Supplemental services	26.98	28.45
Repeating first grade	1.50	2.30

Note. SSIS-CIP = Social Skills Improvement System Classwide Intervention Program. Mean (*SD*) are reported for Age and percentage is reported for all other variables. There were no statistically significant differences ($p < .05$) on any of the demographic variables.

Measures

The measures used to assess the primary student outcomes of interest (social skills, problem behavior, approaches to learning, and academic skills) were identical to those used in the Grade 2 efficacy trial (DiPerna et al., 2015, 2016). Specifically, partici-

pants' social skills and problem behaviors were measured via teacher ratings on the Social Skills Improvement Rating Scales-Teacher Form (SSIS-RST; Gresham & Elliott, 2008) and direct observations using the Cooperative Learning Observation Code for Kids (CLOCK; Volpe & DiPerna, 2010). Participants' approaches to learning (academic motivation and engagement) were measured via teacher ratings on the Academic Competence Evaluation Scales (ACES; DiPerna & Elliott, 2000) and direct observation on the CLOCK. Students' academic skills were assessed via the STAR Reading and Math computerized adaptive tests (Renaissance Learning, 2009, 2010). Finally, the Classroom Assessment Scoring System (CLASS; Pianta, La Paro, & Hamre, 2008) was used to assess the instructional environment in each participating classroom before SSIS-CIP implementation.

Social skills and problem behavior. The SSIS-RST (Gresham & Elliott, 2008) was used to obtain teachers' perspectives of their students' social skills and problem behaviors in the classroom setting. The SSIS-RST Social Skills scale includes 46 items, seven subscales (Communication, Cooperation, Assertion, Responsibility, Empathy, Engagement, and Self-Control), and a total composite. The Problem Behaviors Scale includes 24 items across five subscales (Externalizing, Bullying, Hyperactive-Inattentive, Internalizing, and Autistic Behavior); however, the Autistic Behavior subscale was not analyzed in the current study. Teachers rate each item on the Social Skills and Problem Behaviors scales using a 4-point format ranging from *never* to *almost always*. Psychometric evidence for SSIS-RST scores is consistent with its intended purpose (Gresham & Elliott, 2008), and reliability indices (α = .88–.98) in the current sample are strong (see Table 2).

The CLOCK (Volpe & DiPerna, 2010) is a structured observation protocol that was used to facilitate independent direct observations of student social and problem behavior in the classroom. Specifically, the CLOCK category of Positive Social reflects any social behavior that is permitted during the observation interval, and Interference measures student problem behaviors that distract others or disrupt the functioning of the classroom. Each of these behaviors is observed using a partial interval format with each interval lasting 15 s. Six participants (three boys and three girls) were randomly identified within each classroom, and each was observed on three separate occasions during each data collection window. All direct observations were completed during mathematics to standardize the instructional context. (Both participating districts used Everyday Math curriculum, which features collaborative learning.)

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 hr 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 (κ = .88–.94) across all target behavior domains and paired observations (see Table 2).

Approaches to learning. The ACES (DiPerna & Elliott, 2000) was used to measure teacher perspectives regarding their students' approaches to learning (academic motivation and en-

Table 2
Reliability Indices and Intraclass Correlations for Social Skills, Problem Behaviors, Approaches to Learning, and Academic Skills

Variable	Reliability index		ICC (school)		ICC (class)	
	Pretest	Posttest	Pretest	Posttest	Pretest	Posttest
Social skills						
Social skills composite	.98	.98	.07	.07	.18	.16
Communication	.91	.91	.04	.04	.20	.25
Cooperation	.94	.94	.08	.07	.10	.06
Assertion	.88	.88	.09	.14	.23	.24
Responsibility	.92	.91	.06	.05	.13	.08
Empathy	.93	.95	.01	.04	.24	.21
Social engagement	.93	.94	.05	.04	.19	.20
Self-control	.93	.94	.07	.07	.15	.13
Positive social[a]	.88	.91	.01	.005	.15	.14
Problem behaviors						
Externalizing	.94	.94	.03	.01	.19	.16
Bullying	.90	.92	.04	.01	.20	.17
Hyperactive-inattentive	.90	.91	.06	.06	.15	.13
Internalizing	.89	.88	.04	.02	.32	.37
Interference[a]	.89	.93	.03	.01	.20	.27
Approaches to learning						
Academic motivation	.98	.98	.04	.07	.12	.09
Academic engagement	.96	.95	.04	.07	.11	.13
Engaged time[a]	.92	.94	.04	.08	.28	.32
Academic skills						
Math scaled score	—	—	.14	.11	.09	.09
Reading scaled score	—	—	.06	.09	.07	.07

Note. ICC = Intraclass correlation. Reliability indices are Cronbach's αunless noted otherwise.
[a] Direct observation data (κ agreement index reported for reliability).

gagement) in the classroom. The ACES Academic Motivation subscale includes 11 items that measure a student's approach, persistence, and level of interest regarding academic learning. The Academic Engagement subscale includes eight items that reflect attention and active participation in classroom activities. Items are rated using a 5-point format ranging from *never* to *almost always*. Psychometric evidence for ACES scores is consistent with its intended purpose (DiPerna & Elliott, 2000), and reliability estimates (α = .95–.98) from the current sample are strong (see Table 2).

Direct observation of students' engagement during instruction was also completed as part of the aforementioned CLOCK observations. The CLOCK category of 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 a partial interval format with each interval lasting 15 s.

Academic skills. The STAR Math (Renaissance Learning, 2009) and Reading (Renaissance Learning, 2010) computerized adaptive tests were used to directly measure students' academic skills. STAR Math is composed of a series of multiple choice mathematical problems that assess proficiency with numeration and computation objectives. STAR Reading features vocabulary-in-context items that require students to utilize background information, apply vocabulary, and use active strategies to construct meaning. Each STAR assessment requires approximately 10 min to complete, and both were administered before and after intervention implementation. STAR scores demonstrate high reliability and strong relations with other standardized achievement test scores (e.g., CA Achievement Test, Stanford Achievement Test) as well as teacher ratings of students' academic proficiency (Renaissance Learning, 2009, 2010).

Classroom instructional environment. Each participating classroom was observed once during the first data collection window to determine if there were significant differences in instructional environments across the participating classrooms, and if so, control for them in the tests of the hypotheses. The CLASS K-3 (Pianta et al., 2008) is a structured observation system that yields scores in three domains: Emotional Support, Classroom Organization, and Instructional Support. These broad domains are further differentiated across 10 dimensions (Positive Climate, Negative Climate, Teacher Sensitivity, Regard for Student Perspectives, Behavior Management, Productivity, Instructional Learning Formats, Concept Development, Quality of Feedback, and Language Modeling). Each dimension is rated on a 7-point scale ranging from *low* to *high*. Ratings are assigned after an observer completes an observation "cycle" (20 min of observation followed by 10 min of assigning ratings to dimensions/domains). Psychometric evidence for the CLASS is sound (Hamre, Mashburn, Pianta, & LoCasale-Crouch, 2008) and provides support for its intended purpose.

Observers were formally trained by a CLASS-certified instructor and achieved the CLASS-mastery criterion (>80% accuracy) before completing observations. Consistent with the authors' recommendations, two observation cycles were completed in each classroom to yield representative dimension and domain scores. Domain scores demonstrated acceptable levels of internal consistency (.81–.93). In addition, paired observations were completed for approximately 33% of the classrooms (N = 21), and intraclass correlations between these paired observations were moderate to high (.65–.76) for the CLASS domain scores.

Procedure

Recruitment. Data were collected as part of a multiyear project examining efficacy of the SSIS-CIP. First grade teachers within each elementary school were invited to participate in the project. Upon receiving teachers' consent, letters were distributed to parents requesting consent for their child's participation in the data collection process. A reminder letter was sent approximately four school days after the initial invitation letter.

Data collection. Both the business-as-usual control and treatment classrooms followed the same data collection schedule. Child-level data were collected during 4-week periods before (November–December) and after (March–April) SSIS-CIP implementation (average pre- and posttest interval = 4.8 months). Specifically, teachers completed the SSIS-RST and ACES subscales for all participating children. All participating teachers were compensated for time spent completing these scales. In addition, trained examiners administered the STAR assessments to all students with parental consent. (Verbal assent also was obtained from students before any testing commenced.) Research staff also completed CLOCK observations for the randomly selected subsample (3 boys and 3 girls) from each classroom. As noted previously, each student was observed during mathematics instruction on three separate days during the pre- and postdata collection periods. (If a student was absent on a scheduled observation day, the observer rescheduled the observation for a mathematics period later in the week.)

SSIS-CIP. The SSIS-CIP includes 10 instructional units focused on key classroom social skills 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 focuses on 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).

Intervention teachers (N = 29) completed a 1-day workshop with the lead author before implementation. The first half of the workshop provided a detailed overview of the SSIS-CIP lesson plans, student booklets, and video vignettes. During the second half, participants 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 the formal training, implementing teachers were expected to teach one SSIS-CIP unit (three lessons) per week.

Two complimentary methods were used to evaluate and ensure fidelity of implementation of the SSIS-CIP lessons. First, research staff completed direct observations for approximately 20% of the SSIS-CIP lessons taught in each classroom. Observations were completed approximately every other week to ensure lessons were sampled throughout the implementation period. For each fidelity observation, staff observed an 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 (introduce, define, discuss, identify steps and practice, and model/role-play) using a 4-point scale ranging from *not implemented* (1) to *full implementation* (4). In addition, implementing teachers completed weekly standardized checklists indicating their level of implementation (using the same 4-point scale as the observers) for the five core components of each lesson.

During the implementation period, fidelity (both self-report and independent observations) was monitored to ensure that teachers demonstrated at least 90% fidelity in their implementation of the SSIS-CIP lessons, and all SSIS-CIP teachers consistently met this threshold throughout the implementation period. In addition, the research team periodically checked with all teachers (approximately every other week) to see if they had any implementation questions, make sure nothing had arisen that would adversely impact implementation of the SSIS-CIP lessons, and thank them for their efforts. As a result of the scripted lesson format and these monitoring efforts, the SSIS-CIP program was fully implemented across all classrooms based on summative ratings by teachers ($M = 3.92$, $SD = .16$) and independent observers ($M = 3.97$, $SD = .08$).

Control condition. Teachers in classrooms randomly assigned to the control condition ($N = 30$) continued with their daily approach to managing and promoting positive classroom behavior throughout the duration of the study. Our primary reason for choosing a business-as-usual control condition is that most schools considering use of the SSIS-CIP likely would be doing so because their teachers do not have a similar program already in place. Thus, understanding magnitude of effect relative to typical classroom practices (i.e., locally developed behavior management plans) would be helpful to stakeholders when making decisions about adopting the SSIS-CIP. Approximately 85% of the control teachers reported having an explicit, planned approach to promoting positive behavior in their classroom. These plans primarily focused on the use of reinforcement (verbal praise or systems where students earned points for positive behavior toward tangible rewards) and/or consequences in the classroom. No teacher in the control condition reported use of a structured curriculum focused on the instruction of social skills.

Design and Data Analyses

This study used a multisite Cluster Randomized Trial (CRT) to test the efficacy of SSIS-CIP on each of the key outcome variables. Classrooms were randomly assigned to experimental conditions (SSIS-CIP and business-as-usual control) within schools. Figure 1 depicts the flow of classroom and student participants throughout the study. There was a low percentage of missing data (0.8–4.4% across variables), and these data were missing completely at random (Little's $\chi^2 = 496.76$, $df = 529$, $p > .05$). Given listwise deletion yields unbiased estimates under these conditions (Graham, 2009), cases were deleted listwise for analysis.

Multilevel modeling was used to account for the data structure of students nested within classes ($M = 12$ students/class, $SD = 4$). We first estimated unconditional models to report intraclass correlation (ICC) coefficients that indicated the degree to which the assumption of independence was violated because of the clustering of students in classes in schools (Raudenbush, 1997). Table 2 presents ICCs at both school- and class-levels for all outcome measures. Class-level ICCs for posttest outcome measures ranged from small (.06 for Cooperation) to large (.37 for Internalizing). These levels of ICCs suggested that standard errors might be underestimated if the nested data structure was not taken into account. Therefore, we analyzed a two-level model for each outcome to provide proper *SE* estimates. School-level variances of all posttest outcome measures were mostly small and statistically nonsignificant based on z tests (two-tailed ps $> .05$). However, school-level ICCs for several posttest measures (e.g., .14 for Assertion, .11 for Mathematics) were considered medium-sized (Raudenbush, Spybrook, Liu, & Congdon, 2005). Therefore, we included school indicators (dummy-coded) to control for school differences in all two-level models.

As recommended by What Works Clearinghouse (U.S. Department of Education, What Works Clearinghouse, 2016), we also tested for baseline equivalence between treatment and control conditions with respect to student demographic variables, classroom quality (i.e., CLASS scores), and all pretest measures. We then controlled for variables that showed nonequivalence when evaluating treatment effects. There were no statistically significant differences in CLASS scores between treatment and control classrooms. Moreover, there were no statistically significant differences between treatment and control conditions (based on two-level models) on pretest measures of problem behaviors, academic motivation, academic engagement, and academic skills.

Children in the control condition, however, had higher pretest scores on social skill measures than children in SSIS-CIP condition (Tables 3 and 4). As such, this baseline nonequivalence between conditions was addressed by including the social skills composite pretest (grand-mean centered) as an additional covariate for all outcome measures, including social skill subscales. Although social skill subscale scores were expected to be correlated with the composite, there was no substantial collinearity when the composite pretest was included as a covariate in the social skill subscale models. This was likely because each subscale (as one of seven) only contributed a small fraction to the composite variance. Because of the importance of statistically adjusting for nonequivalent baseline characteristics in the evaluation of treatment effects (U.S. Department of Education, What Works Clearinghouse, 2016) and the lack of multicollinearity, we decided to include the social skill composite pretest as a covariate even for social skill subscale outcomes.

To address the primary research questions regarding SSIS-CIP outcomes, we included both student- and class-level predictors to adjust for their effects. Student-level predictors included pretest scores of social skill composite (grand-mean centered) and the respective outcome measure (group-mean centered), students' gender (1 = male, 0 = female), race (1 = White, 0 = other), receipt of special education services (1 = yes, 0 = no), and receipt of

Table 3
Student-Level Means (SDs) by Measure, Time, and Treatment Condition

Variable	Pretest		Posttest	
	SSIS-CIP	Control	SSIS-CIP	Control
Social skills				
Social skills composite	1.92 (.51)	2.18 (.50)	2.20 (.51)	2.31 (.49)
Communication	2.00 (.62)	2.31 (.54)	2.29 (.56)	2.46 (.52)
Cooperation	1.90 (.67)	2.12 (.67)	2.13 (.68)	2.22 (.66)
Assertion	1.70 (.58)	1.94 (.60)	2.01 (.61)	2.14 (.57)
Responsibility	2.00 (.63)	2.24 (.61)	2.23 (.61)	2.33 (.60)
Empathy	1.97 (.58)	2.27 (.57)	2.31 (.60)	2.38 (.60)
Social engagement	2.00 (.59)	2.25 (.57)	2.30 (.56)	2.41 (.57)
Self-control	1.88 (.59)	2.11 (.62)	2.13 (.59)	2.21 (.63)
Positive social[a]	.36 (.69)	.39 (.67)	.31 (.58)	.28 (.52)
Problem behaviors				
Externalizing	.61 (.57)	.50 (.53)	.57 (.57)	.47 (.54)
Bullying	.34 (.50)	.27 (.46)	.34 (.52)	.24 (.45)
Hyperactive-inattentive	.89 (.67)	.78 (.65)	.80 (.67)	.69 (.64)
Internalizing	.55 (.55)	.51 (.51)	.51 (.53)	.46 (.49)
Interference[a]	.17 (.27)	.27 (.49)	.18 (.35)	.33 (.68)
Approaches to learning				
Academic motivation	3.28 (1.06)	3.42 (1.01)	3.60 (1.00)	3.62 (1.05)
Academic engagement	3.51 (1.02)	3.66 (.96)	3.91 (.88)	3.91 (.91)
Engaged time[a]	74.73 (13.92)	79.50 (13.75)	78.06 (14.49)	79.61 (13.99)
Academic Skills				
Math scaled score	257.00 (153.04)	250.95 (151.80)	356.05 (128.45)	351.32 (132.88)
Reading scaled score	88.35 (90.76)	81.41 (76.59)	152.10 (106.55)	140.10 (96.94)

Note. SSIS-CIP = Social Skills Improvement System Classwide Intervention Program. SSIS-CIP $N = 341$; control $N = 355$ (unless noted).
[a] Direct observation data (SSIS-CIP $N = 157$; control $N = 161$).

supplemental (Title 1) services (1 = yes, 0 = no). The dummy variable predictors were grand-mean centered. Class-level predictors included grand-mean centered class average of pretest scores of the respective outcome measure. Treatment efficacy was tested at the class-level using dummy codes for experimental conditions (1 = SSIS-CIP, 0 = control). In addition to testing for treatment main effects, we examined if SSIS-CIP effects were moderated by pretest scores (both class- and student-levels) by adding product terms to the main effects model.[1]

Our centering approaches were chosen based on recommendations by Enders and Tofighi (2007). Specifically, all student-level covariates (except for the pretest of the outcome measure), including dummy demographic variables, were grand-mean centered to obtain adjusted treatment effects for these covariates. We modeled the pretest of the outcome in both the student- and class-levels because we were interested in the differential effect of pretest at these levels. Although group- and grand-mean centering the student-level pretest in this case would be equivalent (Enders & Tofighi, 2007), we chose to group-mean center the student-level and grand-mean center the class-level because the effect of pretest would be directly decomposed into between- and within-class levels (Raudenbush & Bryk, 2002). Moreover, group-mean centering student-level variables has been recommended in testing cross-level interactions (i.e., between treatment and student-level pretest) to minimize the risk of finding significant interactions that did not exist (e.g., Hofmann & Gavin, 1998).

We estimated multilevel models using the Mixed procedure of SAS (version 9.3) for teacher ratings of social skills and approaches to learning as well as for direct assessments of academic skills (mathematic and reading). We used the SAS Glimmix procedure for teacher ratings of problem behaviors and all direct observation data. Because problem behaviors were observed infrequently and observations consisted of frequency data that were highly skewed, we used Poisson distribution and log link for the Glimmix procedure.

For concerns of incorrect Type 1 error or false discovery rate in testing intervention effects across multiple outcome measures, we followed the What Works Clearinghouse (U.S. Department of Education, What Works Clearinghouse, 2016) recommendation and applied the Benjamini-Hochberg correction (Benjamini & Hochberg, 1995) for teacher-rated social skills, problem behavior, and approaches to learning. Specifically, p values for treatment effects from the main effect models within the same outcome measure (e.g., teacher-rated social skills) were sorted from the lowest to the highest and compared with the corresponding critical p value. Critical p value was computed by multiplying alpha by the rank of the corresponding p value divided by the number of outcomes (i.e., the lowest p value was compared with .05*1/number of outcomes, the second lowest p value was compared to .05*2/number of outcomes, etc.). The highest p value that was less than or equal to the corresponding critical p value was the cutpoint, and findings with p values smaller than or equal to

[1] Consistent with our original proposal to the Institute of Education Sciences, we also examined if SSIS-CIP effects were moderated by student gender, minority status, and special education. Results of these analyses are reported in the supplement to this article.

Table 4
Class-Level Means (SDs) by Measures, Time, and Treatment Condition

Variable	Pretest		Posttest	
	SSIS-CIP	Control	SSIS-CIP	Control
Social skills				
Social skills composite	1.93 (.27)	2.19 (.28)	2.22 (.30)	2.31 (.27)
Communication	2.02 (.34)	2.31 (.28)	2.30 (.34)	2.45 (.31)
Cooperation	1.92 (.33)	2.14 (.32)	2.15 (.31)	2.22 (.29)
Assertion	1.71 (.31)	1.95 (.39)	2.02 (.42)	2.16 (.37)
Responsibility	2.01 (.32)	2.27 (.32)	2.25 (.28)	2.34 (.29)
Empathy	1.97 (.32)	2.27 (.31)	2.31 (.36)	2.39 (.33)
Social engagement	2.01 (.31)	2.25 (.32)	2.33 (.31)	2.42 (.32)
Self-control	1.91 (.31)	2.13 (.32)	2.15 (.36)	2.22 (.32)
Positive social[a]	.36 (.42)	.41 (.32)	.31 (.33)	.28 (.27)
Problem behaviors				
Externalizing	.60 (.29)	.49 (.29)	.55 (.27)	.49 (.28)
Bullying	.33 (.26)	.26 (.26)	.32 (.25)	.25 (.24)
Hyperactive-inattentive	.87 (.34)	.77 (.35)	.78 (.33)	.72 (.34)
Internalizing	.54 (.34)	.52 (.34)	.47 (.33)	.48 (.34)
Interference[a]	.16 (.13)	.28 (.32)	.19 (.23)	.33 (.43)
Approaches to learning				
Academic motivation	3.28 (.46)	3.44 (.57)	3.64 (.44)	3.61 (.54)
Academic engagement	3.54 (.48)	3.66 (.50)	3.94 (.47)	3.90 (.50)
Engaged time[a]	75.39 (9.91)	79.69 (8.16)	77.60 (10.67)	79.56 (10.06)
Academic skills				
Math scaled score	260.63 (84.57)	250.48 (76.78)	359.23 (65.23)	354.83 (64.28)
Reading scaled score	89.59 (42.42)	80.08 (31.62)	151.31 (53.32)	138.41 (46.25)

Note. SSIS-CIP = Social Skills Improvement System Classwide Intervention Program. SSIS-CIP $N = 28$; control $N = 29$.
[a] Direct observation data.

the cutpoint were declared statistically significant after the Benjamini-Hochberg correction.

Given the growing consensus that interpretation of study outcomes should not focus exclusively on statistical significance (e.g., Cumming, 2014; Durlak, 2009; Greenwald, Gonzalez, Harris, & Guthrie, 1996; Wasserstein & Lazar, 2016), we also estimated effect sizes of treatment as compared to the control condition based on the previously specified main effects models. Specifically, we computed the effect size 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 *SD* of the pretest outcome measure. This effect size computation (i.e., using student-level *SD* to standardize the adjusted difference for Hedges' g) followed the guidelines of What Works Clearinghouse (U.S. Department of Education, What Works Clearinghouse, 2016) for results from Hierarchical Linear Model analyses in studies featuring cluster-level assignment. Pooled within-group *SD* of pretest scores was used because pretest scores were not affected by treatment. Moreover, 95% confidence intervals (CIs) were calculated for each effect size to provide insight regarding the precision of the estimate and range of possible effects. We also calculated an improvement index for each outcome variable following the U.S. Department of Education, What Works Clearinghouse (2016) guidelines to help readers better understand the practical impact of the SSIS-CIP intervention. An improvement index indicates the expected percentile rank improvement for an average student in the control group had the student received the intervention.

Results

Given the primary focus of the SSIS-CIP is to promote social skills in the classroom, our first hypothesis was that students in intervention classrooms would demonstrate improved social skills relative to their peers in control classrooms. Parameter estimates for the multilevel main effects models for the social skills domain are reported in Table 5. As shown in the table, two of the social skill domains (empathy and social engagement) were statistically significant ($p < .05$); however, the remaining teacher-rated social skills (communication, cooperation, assertion, responsibility, and self-control) and direct observation variable (positive social) did not demonstrate statistically significant differences ($ps > .05$). After applying the Benjamini-Hochberg (Benjamini & Hochberg, 1995) correction to control for false discovery rate, none of the observed differences met the adjusted threshold criterion for statistical significance. Tests of interactions between baseline level of social skills and intervention condition also did not demonstrate statistical significance ($ps > .05$) in any of the social skills domains (see Table 6). After controlling for pretest scores, gender, race, receipt of special education services, and receipt of supplemental (Title 1) services, SSIS-CIP participation yielded positive effect sizes (adjusted standardized difference) and at least five percentile rank improvement across all teacher-rated social skills subscales, with empathy and social engagement demonstrating the largest positive effects (see Table 7). The direct observation measure (positive social), however, yielded the smallest effect size and improvement index.

Table 5
Mixed Model Estimates (SEs) for SSIS-CIP Effect on Social Skills Outcomes

Predictor	Teacher rating								Direct observation
	Social skills composite	Commun.	Cooper.	Assertion	Responsib.	Empathy	Social engage.	Self-control	Positive social[a]
Intercept	2.13** (.05)	2.32** (.09)	2.06** (.06)	1.87** (.07)	2.16** (.06)	2.21** (.07)	2.25** (.06)	2.06** (.07)	−1.66** (.30)
Covariates									
Student-level pretest	70** (.03)	.21** (.07)	.72**(.05)	.56** (.03)	.65** (.06)	.36** (.06)	.37** (.06)	.61** (.05)	−.26 (.13)
Class-level pretest	−.003 (.11)	.46** (.13)	.64** (.11)	.70** (.11)	.50** (.12)	.52** (.11)	.50** (.12)	.59** (.12)	.78** (.24)
Social skills pretest	—	.45** (.08)	.01(.07)	.07 (.04)	.10 (.08)	.36** (.07)	.35** (.07)	.09 (.06)	.26 (.24)
Gender	−.06* (.02)	−.04 (.03)	−.11**(.03)	−.05 (.03)	−.06* (.03)	−.06 (.03)	−.01 (.03)	−.06 (.03)	.21 (.22)
White	−.001 (.03)	−.01 (.04)	−.01(.04)	.04 (.04)	−.003 (.04)	−.005 (.04)	−.01 (.04)	−.002 (.04)	.19 (.27)
Supp. services	−.01 (.03)	.02 (.03)	−.03(.04)	−.08* (.03)	−.02 (.04)	.03 (.04)	−.02 (.03)	.03 (.04)	.05 (.27)
Special education	.05 (.05)	.01 (.06)	.13(.07)	−.03 (.06)	.11 (.06)	.07 (.07)	−.01 (.06)	.07 (.07)	.42 (.41)
Treatment effect									
SSIS-CIP	.09 (.05)	.09 (.07)	.09 (.05)	.08 (.07)	.08 (.06)	.18 (.07)	.12 (.06)	.09 (.06)	.13 (.23)
	$p = .098$	$p = .199$	$p = .099$	$p = .247$	$p = .147$	$p = .012$	$p = .045$	$p = .152$	$p = .568$
Random effects									
Intercept variance	.03** (.01)	.04** (.01)	.02** (.01)	.04** (.01)	.03** (.01)	.04** (.01)	.03** (.01)	.03** (.01)	.07 (.13)
Residual variance	.09** (.005)	.12** (.01)	.17** (.01)	.12** (.01)	.15** (.01)	.16** (.01)	.13** (.01)	.16** (.01)	—

Note. School indicators are included in the model but not reported. SSIS-CIP = Social Skills Improvement System Classwide Intervention Program; Commun. = communication; Cooper. = cooperation; Responsib. = responsibility; Social engage. = social engagement; Supp. services = supplemental services.
[a] Outcome variable is on log scale.
* $p < .05$. ** $p < .01$.

Our second hypothesis was that children exposed to the SSIS-CIP demonstrate fewer problem behaviors than their peers in control classrooms. Parameter estimates for the multilevel main effects models for problem behavior outcomes are reported in Table 8, and none of the differences between conditions on any of the problem behavior subscales were statistically significant ($ps > .05$). Similarly, tests of interactions between baseline level of problem behaviors and intervention condition also were not statistically significant (see Table 9). Effect size estimates for the problem behaviors subscales were consistently close to 0 with improvement indexes <3 (see Table 7).

Table 6
Mixed Model Estimates (SEs) for SSIS-CIP and Pretest Interaction on Social Skills Outcomes

Predictor	Teacher rating								Direct observation
	Social skills composite	Commun.	Cooper.	Assertion	Responsib.	Empathy	Social engage.	Self-control	Positive social[a]
Intercept	2.14** (.06)	2.33** (.08)	2.06** (.06)	1.89** (.07)	2.16** (.06)	2.20** (.07)	2.25** (.07)	2.07** (.07)	−1.70** (.30)
Covariates									
Student-level pretest	.71** (.04)	.20* (.08)	.71** (.06)	.53** (.04)	.65** (.07)	.39** (.07)	.38** (.07)	.63** (.06)	−.32 (.26)
Class-level pretest	−.08 (.14)	.43* (.17)	.63** (.13)	.60** (.13)	.49** (.14)	.53** (.15)	.50** (.14)	.53** (.15)	1.21* (.55)
Social skills pretest	—	.71** (.04)	.01 (.07)	.07 (.04)	.10 (.08)	.36** (.07)	.35** (.07)	.10 (.06)	.31 (.25)
Gender	−.06* (.02)	−.03 (.03)	−.11** (.03)	−.05 (.03)	−.06* (.03)	−.06* (.03)	−.02 (.03)	−.06 (.03)	.21 (.22)
White	−.001 (.03)	−.01 (.04)	−.01 (.04)	.05 (.04)	−.003 (.04)	−.004 (.04)	−.01 (.04)	−.002 (.04)	.22 (.28)
Supp. services	−.01 (.03)	.02 (.03)	−.04 (.04)	−.08* (.03)	−.02 (.04)	.03 (.04)	−.02 (.03)	.03 (.04)	.05 (.27)
Special education	.05 (.05)	.01 (.06)	.13 (.07)	−.03 (.06)	.10 (.07)	.07 (.07)	−.01 (.06)	.08 (.07)	.41 (.42)
Treatment effect									
SSIS-CIP	.09 (.05)	.09 (.07)	.09 (.06)	.09 (.07)	.08 (.06)	.18* (.07)	.12* (.06)	.09 (.06)	.21 (.25)
Interaction effect									
SSIS-CIP*Student- level pretest	−.02 (.05)	.01 (.05)	.02 (.05)	.06 (.05)	.001 (.05)	−.05 (.06)	−.03 (.06)	−.05 (.06)	.08 (.30)
	$p = .773$	$p = .806$	$p = .685$	$p = .281$	$p = .982$	$p = .382$	$p = .589$	$p = .424$	$p = .781$
SSIS-CIP*Class- level pretest	.18 (.19)	.05 (.22)	.02 (.17)	.29 (.20)	.02 (.18)	−.03 (.20)	.003 (.19)	.12 (.19)	−.53 (.62)
	$p = .349$	$p = .816$	$p = .883$	$p = .156$	$p = .893$	$p = .865$	$p = .987$	$p = .541$	$p = .399$
Random effects									
Intercept variance	.03** (.01)	.04** (.01)	.02** (.01)	.04** (.01)	.03** (.01)	.04** (.01)	.03** (.01)	.03** (.01)	.08 (.14)
Residual variance	.09** (.005)	.12** (.01)	.17** (.01)	.12** (.01)	.15** (.01)	.16** (.01)	.13** (.01)	.16** (.01)	—

Note. School indicators are included in the model but not reported. SSIS-CIP = Social Skills Improvement System Classwide Intervention Program; Commun. = communication; Cooper. = cooperation; Responsib. = responsibility; Social engage. = social engagement; Supp. services = supplemental services.
[a] Outcome variable is on log scale.
* $p < .05$. ** $p < .01$.

Table 7
Standardized Group Differences, 95% Confidence Intervals, and Improvement Indices

Variable	Effect size[a]	95% Confidence interval	Improvement index (%)
Social skills			
Social skills composite	.18	[.03, .33]	7.14
Communication	.16	[.01, .30]	6.36
Cooperation	.14	[−.01, .29]	5.57
Assertion	.13	[−.02, .28]	5.17
Responsibility	.14	[.00, .29]	5.57
Empathy	.31	[.16, .46]	12.17
Social engagement	.21	[.06, .36]	8.32
Self-control	.15	[.00, .3]	5.96
Positive social[b]	.05	[−.17, .27]	1.99
Problem behaviors			
Externalizing	−.04	[−.19, .11]	−1.60
Bullying	.01	[−.14, .16]	.40
Hyperactive-inattentive	−.03	[−.18, .12]	−1.20
Internalizing	−.03	[−.18, .12]	−1.20
Interference[b]	−.07	[−.29, .15]	−2.79
Approaches to learning			
Academic motivation	.17	[.02, .32]	6.75
Academic engagement	.17	[.02, .32]	6.75
Engaged time[b]	.13	[−.09, .35]	5.17
Academic skills			
Math scaled score	.04	[−.11, .19]	1.60
Reading scaled score	.07	[−.08, .22]	2.79

[a] Standardized difference adjusted for pretest and other student- and class-level covariates. [b] Direct observation data.

The third hypothesis was that children exposed to the SSIS-CIP demonstrate improved approaches to learning (i.e., academic motivation and engagement), and parameter estimates for the multilevel main effects models for the approaches to learning domain are reported in Table 10. Teacher-ratings of academic motivation and engagement were statistically significant (*ps* < .05), and remained statistically significant after the Benjamini-Hochberg correction. Neither direct observation of engaged time (see Table 10) nor tests of interactions between pretest level of skills and intervention (see Table 11) were statistically significant. Effect size estimates indicated positive effects of SSIS-CIP exposure on academic motivation and engagement with improvement indexes >5 (see Table 7). Finally, our fourth hypothesis was that SSIS-CIP students demonstrate improved academic skills (reading and mathematics) relative to their peers. Parameter estimates for direct (see Table 10) and interaction effects (see Table 11) were not statistically significant for either reading or mathematics. Effect size estimates were slightly positive but close to 0, with improvement indexes below 3 (see Table 7).

Discussion

The purpose of this project was to examine social, behavioral, and academic outcomes resulting from implementation of a universal social skills program in primary classrooms. Specifically, first grade classrooms within six schools were randomly assigned to treatment or business-as-usual control conditions. Teachers in the classrooms assigned to treatment were formally trained and implemented the SSIS-CIP over a 12-week period. Outcomes were assessed via teacher ratings and direct observations of classroom behavior as well as computer-adaptive tests of reading and mathematics. Multiple indices (*p* values, effect sizes, CIs, and improvement indexes) were examined to draw conclusions about hypotheses.

SSIS-CIP Outcomes for Students in First Grade

With regard to the first hypothesis, none of the observed differences across the social skills subscales met a $p < .05$ threshold after correcting for false discovery rate using the Benjamini-Hochberg procedure. Effect sizes, CIs, and improvement indexes, though, suggest that the SSIS-CIP generally has small positive

Table 8
Multilevel Model Estimates (SEs) for SSIS-CIP Effect on Problem Behaviors Outcomes

	Teacher rating				Direct observation
Predictor	Externalizing	Bullying	Hyperactive-inattentive	Internalizing	Interference
Intercept	−.94** (.12)	−1.76** (.19)	−.52** (.10)	−1.05** (.14)	−1.97** (.39)
Covariates					
Student-level pretest	.87** (.13)	.82** (.15)	.73** (.10)	.68** (.15)	.36 (.23)
Class-level pretest	1.27** (.21)	1.73** (.32)	.96** (.17)	1.61** (.19)	2.47** (.51)
Social skills pretest	−.21 (.17)	−.63** (.20)	−.11 (.14)	−.40* (.16)	−.53* (.27)
Gender	.14 (.12)	−.07 (.16)	.20* (.10)	.05 (.12)	.70** (.26)
White	.08 (.13)	.09 (.18)	.08 (.11)	.003 (.14)	.17 (.29)
Supplemental services	.04 (.12)	−.04 (.17)	.07 (.10)	.01 (.12)	−.06 (.29)
Special education	−.14 (.21)	.04 (.30)	−.13 (.18)	−.01 (.22)	−.75 (.53)
Treatment effect					
SSIS-CIP	−.05 (.11)	.03 (.17)	−.03 (.10)	−.04 (.12)	−.18 (.32)
	$p = .652$	$p = .849$	$p = .751$	$p = .722$	$p = .578$
Random effects					
Intercept variance	<.0001	.05 (.070)	<.0001	.02 (.03)	.18 (.15)

Note. SSIS-CIP = Social Skills Improvement System Classwide Intervention Program. Outcome variables are on log scale. School indicators are included in the model but not reported.
* $p < .05$. ** $p < .01$.

Table 9
Multilevel Model Estimates (SEs) for SSIS-CIP and Pretest Interaction Effect on Problem Behaviors Outcomes

Predictor	Teacher rating			Direct observation	
	Externalizing	Bullying	Hyperactive-inattentive	Internalizing	Interference
Intercept	−.97** (.13)	−1.74** (.19)	−.55** (.11)	−1.01** (.14)	−1.91** (.39)
Covariates					
Student-level pretest	1.00** (.17)	.79** (.20)	.82** (.12)	.66** (.21)	.36 (.25)
Class-level pretest	1.24** (.31)	1.35** (.50)	.98** (.24)	1.38** (.28)	2.28 (.53)
Social skills pretest	−.21 (.17)	−.65** (.20)	−.12 (.14)	−.43** (.16)	−.52 (.27)
Gender	.16 (.12)	−.09 (.16)	.20* (.10)	.05 (.12)	.70 (.26)
White	.08 (.13)	.09 (.18)	.10 (.11)	−.001 (.14)	.19 (.29)
Supplemental services	.02 (.12)	−.04 (.17)	.06 (.10)	.01 (.12)	−.06 (.29)
Special education	−.19 (.22)	.04 (.30)	−.16 (.18)	−.003 (.22)	−.77 (.54)
Treatment effect					
SSIS-CIP	.01 (.13)	−.05 (.20)	.02 (.11)	−.12 (.15)	−.23 (.32)
Interaction effect					
SSIS-CIP*Student-level pretest	−.20 (.17)	.05 (.24)	−.14 (.13)	.01 (.23)	.07 (.67)
	$p = .236$	$p = .841$	$p = .288$	$p = .957$	$p = .916$
SSIS-CIP*Class-level pretest	.03 (.39)	.61 (.62)	−.03 (.29)	.47 (.41)	1.83 (1.69)
	$p = .941$	$p = .324$	$p = .905$	$p = .252$	$p = .281$
Random effects					
Intercept variance	<.0001	.05 (.07)	<.0001	.01 (.03)	.17 (.15)

Note. SSIS-CIP = Social Skills Improvement System Classwide Intervention Program. Outcome variables are on log scale. School indicators are included in the model but not reported.
* $p < .05$. ** $p < .01$.

effects on the social skills of first grade students. Specifically, communication, cooperation, assertion, responsibility, and self-control all yielded effect sizes between .13–.16 with CIs ranging from approximately 0–.30 and improvement indexes in the 5–6 range. Social engagement and empathy had somewhat larger effect size estimates, higher CI bounds, and improvement indexes. (Conversely, direct observation of positive social behavior demonstrated a smaller effect size and improvement index.) The observed pattern of findings in first grade is similar to the pattern observed in the second grade trial of the SSIS-CIP (DiPerna et al., 2015). It is important to note, though, that effect sizes and improvement indexes for the first grade sample are consistently smaller (by approximately half) than those observed for second grade. The lone exception is empathy, where the effect sizes and corresponding CIs are almost identical (and more moderate in magnitude) across both trials.

Although the adjusted standardized group differences for social skills would be classified as small under Cohen's, 1988 guidelines, methodologists (Durlak, 2009; Ferguson, 2009) caution against rigid application of such guidelines and encourage interpretation

Table 10
Multilevel Model Estimates (SEs) for SSIS-CIP Effect on Approaches to Learning and Academic Skills Outcomes

Predictor	Teacher rating		Direct observation	Direct assessment	
	Academic motivation	Academic engagement	Engaged time[a]	Math	Reading
Intercept	3.38** (.08)	3.64** (.09)	4.35** (.03)	365.02** (12.01)	158.17** (6.06)
Covariates					
Student-level pretest	.69** (.04)	.57** (.03)	.002* (.001)	.45** (.03)	.84** (.03)
Class-level pretest	.65** (.09)	.61** (.09)	.01** (.002)	.57** (.09)	.98** (.09)
Social skills pretest	.10 (.08)	.16* (.06)	.09** (.02)	45.17** (7.94)	19.67** (4.94)
Gender	−.15** (.05)	−.08 (.04)	−.04** (.01)	15.19* (7.13)	−8.59 (4.54)
White	−.03 (.06)	.08 (.06)	.04* (.02)	14.66 (9.25)	13.71* (5.88)
Supplemental services	−.19** (.06)	−.10 (.05)	.01 (.02)	−30.88** (8.50)	−28.03** (5.28)
Special education	−.05 (.10)	−.06 (.09)	−.01 (.03)	−21.36 (14.95)	−5.03 (9.52)
Treatment effect					
SSIS-CIP	.18 (.08)	.17 (.08)	.02 (.03)	5.61 (10.04)	5.87 (5.49)
	$p = .025$	$p = .035$	$p = .389$	$p = .579$	$p = .290$
Random effects					
Intercept variance	.04** (.01)	.06** (.02)	.01 (.002)	649.75** (265.82)	1180.63 (75.92)
Residual variance	.35** (.02)	.29** (.02)	—	7820.35** (438.76)	3212.35** (179.73)

Note. SSIS-CIP = Social Skills Improvement System Classwide Intervention Program. School indicators are included in the model but not reported.
[a] Outcome variable is on log scale.
* $p < .05$. ** $p < .01$.

Table 11
Multilevel Model Estimates (SEs) for SSIS-CIP and Pretest Interaction Effect on Approaches to Learning and Academic Skills Outcomes

Predictor	Teacher rating		Direct observation	Direct assessment	
	Academic motivation	Academic engagement	Engaged time[a]	Math	Reading
Intercept	3.38** (.08)	3.64** (.09)	77.23** (2.43)	365.22** (12.16)	158.40** (6.05)
Covariates					
Student-level pretest	.72** (.05)	.58** (.04)	.11 (.08)	.48** (.04)	.83** (.04)
Class-level pretest	.65** (.10)	.55** (.12)	.74** (.18)	.58** (.12)	1.08** (.15)
Social skills pretest	.10 (.08)	.16* (.06)	6.44** (1.49)	45.46** (7.95)	19.58** (4.95)
Gender	−.16** (.05)	−.08 (.04)	−2.94* (1.23)	14.48* (7.15)	−8.37 (4.56)
White	−.04 (.06)	.08 (.06)	3.00 (1.60)	14.43 (9.25)	13.48* (5.90)
Supplemental services	−.19** (.06)	−.10 (.05)	.37 (1.57)	−30.66** (8.50)	−28.16** (5.29)
Special education	−.03 (.10)	−.05 (.09)	−.33 (2.68)	−20.66 (14.97)	−5.07 (9.56)
Treatment effect					
SSIS-CIP	.18* (.08)	.17* (.08)	1.97 (2.08)	5.71 (10.14)	5.83 (5.48)
Interaction effect					
SSIS-CIP*Student-level pretest	−.06 (.05)	−.01 (.05)	.07 (.11)	−.07 (.05)	.02 (.06)
	$p = .204$	$p = .805$	$p = .542$	$p = .186$	$p = .791$
SSIS-CIP*Class-level pretest	.01 (.16)	.14 (.17)	−.20 (.24)	−.02 (.13)	−.14 (.16)
	$p = .938$	$p = .434$	$p = .420$	$p = .847$	$p = .390$
Random effects					
Intercept variance	.05** (.02)	.06** (.02)	34.96** (11.46)	675.77** (275.00)	115.86 (77.10)
Residual variance	.35** (.02)	.29** (.02)	105.52** (9.38)	7811.71** (438.74)	3219.67** (180.41)

Note. SSIS-CIP = Social Skills Improvement System Classwide Intervention Program. School indicators are included in the model but not reported.
[a] Outcome variable is on log scale.
* $p < .05$. ** $p < .01$.

relative to methodology and previous studies of similar interventions. Relative to randomized trials of other universal SEL programs, SSIS-CIP effect sizes based on teacher ratings of students' social skills are similar in magnitude to those reported for Second Step ($g = -.016$–$.126$; Low et al., 2015). Direct comparison with IYCD is difficult, as randomized trials evaluating its efficacy have included implementation of other Incredible Years interventions in conjunction with IYCD and outcome measures have not included teacher ratings of social skills (Baker-Henningham et al., 2009; Webster-Stratton et al., 2008). The 95% CI ranges for all of the SSIS-CIP social skills outcomes overlap with the mean effect sizes for student prosocial behavior reported by two recent meta-analyses (Durlak et al., 2011; January et al., 2011).

With regard to problem behaviors, results were inconsistent with our hypothesis that exposure to the SSIS-CIP would yield reductions in these behaviors. All tests of significance in this domain were nonsignificant with effect sizes and improvement indexes close to 0, and 95% effect size CIs that extended in both the positive and negative directions. Similar to the social skills domain, the problem behavior effect sizes observed in the first grade trial are smaller than those observed in the second grade trial that ranged from −.08 (improvement index = 3.19) for bullying to −.24 (improvement index = 9.48) for internalizing behavior (DiPerna et al., 2015). Relative to previous research, the SSIS-CIP first grade effect sizes are lower than those reported for problem behavior in the Durlak et al. (2011) meta-analysis ($g = .22$–$.24$). The studies in their review, however, included a much broader range of behavioral (e.g., noncompliance, aggression, and school suspensions) and emotional problems (e.g., stress, depression, and anxiety) than the measures in the current study. Results from the current study are more similar to those from the recent efficacy trial for Second Step (Low et al., 2015) in which specific problem behavior outcomes were measured through teacher ratings and direct observations. The Second Step effect size for conduct problems ($g = -.04$) was equivalent to SSIS-CIP effect size for externalizing behaviors, and although the other Second Step problem behavior outcomes effect sizes were larger ($-.109 < g < -.067$), they still fell within the 95% CI ranges for similar problem behavior outcomes in the current study. Similar to the current study, Webster-Stratton et al. (2008) reported a small mean effect size (.032) for IYCD on conduct problems measured via direct observation.

Our third hypothesis was that SSIS-CIP improves young students' approaches to learning (academic motivation and engagement), and results were consistent with this hypothesis. Teacher ratings of academic motivation and engagement remained statistically significant after the Benjamini-Hochberg correction and yielded small effect sizes with positive CIs and improvement indexes. As with the social skills domain, the first grade effect sizes and improvement indexes based on teacher ratings of academic motivation and engagement were approximately half the magnitude of those in the second grade trial (DiPerna et al., 2016). In contrast to all other findings, the effect size for direct observation of engagement in instruction was larger in the current study (.13) than the second grade trial (.03). Unfortunately, few CRTs of other universal SEL interventions (Durlak et al., 2011; January et al., 2011; Nelson et al., 2003) have assessed these intermediary variables hypothesized to link classroom behavior to academic outcomes. Low et al. (2015) did observe a positive effect of Second Step on a "social-emotional skills for learning" variable, which appears to overlap with the skills and attitudes of the academic motivation and engagement variables in the current

study. The observed Second Step effect size (g = .114) was slightly smaller in magnitude than (but still within the 95% CI of) the effect sizes observed in the current study.

The final hypothesis was that students exposed to the SSIS-CIP curriculum would demonstrate improved academic skills relative to their peers in nonimplementing classrooms. Results, however, did not support this hypothesis as all tests were nonsignificant with effect sizes and improvement indexes close to 0. In addition, the 95% CIs for both reading and mathematics effect sizes ranged from positive to negative. These findings were consistent with the academic skills effect sizes reported in the second grade trial (DiPerna et al., 2016); however, they are smaller than the mean effect sizes reported in the Durlak et al. (2011) meta-analysis (g = .27). One possible explanation for these differences is that Durlak et al.'s review included studies that used school grades as well as standardized tests as outcome measures. Additionally, the reviewed studies focused on a wide range of programs, including multicomponent programs that supplemented teacher-facilitated programs with parent or school-wide initiatives. It is unknown if some of these components may have had an academic enrichment focus, as has been the case in some studies of SEL programs (e.g., Bradshaw et al., 2009). Finally, given that SEL and approaches to learning may set the foundation for development of school readiness and academic achievement (Blair & Raver, 2015), academic outcomes may be more apparent in research evaluating the long-term effects of a universal social-emotional program (e.g., Nelson et al., 2003).

In addition to testing for main effects of SSIS-CIP implementation in each of the four outcome domains (social skills, problem behaviors, approaches to learning, and academic skills), we tested interactions between initial skill level at pretest (at both the student and classroom level) and intervention condition within each analysis. In the second Grade SSIS-CIP trial (DiPerna et al., 2015, 2016) as well as studies of IYCD (Webster-Stratton et al., 2008) and Second Step effectiveness (Low et al., 2015), students with lower levels of initial skills demonstrated larger positive effects relative to their peers with higher levels of initial skills. In the current study, however, there were no statistically significant initial skill level by treatment interactions within any of the four skill domains for first grade students. Although this finding must be replicated, it suggests that, though the SSIS-CIP effects may be smaller in magnitude for younger students in the primary grades, they also may be more universally distributed throughout the classroom.

Limitations and Directions for Future Research

There are several limitations to the current study that also provide directions for future research. First, though the study included a sufficient number of classrooms and students to detect small effects, these participants were drawn from a limited number of schools across two school districts (rural and small urban). As such, replication of the current findings with an additional sample of first grade classrooms and schools is necessary. In addition, although the different versions of the SSIS-CIP curriculum for the intermediate (Grades 3–5) and preschool levels are similar to the version tested in this study, efficacy trials focused on those versions are necessary to determine if the impact of the curriculum is similar across developmental levels.

Beyond these design and replication considerations, it also is important to note that the student observation system used in the current study (CLOCK) focused on molar (broader) classes of behavior to make it feasible for staff to observe student behavior across all three outcome domains of interest (social skills, problem behaviors, and academic engagement). Despite using this approach and completing multiple direct observations for each student, the means for these variables were still low. In addition, although the effect size CIs overlapped between CLOCK scores and corresponding scores from the teacher rating scales, the strongest evidence for the social skills and approaches to learning predictions was based on teacher report via measures well-aligned with the SSIS-CIP program. Given teachers are not blind to treatment condition, future studies of the SSIS-CIP may benefit from using an observation protocol that assesses the specific social skills targeted by the SSIS-CIP instructional units (e.g., empathy, assertion, and self-control). Finally, the current study focused on immediate outcomes resulting from the SSIS-CIP. Studies examining follow-up data, as well as resources required for implementation, are necessary to better understand the benefits and costs associated with this universal program.

Conclusion

Results from the current study suggest that exposure to the SSIS-CIP curriculum has small positive effects on first graders' social skills and approaches to learning. Although the observed effect sizes in these domains were consistent with those reported for other classwide programs (e.g., January et al., 2011; Low et al., 2015), they were approximately half the magnitude of those observed when the SSIS-CIP was implemented with second-grade students (DiPerna et al., 2015, 2016). Effects on problem behaviors were negligible and lower than the effects observed both in the second Grade SSIS-CIP trial (DiPerna et al., 2015) and Durlak et al.'s (2011) meta-analyses of other universal SEL interventions. Academic skill effect sizes also were negligible, which was consistent with outcomes of the second grade trial but smaller than other universal SEL interventions (Durlak et al., 2011).

The pattern of observed findings across first and second grade suggests that the SSIS-CIP Early Elementary version yields positive effects in the social skills and approaches to learning domains with the effects being larger when implemented in second grade. It is important to note, though, that the 95% CIs for the effect sizes demonstrate overlap across all domains. Thus, it is possible that the observed differences in magnitude between first and second grade are not true differences but because of random variation across trials. As such, additional trials are necessary to determine if the SSIS-CIP effects replicate with new samples of first and second-grade students (e.g., Makel & Plucker, 2014). SSIS-CIP implementation appears to have negligible immediate effects on first grade students' problem behaviors and academic skills, though again given the 95% CIs around the effect sizes in the current study, replication is necessary to confirm these findings. If the pattern of findings across the first- and second-grade trials are replicated in future studies, educators and administrators contemplating adoption of the SSIS-CIP should consider prioritizing second grade for implementation of the program. In addition, researchers studying other universal SEL programs for young

children should test for potential developmental differences in program effectiveness.

References

Ashdown, D. M., & Bernard, M. E. (2012). Can explicit instruction in social and emotional learning skills benefit the social-emotional development, well-being, and academic achievement of young children? *Early Childhood Education, 39,* 397–405. http://dx.doi.org/10.1007/s10643-011-0481-x

Baker-Henningham, H., Walker, S., Powell, C., & Gardner, J. M. (2009). A pilot study of the Incredible Years Teacher Training programme and a curriculum unit on social and emotional skills in community preschools in Jamaica. *Child: Care, Health and Development, 35,* 624–631. http://dx.doi.org/10.1111/j.1365-2214.2009.00964.x

Barrish, H. H., Saunders, M., & Wolf, M. M. (1969). Good behavior game: Effects of individual contingencies for group consequences on disruptive behavior in a classroom. *Journal of Applied Behavior Analysis, 2,* 119–124. http://dx.doi.org/10.1901/jaba.1969.2-119

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.

Bierman, K. L., Domitrovich, C. E., Nix, R. L., Gest, S. D., Welsh, J. A., Greenberg, M. T., . . . Gill, S. (2008). Promoting academic and social-emotional school readiness: The head start REDI program. *Child Development, 79,* 1802–1817. http://dx.doi.org/10.1111/j.1467-8624.2008.01227.x

Blair, C., & Raver, C. C. (2015). School readiness and self-regulation: A developmental psychobiological approach. *Annual Review of Psychology, 66,* 711–731. http://dx.doi.org/10.1146/annurev-psych-010814-015221

Bradshaw, C. P., Zmuda, J. H., Kellam, S. G., & Ialongo, N. S. (2009). Longitudinal impact of two universal preventive interventions in first grade on educational outcomes in high school. *Journal of Educational Psychology, 101,* 926–937. http://dx.doi.org/10.1037/a0016586

Bub, K. L. (2009). Testing the effects of classroom supports on children's social and behavioral skills at key transition points using latent growth modeling. *Applied Developmental Science, 13,* 130–148. http://dx.doi.org/10.1080/10888690903041527

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

Collaborative for Academic, Social, and Emotional Learning. (2016, October 5). *Approaches to social-emotional learning*. Retrieved from http://www.casel.org/what-is-sel/approaches/

Committee for Children. (1992). *Second Step: A violence prevention curriculum, grades 1–3* (2nd ed.). Seattle, WA: Author.

Cumming, G. (2014). The new statistics: Why and how. *Psychological Science, 25,* 7–29. http://dx.doi.org/10.1177/0956797613504966

Denham, S. A., Blair, K. A., DeMulder, E., Levitas, J., Sawyer, K., Auerbach-Major, S., & Queenan, P. (2003). Preschool emotional competence: Pathway to social competence? *Child Development, 74,* 238–256. http://dx.doi.org/10.1111/1467-8624.00533

Denham, S. A., Way, E., Kalb, S. C., Warren-Khot, H. K., & Bassett, H. H. (2013). Preschoolers' social information processing and early school success: The challenging situations task. *British Journal of Developmental Psychology, 31,* 180–197. http://dx.doi.org/10.1111/j.2044-835X.2012.02085.x

DiPerna, J. C., & Elliott, S. N. (2000). *Academic Competence Evaluation Scales*. San Antonio, TX: The 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,* 123–141. http://dx.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,* 189–203. http://dx.doi.org/10.1002/pits.21891

DiPerna, J. C., Volpe, R. J., & Elliott, S. N. (2005). A model of academic enablers and mathematics achievement in the early grades. *Journal of School Psychology, 43,* 379–392. http://dx.doi.org/10.1016/j.jsp.2005.09.002

Durlak, J. A. (2009). How to select, calculate, and interpret effect sizes. *Journal of Pediatric Psychology, 34,* 917–928. http://dx.doi.org/10.1093/jpepsy/jsp004

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,* 405–432. http://dx.doi.org/10.1111/j.1467-8624.2010.01564.x

Durlak, J. A., Weissberg, R. P., & Pachan, M. (2010). A meta-analysis of after-school programs that seek to promote personal and social skills in children and adolescents. *American Journal of Community Psychology, 45,* 294–309. http://dx.doi.org/10.1007/s10464-010-9300-6

Early, D., Pianta, R., & Cox, M. (1999). Kindergarten teachers and classrooms: A transition context. *Early Education and Development, 10,* 25–46. http://dx.doi.org/10.1207/s15566935eed1001_3

Elliott, S. N., & Gresham, F. M. (2007). *Social Skills Improvement System: Classwide Intervention Program*. Minneapolis, MN: Pearson Assessments.

Enders, C. K., & Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: A new look at an old issue. *Psychological Methods, 12,* 121–138. http://dx.doi.org/10.1037/1082-989X.12.2.121

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

Flay, B. R., & Allred, C. G. (2003). Long-term effects of the Positive Action program. *American Journal of Health Behavior, 27*(Suppl. 1), S6–S21. http://dx.doi.org/10.5993/AJHB.27.1.s1.2

Flay, B. R., Allred, C. G., & Ordway, N. (2001). Effects of the Positive Action program on achievement and discipline: Two matched-control comparisons. *Prevention Science, 2,* 71–89. http://dx.doi.org/10.1023/A:1011591613728

Graham, J. W. (2009). Missing data analysis: Making it work in the real world. *Annual Review of Psychology, 60,* 549–576. http://dx.doi.org/10.1146/annurev.psych.58.110405.085530

Greenberg, M. T., Domitrovich, C., & Bumbarger, B. (2001). The prevention of mental disorders in school-aged children: Current state of the field. *Prevention & Treatment, 4,* 1–62. http://dx.doi.org/10.1037/1522-3736.4.1.41a

Greenwald, A. G., Gonzalez, R., Harris, R. J., & Guthrie, D. (1996). Effect sizes and p values: What should be reported and what should be replicated? *Psychophysiology, 33,* 175–183. http://dx.doi.org/10.1111/j.1469-8986.1996.tb02121.x

Gresham, F., & Elliott, S. N. (2008). *Social Skills Improvement System Rating Scales*. Minneapolis, MN: Pearson Assessments.

Grossman, D. C., Neckerman, H. J., Koepsell, T. D., Liu, P. Y., Asher, K. N., Beland, K., . . . Rivara, F. P. (1997). Effectiveness of a violence prevention curriculum among children in elementary school. A randomized controlled trial. *Journal of the American Medical Association, 277,* 1605–1611. http://dx.doi.org/10.1001/jama.1997.03540440039030

Hamre, K. B., Mashburn, A. J., Pianta, R. C., & LoCasale-Crouch, J. (2008). *Classroom Assessment Scoring System, Pre-K: Technical appendix*. Baltimore, MD: Brookes.

Heckman, J. J. (2006). A broader view of what education policy should be. In N. F. Watt, C. Ayoub, R. H. Bradley, J. E. Puma, & W. A. LeBoeuf (Eds.), *The crisis in youth mental health: Early intervention programs and policies* (Vol. 4, pp. 3–26). Westport, CT: Praeger.

Hofmann, D. A., & Gavin, M. B. (1998). Centering decisions in hierarchical linear models: Implications for research in organizations. *Journal of Management, 24,* 623–641. http://dx.doi.org/10.1177/014920639802400504

Hughes, J. N., & Kwok, O. M. (2006). Classroom engagement mediates the effect of teacher-student support on elementary students' peer acceptance: A prospective analysis. *Journal of School Psychology, 43,* 465–480. http://dx.doi.org/10.1016/j.jsp.2005.10.001

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,* 242–256.

Linares, L. O., Rosbruch, N., Stern, M. B., Edwards, M. E., Walker, G., Abikoff, H. B., & Alvir, J. (2005). Developing cognitive-social-emotional competencies to enchance academic learning. *Psychology in the Schools, 42,* 405–417. http://dx.doi.org/10.1002/pits.20066

Low, S., Cook, C. R., Smolkowski, K., & Buntain-Ricklefs, J. (2015). Promoting social-emotional competence: An evaluation of the elementary version of Second Step®. *Journal of School Psychology, 53,* 463–477. http://dx.doi.org/10.1016/j.jsp.2015.09.002

Makel, M. C., & Plucker, J. A. (2014). Facts are more important than novelty: Replication in the education sciences. *Educational Researcher, 43,* 304–316. http://dx.doi.org/10.3102/0013189X14545513

Makel, M. C., Plucker, J. A., & Hegarty, B. (2012). Replications in psychology research: How often do they really occur? *Perspectives on Psychological Science, 7,* 537–542. http://dx.doi.org/10.1177/1745691612460688

McClelland, M. M., Acock, A. C., & Morrison, F. J. (2006). The impact of kindergarten learning-related skills on academic trajectories at the end of elementary school. *Early Childhood Research Quarterly, 21,* 471–490. http://dx.doi.org/10.1016/j.ecresq.2006.09.003

Miles, S. B., & Stipek, D. (2006). Contemporaneous and longitudinal associations between social behavior and literacy achievement in a sample of low-income elementary school children. *Child Development, 77,* 103–117. http://dx.doi.org/10.1111/j.1467-8624.2006.00859.x

Nelson, G., Westhues, A., & MacLeod, J. (2003). A meta-analysis of longitudinal research on preschool prevention programs for children. *Prevention & Treatment, 6,* 1–35. http://dx.doi.org/10.1037/1522-3736.6.1.631a

Pianta, R. C., La Paro, K. M., & Hamre, B. K. (2008). *The Classroom Assessment Scoring System Manual (K-3)*. Baltimore, MD: Brookes.

Raudenbush, S. W. (1997). Statistical analysis and optimal design for cluster randomized trials. *Psychological Methods, 2,* 173–185. http://dx.doi.org/10.1037/1082-989X.2.2.173

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.

Raudenbush, S. W., Spybrook, J., Liu, X., & Congdon, R. (2005). Optimal design for longitudinal and multilevel research: Documentation for the "Optimal Design" software [Computer program] (Version March 09, 2005).

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

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

Shonkoff, J. P., & Philips, D. A. (Eds.). (2000). *From neurons to neighborhoods: The science of early childhood development*. Washington, DC: National Academy Press.

U.S. Department of Education, Institute of Education Sciences, What Works Clearinghouse. (2016, September). *What works clearinghouse: procedures and standards handbook* (Version 3.0). Retrieved from http://whatworks.ed.gov

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

Wasserstein, R. L., & Lazar, N. A. (2016). The ASA's statement on p-values: Context, process, and purpose. *The American Statistician, 70,* 129–133. http://dx.doi.org/10.1080/00031305.2016.1154108

Webster-Stratton, C., Jamila Reid, M., & Stoolmiller, M. (2008). Preventing conduct problems and improving school readiness: Evaluation of the Incredible Years Teacher and Child Training Programs in high-risk schools. *Journal of Child Psychology and Psychiatry, 49,* 471–488. http://dx.doi.org/10.1111/j.1469-7610.2007.01861.x

Webster-Stratton, C., & Reid, M. J. (2004). Strengthening social and emotional competence in young children-The foundation for early school readiness and success: Incredible Years Classroom Social Skills and Problem-Solving Curriculum. *Infants and Young Children, 17,* 96–113. http://dx.doi.org/10.1097/00001163-200404000-00002

Wentzel, K. R., & Asher, S. R. (1995). The academic lives of neglected, rejected, popular, and controversial children. *Child Development, 66,* 754–763. http://dx.doi.org/10.2307/1131948

Received October 19, 2015
Revision received January 5, 2017
Accepted January 9, 2017 ■