Note. This article has been accepted for publication in the *Journal of Educational Effectiveness,* published by Taylor & Francis.

Does Universal SEL Work Under Typical Implementation Practices?

Outcomes of a First Grade Effectiveness Trial

James C. DiPerna, Susan Crandall Hart, Pui-Wa Lei, Tianying Sun, Hui Zhao,

Kyle Husmann, Xinyue Li

The Pennsylvania State University

Accepted Date: July 19, 2023

The research reported here was supported, in whole or in part, by the Institute of Education Sciences, U.S. Department of Education, through Grant R305A170047 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.

The research reported here was pre-registered (#1833.1v1) in the *Registry of Efficacy and Effectiveness Studies*.

Correspondence regarding concerning this article should be addressed to James C. DiPerna, Department of Educational Psychology, Counseling, and Special Education, The Pennsylvania State University, 125 CEDAR Building, University Park, PA 16802. Email:jdiperna@psu.edu. Phone

Abstract

The purpose of this study was to explore the effectiveness of a universal classroom-based social-emotional learning (SEL) program for first grade students. Forty classrooms from 13 elementary schools across 3 states participated in the trial. Teachers in classrooms randomly assigned to the implementation condition were provided with free access to all curricular materials from the SSIS SEL Classwide Intervention Program (Elliott & Gresham, 2017). Consistent with the goals of an effectiveness trial, teachers made local decisions regarding how they implemented the SSIS SEL CIP within their own classroom. Measures of participating students' (*N*= 365) social behavior (positive and negative) were collected before and after program implementation in all participating classrooms (treatment and business-as-usual control). Overall, results were mixed relative to hypotheses. Main effects tended to be non-significant and negligible in size; however, several interactions indicated positive outcomes resulting from program exposure (increases in prosocial behavior, decreases in negative social behavior) for students with lower skills at baseline. In addition to replication, future research directions include identifying critical program components and approaches to implementation that optimize program effectiveness under authentic conditions.

Does Universal SEL Work Under Typical Implementation Practices?

Outcomes of a First Grade Effectiveness Trial

Implementation of social-emotional learning (SEL) programs in U.S. elementary schools is rapidly increasing. In a 2021 national survey of SEL practices, 93% of districts reported implementing some form of SEL, with many (68%) implementing formal SEL programs. Not surprisingly, spending on SEL curricula also has increased 80% since 2019 (Bryant et al., 2021). Responding to the dual pandemics of COVID-19 and systemic racial injustice (J. M. Jones, 2021), teachers and administrators have reported prioritizing efforts to promote students' social, emotional, and behavioral health in school (Bryant et al., 2021; McGraw Hill, 2021; Zieher et al., 2021). Concern also has been heightened during the pandemic for children in the early elementary grades, particularly those from marginalized communities, as their formal schooling has been significantly disrupted during a critical developmental period (D'Souza, 2021; Greenberg, 2021; Kamenetz et al., 2021). As such, many school leaders are urgently seeking data to inform efforts to select, adopt, and implement evidence-based programs intended to improve the social-emotional outcomes of their primary grade students.

Unfortunately, such data are not always readily available or easily interpretable. Results of large meta-analytic studies suggest that SEL interventions improve student social-behavioral competencies in both the short- and long-term (Durlak et al., 2011; Taylor et. al., 2017). Randomized controlled efficacy trials of individual programs, however, have yielded smaller and more varied causal impacts (Jones & Doolittle, 2017), and only about one-third of SEL publishers have conducted independent evaluations of their programs with some form of a comparison group (Bryant et al., 2021). In contrast to highly controlled efficacy studies that often involve researcher support for training and/or implementation, effectiveness trials evaluate

programs *under routine conditions* - the everyday practices and conditions occurring in heterogenous classroom and school settings under typical implementation support (Al-Ubaydli et al., 2019; Dettmer et al., 2017). Without information about how SEL programs translate and operate at scale across diverse populations and at specific grade levels of need, educators are missing a critical piece of the evidence base of SEL. To address this need, the purpose of this study was to investigate the effectiveness of a classwide SEL program on first-grade students' social and behavioral outcomes across a diverse school sample when implemented under routine conditions.

Social-Behavior Efficacy of Primary Grade (K-1) SEL

Synthesizing evidence regarding the effects of SEL is complicated by differences in how SEL has been defined, conceptualized, and implemented in schools (Humphrey et al., 2011). A widely adopted definition of SEL from the Collaborative for Academic, Social, and Emotional Learning (CASEL) describes a process of promoting positive learning environments and fostering the development of five competencies: self-awareness, self-management, social awareness, relationship skills, and responsible decision-making (Osher et al., 2016). Other frameworks emphasize improving social/relational/interpersonal and intrapersonal (emotional and cognitive regulation) skills (S. M. Jones & Bouffard, 2012). Promoting SEL in schools can range from small practices embedded into existing instruction, to whole-school approaches infused into all aspects of the school day, to curriculum-based SEL programs (Elias, 2019). With respect to the latter approach, stand-alone, manualized, direct instruction curricula designed to explicitly teach SEL skills can be integrated into a universal Tier 1 level within a school-based multi-tiered system of supports (U. S. Department of Education, 2021). Often taught by the classroom teacher in elementary schools, these primary prevention programs are intended to

promote students' positive social-emotional and learning-related behavioral skills and reduce problematic behaviors.

The primary grades represent a promising developmental period for maximizing the benefits of universal prevention programs aimed at promoting positive social behavior. Potentially malleable social-behavioral factors in the early grades have been linked to concurrent and subsequent student outcomes related to prosocial behavior (e.g., Hamre & Pianta, 2005; Morgan et al., 2008; Rabiner et al., 2011; Spivak & Farran, 2012). Young students' prosocial skills predict a variety of young adult outcomes including educational attainment, public assistance, crime, and substance abuse, as well as societal benefits to the economy and workforce (D. E. Jones et al., 2015; 2017). Findings from the field of neuroscience suggest that early prevention is more effective than remediation later in life; the brain is more plastic and can rewire much easier in childhood compared to adulthood (National Scientific Council on the Developing Child, 2007). The importance of early intervention has been further underscored by the COVID-19 crisis: while young students, especially those from communities that are underserved by mental health and behavioral health resources, have been disproportionately impacted during the pandemic, school-based supports can serve in a protective function to support such students (D'Souza, 2021; Herbers et al., 2021).

Meta-analyses of studies evaluating school-based interventions intended to improve prosocial behavior have demonstrated a positive impact in aggregate but suggest some moderation by age or grade. According to these reviews, average effect sizes for prosocial behavior are typically in the small-to-medium range for immediate outcomes (ES =.15 - .39; Durlak et al., 2011; January et al., 2007; Sklad et al., 2012) and smaller in magnitude at follow-up (ES = .12 - .13; Sklad et al., 2012; Taylor et al., 2017). In a seminal meta-analysis of 213

school-based SEL programs, Durlak et al. (2011) reported small effects on improving prosocial behavior and reducing conduct problems (ES = .24 and .22 respectively); results indicated that there was a small inverse relationship between skill outcomes and student age (r = -.27). Similarly, a meta-analysis of 28 studies of classroom-based social skills programs found larger positive effects when interventions were provided in preschool and Kindergarten (d = .55) compared to later grades (d = .12 - .23; January et al., 2011). With respect to follow-up outcomes, across 82 longitudinal studies of school-based SEL, interventions focused on children ages 5-10 yielded larger effect sizes (ES = .27) as compared to those for early adolescent (ES = .12) and adolescent (ES = .18) age groups (Taylor et al., 2017). While these reviews capture a wide range of program types, study designs, and outcome measures, accumulated evidence suggests a positive impact that appears to be more pronounced in the early grades.

Rigorous randomized-controlled efficacy studies of universal programs aimed at improving SEL competencies of primary grade students have yielded mixed results, even when programs are implemented under ideal conditions. For example, a study of Promoting Alternative Thinking Strategies (PATHS) found small-to-medium positive main effects for teacher-rated outcomes including authority acceptance (d=.24), concentration (d=.12), and social competence (d=.34) after 3 years of sustained exposure starting in first grade (Conduct Problems Prevention Research Group, 2010). In that study, teachers were provided extensive support from researchers including a 2-day training as well as weekly consultation, team-teaching, and lesson demonstration from research staff (an average of 1-1.5 hours per week in each classroom). A later independent evaluation of the impact of PATHS on first-grade students - during which training, ongoing professional development, and coaching were provided - found a statistically significant decline in teacher-rated aggressive behavior 2 years post-implementation (d=.42-.46)

but not immediately after the intervention. In addition, no positive effects were found on immediate or follow-up ratings of prosocial behavior (Malti et al., 2013). An evaluation of the INSIGHTS program, delivered by trained and supervised facilitators (rather than classroom teachers) to groups of K-1 parents, teachers, and students over 2 years, found an increase in a direct measure of sustained attention (g = .39) and a decrease in teacher-rated problem behaviors (ES = -.54; O'Connor et al., 2014). In a study of Second Step where teacher training and implementation monitoring were conducted by the research team, but their day-to-day involvement was limited to maximize generalizability of results, there were negligible-to-small main effects on teacher-rated behavioral outcomes of K-2 students (g = .02 to .13). Positive effects were larger for students with lower initial skill levels (Low et al., 2015).

The Social Skills Improvement System Classwide Intervention Program (SSIS-CIP; Elliott & Gresham, 2007), a classroom-based curriculum designed to improve student social skills across the five social-emotional competency areas identified by CASEL, has also been evaluated when implemented under ideal conditions with early grade students. Findings from an efficacy trial indicated that the SSIS-CIP had small positive effects on students' social and learning-related behavior in a sample of 59 first grade classrooms (696 students; 33% racial-ethnic minority students) within two districts (one small urban, one small rural) in the Northeast (DiPerna et al., 2018). Researcher-provided implementation support included a daylong training, an implementation schedule, and implementation monitoring. Small-to-medium sized positive main effects for all teacher-rated social skills and approaches to learning outcomes were found, with effect sizes ranging from .13 to .31. Six behavioral outcomes assessed in the study (overall social skills, communication, empathy, social engagement, academic motivation, academic engagement) were found to have main effect size confidence intervals that did not include zero,

though only two outcomes remained statistically significant after applying a correction to control for false discovery rate. Moderation tests revealed no skill level by treatment interactions. Teachers generally found the curriculum relevant, feasible, and effective (Wollersheim Shervey et al., 2017), and the average cost per student to implement the program was about $19 (Hunter et al., 2018). It is unknown, however, if findings would be similar when the program is implemented under routine conditions in the absence of researcher support.

Effectiveness Research to Inform SEL at Scale

To maximize the benefit of school-based prevention programs, evidence-based practices must be implemented and evaluated without extensive researcher support. When programs are translated into school settings for typical use and adopted at scale, their positive benefits reach more students, including those in different school communities and contexts (Aarons et al., 2017; Horner et al., 2017). A current challenge for administrators and educators is selecting an SEL program that will be effective in meeting their local needs. The 2015 Every Student Succeeds Act provided funding streams to support implementation of evidence-based SEL in schools; however, educators were provided limited guidance on how to select programming that not only had evidence to justify its use but also the greatest probability of positively impacting students (Wrabel et al., 2019). Increased focus on issues of racial injustice and challenges caused by COVID-19 have intensified the demand for social-emotional programming and led to a proliferation of programs focused on SEL. Unfortunately, though, schools may have difficulty discerning between the many available options given the limitations of the research base currently available (Bryant et al., 2021).

The research-to-practice translation continuum includes efficacy studies, followed by replication, effectiveness, and/or implementation research (Curran et al., 2012; Flay et al., 2005).

While efficacy and replication trials provide an opportunity to isolate and confirm intervention effects when SEL programs are implemented in a well-controlled manner, effectiveness trials examine whether an intervention will demonstrate the same level of efficacy when it is implemented in routine conditions (i.e., typical implementation support, everyday practice occurring in schools, and heterogeneity of the target populations; Al-Ubaydli et al., 2019; Dettmer et al., 2017). As evidence builds indicating implementation impacts SEL outcomes (e.g., Domitrovich et al., 2010; Durlak, 2015; Durlak & DuPre, 2011; Evans et al., 2015; Greenberg, 2010), evaluating programming under routine rather than optimal conditions is critical. As of 2022, the Institute of Education Sciences, the U.S. Department of Education's Research arm, had funded 267 research grants within their programs focused on improving the social-emotional, behavioral, and/or character outcomes of students. Of those, less than 2% were effectiveness or scale-up studies of school-based interventions, and less than 1% (n = 2) evaluated universal programs (U.S. Department of Education, 2022). Increasing the number of rigorous effectiveness trials could improve the speed with which knowledge garnered from education science can transfer into and benefit schools (Sanetti et al., 2019).

Purpose, Rationale, & Hypotheses

Given the lack of studies evaluating formal SEL skills programs when implemented in a real-world context, the purpose of this study was to examine effectiveness of a classroom-based program intended to improve positive prosocial learning-related skills. Specifically, we examined social-behavioral outcomes associated with the Social Skills Improvement System Social Emotional Learning Edition Classwide Intervention Program (SSIS SEL CIP; Elliott & Gresham, 2017) when implemented under routine conditions (i.e., implementation decisions made and executed locally by school personnel rather than by the research team). Using data

from our pre-registered multi-site cluster randomized effectiveness trial, this paper examines three hypotheses. First, we hypothesized that students in first-grade classrooms implementing the SSIS SEL CIP would demonstrate higher rates of prosocial behavior at posttest, as measured by both teacher ratings and independent observers, compared to children in non-implementing (waitlist control) classrooms. Second, first-grade students in the SSIS SEL CIP condition were expected to demonstrate lower rates of negative social behaviors at posttest compared to their peers in control classrooms. Third, we hypothesized that social behavior effects from exposure to the SSIS SEL CIP under routine conditions would be stronger for students with lower initial levels of skills.

Method

Participants

We collected data for the present study with a sample of schools, teachers, and students who participated during the 2018-19 school year[1]. A total of 365 first-grade students and 40 teachers in 13 elementary schools ultimately participated in the study. Schools were in seven districts across three states/sites. Ten schools were split between two states in the Midwest: five schools (10 classrooms) were within five different remote rural districts, and five schools (13 classrooms) were in one large suburban district. The three remaining schools (17 classrooms) were in one very large district in a Southeastern state: one school each in a large city, midsize city, and large suburb according to Census locale classification. All participating schools were public schools. One was a charter school, and a second school offered both a Spanish Immersion Magnet program. All but one of the schools (92%) were eligible for Title I funds. In over half of

[1] We originally planned to work with new cohorts of schools, teachers, and students annually through the 2021-22 school year. Due to the onset of the pandemic midway through the second year of our project and its continuation through the initial project end date, the current sample represents approximately 33% of the original target sample.

the schools, more than 75% of the student population qualified for free- or reduced-price lunch. Six of the schools had a majority White student population, four had a majority Black student population, and the remaining three had a majority Hispanic/Latina population. The schools' total student enrollment ranged from 51 to 756, with a median of 355 students per school[2]. At the district level ($N = 7$), the percentage of students receiving special education services ranged from 13- 24%, and the percentage of students identified as English Language Learners ranged from 1-20%.

All participating teachers were female; 80% were White, 5% were Black, and 15% reported having Hispanic, Latine, or Spanish origin. While most teachers (88%) identified English as their primary language, 13% reported Spanish as their primary language. In terms of highest level of attained education, 63% had bachelor's degrees while the remaining teachers had master's degrees. Most of the teachers (73%) had certification in general education, 28% had special education certification, and 15% had dual language certification. On average, teachers reported 16 years of experience in teaching and 4.5 years in their current school ($SD = 10$, range 1 - 35 years). At the classroom level, total student enrollment in the participating first-grade classrooms ranged from 11 to 26.

Students were 52% male, 48% female, 44% White, 25% Black, 23% Hispanic/Latine, 4% multiracial, and 3% Asian. With respect to primary language, most students (94%) spoke English, 4% spoke Spanish, and 2% spoke another language. About 18% of students who spoke English as a primary language also spoke another secondary language. At the beginning of the year, 7% of students received special education services, and 23% received supplemental

[2] School-level demographic data collected from the U.S. Department of Education, National Center for Education Statistics, Common Core of Data (2018-2019)

services. All participants were treated in accord with the American Psychological Association's ethical principles.

Measures

The measures used to assess the social-behavioral outcomes associated with the SSIS SEL CIP paralleled those used in the previous efficacy trial. Student data were collected via teacher report and observations by research staff at two time points: prior to implementation in the fall of 2018 (pretest) and in the spring of 2019 (posttest). Teachers in the treatment condition also reported implementation information weekly and at the end of the school year.

Social Skills and Problem Behavior

Behavior rating scales. Teacher ratings were used to measure pretest and posttest student social skills and problem behaviors. Specifically, teachers completed the Social Skills Improvement System Rating Scales – Teacher Form (SSIS-RST; Gresham & Elliott, 2008) for all students with parental consent to participate in data collection. An overall composite of social skills and seven subscales (Communication, Cooperation, Assertion, Responsibility, Empathy, Engagement, and Self-Control) were assessed with 46 items. Five problem behaviors subscales (Externalizing, Bullying, Hyperactivity/Inattention, Internalizing, and Autism Spectrum) were assessed with 24 items. Ratings used a 4-point scale of *Never* to *Almost Always;* subscale and composites were calculated as mean scores. Scores from these scales have been widely used in research and practice for both normative- and criterion-referenced decision making (Gresham et al., 2010). Further, research has demonstrated that many of the SSIS items align with CASEL's social emotional learning framework (Gresham & Elliott, 2017; Gresham et al., 2018).

Direct observation. Systematic observations were also used to assess a subset of participating students' social behavior within the classroom. Observers used an updated version

of the Cooperative Learning Observation Code for Kids (CLOCK-2; Volpe & DiPerna, 2018) to conduct multiple observations per student at both pretest and posttest. Using partial interval recording, research staff observed each student individually and coded the occurrence of two types of social behavior during 15-second time intervals across 12-minute observation periods during academic instruction. Observers coded Positive Social when the student engaged in any appropriate social behavior that is permitted at that time. Examples of positive social behaviors include talking with a peer about an assignment, sharing work materials, helping a peer with their work, asking a teacher for help with an assignment, raising their hand to participate, responding appropriately to an adult request, and saying "please" or "thank you." Observers coded Instructional Interference when a student's behaviors were potentially distracting to others or interfering with academic tasks. Examples of instructional interference behavior included tantrums, interrupting group activities, calling out during instruction, asking an irrelevant question, or making a loud noise.

Proportions were calculated for each outcome (number of behaviors observed across intervals/total number of observation intervals) and averaged across the observations conducted during each data collection period. Approximately 14% of the observations were conducted with two observers present to monitor interrater reliability. At pretest and posttest, interrater reliability was .80 and .78 for Positive Social and .93 and .97 for Instructional Interference (prevalence and bias adjusted Kappa; Byrt et al., 1993; Sim & Wright, 2005).

Classroom Environment

The Classroom Assessment Scoring System (CLASS K-3; Pianta et al., 2008) is a structured observation protocol for assessing the classroom instructional environment. Using a 7-point scale ranging from *Low* (1-2)*, Middle* (3-5), to *High* (6-7), classrooms are rated on 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). Coded dimensions are aggregated to produce three domain scores: Emotional Support (i.e., teachers' warmth and sensitivity toward students), Classroom Organization (i.e., teachers' use of effective behavior management and varied learning modalities), and Instructional Support (i.e., teachers' use of strategies that develop higher-order thinking and language skills). The psychometric properties of scores from the CLASS have been examined extensively (Briesch et al., 2018). There are strong theoretical and conceptual underpinnings to support the three-factor structure of the CLASS K-3 (Hamre et al., 2007), and, in early elementary grades, research has generally supported this factor structure (Pianta et al., 2008; Sandilos et al, 2016). In the present study, internal consistency was high for the three domains ranging from .87 - .92. Inter-rater agreement (within-1-point) ranged from 85-92% across CLASS domains. Two observation cycles (20 min observation + 10 min coding per cycle) were completed in each classroom at pretest.

Weekly Implementation Surveys

Teachers in the treatment condition reported their SSIS SEL CIP implementation practices via brief online surveys distributed weekly via email on Fridays from January to June 2019. Specifically, teachers reported the number of lessons that they taught during the week as well as provided information about their experience teaching the lessons. To assess overall program completion, we summed the number of lessons taught per teacher across the implementation period. Teachers responded to the question "How closely did you adhere to the lesson plan?" for each lesson they taught each week on a scale of 1 (*Not Implemented*) to 5

(*Completely Implemented*). Overall adherence scores were calculated by averaging across all lessons and all surveys completed.

Study Procedures

Recruitment

The study was approved by the university's Institutional Review Board. Consistent with the goal of testing the effectiveness of the SSIS SEL CIP, we recruited districts that were considering adopting a universal SEL program as part of their routine/typical practice. We distributed information about the trial online and through professional networks. If schools indicated they wanted to participate after learning the details of the project, we subsequently discussed further details, addressed additional questions, and established plans for participation via individual conversations over email and phone. Prior to enrolling a school site into the study, we sought and received permission to conduct the research according to school district protocols as applicable.

Figure 1 displays the flow of participants through study phases. At the beginning of the 2018-2019 school year, all first-grade teachers within participating schools were invited to join the project. Teachers provided informed consent through paper or digital forms, and 40 (97.5% of all first-grade classroom teachers) agreed to participate in the project. In these classrooms, we asked the teachers to send home a paper invitation letter and consent form to the families of their students. Digital consent forms were also available upon request. Across all schools, teachers sent home forms to families of 789 students; they also sent home a follow-up reminder letter approximately one week later. Approximately 66% (n = 518) of the forms were signed and returned. At the end of the recruitment period, a total of 415 forms were returned that indicated

active parental agreement to participate in data collection associated with the project. This represented 53% of all forms distributed, and 80% of all forms returned.

Given the cost of data collection and available resources, our target number of students per classroom for participation in data collection was 10. When the number of students with parental consent to participate in data collection exceeded this number, we randomly selected 10 students (with gender stratification when possible) to participate in the data collection. Prior to data collection, our field-based research staff asked students for their verbal assent to participate in the data collection activities associated with the project; 99% provided affirmative consent. In total, 365 students participated in data collection at pretest (M = 9 students per classroom); because 12 students moved from their participating school during the year (3% of sample), 353 students remained at posttest.

After the consent process was completed, schools were randomly assigned either to treatment (SSIS SEL CIP implementation) or business-as-usual waitlist control condition in first grade. Random assignment at the school level (rather than within school at the classroom level) allowed teachers to plan in grade-level teams, which teachers and administrators indicated was their preference. Six schools (19 classrooms) were assigned to the treatment condition and seven schools (21 classrooms) were in the control.

Data Collection

Both treatment and control teachers completed the SSIS rating scales for each participating student in their classroom prior to program implementation (November-December) as well as at the end of the academic year (May-June). We distributed questionnaires online, and compensated teachers for their time spent completing them. During the implementation period, treatment teachers also completed brief weekly online surveys documenting the number of

lessons they taught during the previous week; control teachers documented any SEL approaches that they implemented as part of their business-as-usual practice. All teachers also provided information about themselves, their classroom composition, and their implementation of the program (if applicable) on questionnaires at the beginning and end of the school year. Finally, trained field staff completed periodic observations of program implementation. All observations were scheduled based on teachers' preferences to minimize disruption to the classroom environment.

CLOCK-2 observations were conducted for a subset of students participating in data collection. We randomly selected six participating students per classroom; a gender balance of three girls and three boys was selected when there were at least three students of each gender with permission to participate in data collection within a classroom. A total of 225 students were selected for CLOCK observations, and 216 remained at posttest. An average of 5 observations were conducted per student during both the pretest and posttest periods. To standardize the observation context, observations were conducted on different days during literacy (reading/writing) and math instruction. Prior to conducting live CLOCK-2 observations, field staff members reached a mastery criterion of 80% agreement when observing two videotaped lessons. During data collection, 431 observations (approximately 14% of all observations) were conducted with two coders to monitor interrater reliability.

A CLASS observation (two cycles) was conducted at pretest. Observers were certified CLASS observers who completed an additional mastery activity with a certified CLASS trainer and achieved 80% accuracy before completing live classroom observations. In addition, a subsample of classrooms (21%) was rated by two observers to ensure interrater reliability.

A total of 20 field staff members facilitated student assent, conducted observations, and interfaced with school staff for the duration of the project. We hired local data collectors from the communities surrounding participating schools. They were all female, 70% were White, and 30% were Black. About 15% reported Hispanic/Latine ethnicity. While all data collectors spoke English as their primary language, 30% were bilingual. With respect to the highest level of attained education, 5% had some college, 45% had bachelor's degrees, and 50% had at least a master's degree. They had previous experience in the fields of education, research, and human services and reported an average of 16 years (range 0 – 40) of professional experience.

Field staff were trained during a 2-day in-person local training, which focused on ensuring human subject research protections, collaborating successfully and responsively with school partners, conducting quality CLOCK-2 observations, and collecting reliable and valid data. During the observation portion of the training, we taught data collectors about interval recording methods and the CLOCK-2 protocol. We also presented operational definitions, examples, and non-examples of all behavior categories. We demonstrated a CLOCK-2 observation using a video of classroom instruction and then had data collectors practice in small groups and independently using additional video examples. Observers were required to meet an 80% mastery criterion prior to conducting live observations for the study. In addition, live synchronous follow-up training was provided online, and observers were required to again meet an 80% agreement criterion after watching a videotaped lesson. If observers failed to gain mastery, individual feedback was provided along with one additional opportunity to achieve the master criterion. In addition, university research staff supported field staff throughout data collection with ongoing consultation via phone and email. For example, staff were required to conduct a portion of their observations with a secondary observer present. When ratings differed

between the two observers, they discussed their codes and consulted with us to resolve discrepancies.

SSIS SEL CIP Implementation

The SSIS SEL CIP (Gresham & Elliott, 2017) is a universal classroom-based program designed to promote students' positive classroom behavior. The manualized program includes units focused on discrete social behaviors that a nationally representative sample of teachers identified as being important to success in the classroom. The SSIS SEL CIP includes 10 core units focused on foundational skills as well as additional supplemental units focused on more developmentally advanced skills. The 10 foundational skills of the core units include: listen to others, say please and thank you, follow the rules, pay attention to your work, ask for help, take turns when you talk, get along with others, stay calm with others, do the right thing, and do nice things for others. Teachers in the study were also provided access to 13 additional advanced skill units that covered more complex skills such as own your actions, stand up for others, and listen to different ideas. The SSIS SEL CIP core and advanced units also have been aligned with the five social-emotional competency areas identified by CASEL (Elliott & Gresham, 2017).

Each unit consists of three scripted lessons taught using six complementary instructional approaches (tell, show, do, practice, monitor progress, generalize). Teachers introduce each skill by defining it, describing the steps needed to demonstrate the skill, and modeling the skill. Students then practice identifying the skill by watching video clips and using the skill through role plays. Teachers also guide students in thinking about their progress toward mastering the skill and ways they can demonstrate the skill outside of school. Program materials include a teacher manual with lesson scripts, accompanying slides, brief video clips, skill step posters, role play cards, and emotion cue cards.

In the fall, teachers in the treatment condition were provided with a printed manual and given access to the other program materials for the core and advanced units via a secure project website. All materials were provided free of charge for participating schools; teachers in the waitlist control condition received their materials after posttest (June) for use in future school years. While active parental/guardian consent was needed for students to participate in data collection, all students took part in the lessons as part of the classroom instruction given the universal delivery of the program. Consistent with the goal of evaluating the effectiveness of the SSIS SEL CIP when implemented under the typical practices of teachers and routine conditions in schools, no formal professional development, consultation, implementation support, or training of any kind was provided by research staff to participating teachers. As such, we asked teachers to share how they approached preparing to teach the SSIS SEL CIP units. The five most common approaches were: planned individually (73%), planned with colleagues (39%), watched SSIS SEL CIP training video provided by publisher (39%), attended a formal training provided by school/district (17%).

In total, teachers had access to 69 social skill lessons: 30 lessons from 10 core units and 39 lessons from 13 advanced units). On average, implementing teachers reported teaching 25 total lessons; however, there was significant variability (*SD* = 8). The average number of core lessons completed was 22 (*SD* = 7); teachers taught 6 completed core units (3 lessons per unit) on average (*SD* = 3). For advanced skills, teachers taught an average of 2 lessons (*SD* = 4) and 0.5 units (*SD* = 1). Over 80% of teachers *(n* = 16) reported starting implementation in January, though two teachers started in February and one teacher did not begin until March. Almost all teachers ended implementation in April (42%) or May (47%), except for two who stopped in March and June, respectively.

During the implementation period, field staff conducted periodic observations of treatment teachers' implementation of program lessons to monitor fidelity of implementation (*M* = 5 observations per teacher). For each observation, staff recorded which lesson steps were implemented via a structured checklist; on average, teachers completed 75% of the lesson steps across observations conducted (*SD* = 14%). Staff also provided summative ratings of teachers' level of implementation for each lesson component on a scale from *Not Implemented* (1) to *Completely Implemented* (5). Across all observations and lesson components, observers rated teachers' implementation to be in the *Mostly Implemented* (4) range (*M* = 3.83, *SD* = .59). Teachers' mean rating of their own lesson adherence as reported on weekly surveys was nearly identical to observer rating (*M* = 3.86, *SD* = .46). Consistent with the primary goal of testing the SSIS SEL CIP under routine conditions, staff did not share information back with teachers or attempt to change their practice in any way[3].

Design and Analysis Plan

We used a *Multi-Site Cluster Randomized Trial (CRT)* to evaluate the effectiveness of the SSIS SEL CIP in Grade 1 classrooms under routine conditions. In this design, schools were randomly assigned to treatment conditions within sites (matched schools by demographics and region). Hierarchical Linear Modeling (HLM; Raudenbush & Bryk, 2002) was used to account for the clustering of students within classrooms within schools. Missing data analysis was first conducted to determine missing data treatment. Because missing was less than 5%, listwise deletion was adopted. Variables statistically related to missing were controlled in the analysis models. In addition, baseline equivalence between treatment conditions was assessed for each of

[3] For more details about the implementation choices and approaches of teachers in this trial, including program completion, lesson activity adherence, verbal adherence to lesson plans, lesson delivery quality, and factors influencing implementation, see Hunter et al., 2022.

the outcome measures and demographic variables. Statistically significant nonequivalent baseline measures were included as covariates in the HLM models to mitigate potential bias for all outcome analyses.

We analyzed a 3-level random-intercepts HLM model in which students were nested in classrooms and classrooms were nested within schools for each of the outcome variables. Due to the small sample size at the site/region level, regions were analyzed as fixed effects. We included gender (1=male, 0=female), race/ethnicity (1=white, 0=racial-ethnic minority), receipt of supplementary services (1=yes, 0=no) or special education (1=yes, 0=no), problem behavior composite (due to baseline nonequivalence), and group-mean centered pretest scores at student level. At the classroom level, group-mean centered class average pretest scores and CLASS instructional support (due to baseline imbalance) were included. Treatment effect was tested using dummy codes (1 = SSIS SEL CIP, 0 = control) with grand-mean centered school average pretest scores, school size (1 = large or 401-756 total student enrollment, 0=small or 129-400), and total enrollment of racial-ethnic minority students (1=large or 60.01%-88.02%, 0=small or 5.43% - 60%) controlled for at the school level.

Potential interaction effects between treatment and demographic variables and different levels of the pretest measures were also explored by adding appropriate product terms to the model. Statistically significant interactions were plotted to examine the patterns of interaction and nonsignificant ones were dropped from the final interaction model. The Benjamini-Hochberg (1995) correction was applied by outcome domain to control for false discovery rate due to multiple comparisons. In addition, we estimated effect sizes of SSIS SEL CIP as a standardized adjusted mean difference between treatment groups (adjusting for pretest scores and other covariates). Specifically, we divided the coefficient for treatment condition from the main-effects

HLM model by the unadjusted student-level *pooled* within-group standard deviation of the pretest outcome measure. We also computed an improvement index that indicated the percentile rank change expected for an average student in the control group if they had received the intervention (What Works Clearinghouse, 2022).

We used the Mixed procedure of SAS to estimate all HLM models for Social Skills outcomes. Due to low frequency of problem behaviors observed in this sample, we used the GLIMMIX procedure of SAS with Poisson distribution and log link (i.e., Poisson regression) to analyze the problem behavior outcomes. Effect sizes for problem behavior measures were computed for the original scale at the mean of all covariates (by converting the least square treatment means from the Poisson regression model from the log scale back to the original scale before standardizing).

Results

There was no attrition at the school and class (teacher) levels. Total student-level attrition (i.e., no outcome data from a student who participated in pretest data collection) was 5.1%, and differential attrition across conditions was negligible (.01%). Of the enrolled sample (N = 365 students), 3.3% missed posttest due to moving. Missing completely at random was not rejected by Little's MCAR test (chi-square =18.35, df=16, p=.30). There were no other missing data aside from those intentionally not collected due to the random selection of participant subsamples for direct observation. As such, listwise deletion was used to handle missing data, and the final analysis sample size was 353.

Tables 1 and 2 present the demographic characteristics and descriptive statistics of measures for the final analysis sample by treatment condition. Student-level baseline nonequivalence between conditions was present for percentage of White students (55.4% in

control compared to 30.5% in treatment), percentage of boys (57.5% vs. 46.7% in control and treatment, respectively), percentage of English Language Learners (9.1% in control vs. 2.99% in treatment), percentage of schools with predominantly racial-ethnic minority students (36.0% in control, 83.8% in treatment), problem behavior composite (higher in treatment condition), hyperactive or inattentive behaviors (higher in treatment), and internalizing behaviors (higher in treatment). Mean level of instructional support was also lower in the treatment group than in the control at baseline. As such, the demographic variables, baseline problem behavior composite and level of instructional support were included as covariates in the HLM models.

Class- and school-level Intra-class correlations (ICC) calculated for both pretest and posttest outcome measures are shown in Table 3. Class-level ICCs for posttest scores and school-level ICCs for direct observation measures were generally large. Therefore, three-level random-intercepts models were analyzed to account for the dependency of observations due to clustering. However, school-level variance for some outcomes (assertion, self-control, problem behavior composite, externalizing, hyperactivity-inattentive, and internalizing) became zero after entering predictors and was removed from the final model. Main effect model estimates for social skill outcomes appear in Table 4. As expected, pretest scores of the corresponding social skill domain (student- and class-level) and the problem behavior composite were significant predictors of most teacher-rated social skill outcomes. There was no statistically significant evidence of treatment main effect after controlling for pretest measures and other covariates. However, some statistically significant interactions of participation in the SSIS SEL CIP were observed (Table 5), and those remaining significant after Benjamini-Hochberg correction were plotted.

Figures 2 and 3 show the interaction between treatment and student-level baseline score for teacher ratings of responsibility and direct observation of positive social behavior,

respectively. The SSIS SEL CIP tended to benefit students with lower initial skills in these domains but not for students with higher initial skills (> .5 SD above the mean for responsibility and above the mean for positive social). However, adjusted treatment group differences in these domains within the observed baseline score range (plotted in the figures) were not statistically significant at the .05 level. The observed interaction between treatment condition and race was similar for communication (Figure 4) and social engagement (Figure 5) in that participation in the SSIS SEL CIP appeared to improve communication and social engagement scores slightly for White students but not for racial-ethnic minority students.

For problem behavior outcomes, student- and class-level pretest scores were also largely significant predictors of the corresponding teacher-rated posttest scores. There were no statistically significant main effects of treatment on problem behavior outcomes (Table 6). However, the interaction between student-level pretest and treatment was statistically significant for all teacher ratings of problem behavior outcomes but internalizing after the Benjamini-Hochberg adjustment, and the pattern indicated some benefit of SSIS SEL CIP participation in reducing problem behaviors for students with high baseline scores. Specifically, adjusted treatment group differences were statistically significant for students with baseline scores greater than 3.64 standard deviations above the mean for overall problem behavior (Figure 6), 4.36 standard deviations above the mean for externalizing behavior (Figure 7), and 5.33 standard deviations above the mean for bullying (Figure 8). (For bullying, significant differences also were observed for baseline scores less than 1.76 standard deviations below mean.) The regions of significant treatment group differences in adjusted hyperactivity-inattention, however, were outside the observed baseline score range (Figure 9). The interaction between student-level pretest and treatment was also statistically significant on direct observations of instructional interference

(Figure 10). In contrast to teacher ratings of problem behaviors, participation in SSIS SEL CIP tended to reduce instructional inference for students with low to moderately high initial scores (though adjusted treatment group differences were statistically significant only for baseline scores lower than .14 standard deviations below the mean) but not for students with high initial scores (> 2 SD above the mean).

Overall, effect sizes of the SSIS SEL CIP were small on positive social skill outcomes and teacher-rated problem behavior outcomes ($|g| \leq .11$; improvement indices < 4.4%; see Table 8). The effect size of the SSIS SEL CIP in reducing observed instructional interference ($g = -.23$) was somewhat larger and close to the threshold considered practically meaningful. The improvement index of -9.1 suggested that an average student in the control condition would have ranked about nine percentiles lower on instructional interference had they been exposed to the SSIS SEL CIP.

Discussion

The purpose of this randomized trial was to examine the effectiveness of a universal program intended to promote young students' prosocial development in primary classrooms. Specifically, first grade classrooms from participating schools were randomly assigned (by school) to the treatment (SSIS SEL CIP) or a business-as-usual control condition. Consistent with the study's primary goal of evaluating student outcomes when the SSIS SEL CIP is implemented in classrooms under routine conditions with typical levels of training and support, implementation teachers received no formal training or implementation guidance from the research team. A priori hypotheses were that students in implementing classrooms would demonstrate greater social skill proficiency and less negative social behavior at the end of the school year relative to their peers in business-as-usual comparison classrooms. Results were mixed relative to these hypotheses. Main effects tended to be non-significant and negligible in

size; however, several interactions indicated positive outcomes resulting from program exposure (increases in prosocial behavior, decreases in negative social behavior) for students with lower skills at baseline.

Prosocial Behavior Outcomes

In regard to students' prosocial behavior, findings from the current study were inconsistent with our original hypotheses in that there were no statistically significant main effects indicating the SSIS SEL CIP yielded greater improvements in students' behavior relative to students in the control classrooms. These findings also run somewhat counter to an efficacy trial in first grade classrooms (DiPerna et al., 2018) with the original edition of the SSIS Classwide Intervention program (Elliott & Gresham, 2008). Results of this prior trial indicated small-moderate positive effects in several prosocial domains (e.g., engagement, empathy), though they did not remain statistically significant after Benjamini-Hochberg adjustment.

The significant interactions between exposure to the SSIS SEL CIP and student-level baseline skills (responsibility and observed positive social behavior) in the current study were not present in the previous first grade efficacy trial (DiPerna et al., 2018). Similar interactions at the class-level (i.e., those with lower average skills at pretest demonstrated greater posttest gains from program exposure) were noted in these prosocial domains (as well as several others) in a prior efficacy trial featuring second grade classrooms (DiPerna et al., 2015). Interactions indicating small differences in outcomes (communication, social engagement) based on race in the current study also were not observed in previous efficacy trials (DiPerna et al., 2015, 2018) with the original edition of the SSIS CIP.

Negative Behavior Outcomes

With regard to negative social behavior, main-effect findings also were inconsistent with our original hypotheses of overall reductions in such behavior after exposure to the SSIS SEL CIP. Nonetheless, there were two statistically significant interactions demonstrating reductions in negative social (externalizing and bullying-related) behavior for students who exhibited higher levels of such behaviors at baseline. A third statistically significant interaction indicated that students with low-moderate baseline levels of disruptive/interfering behaviors in the classroom demonstrated reduced levels of such behaviors after participating in the SSIS SEL CIP; however, students with more extreme levels of interference behavior did not demonstrate significant reductions.

In the prior first grade efficacy trial (DiPerna et al., 2018), there were no statistically significant effects (main or interaction) in the problem behavior domains. In the prior second grade efficacy trial (DiPerna et al., 2015), there were statistically significant reductions (small-moderate effect) on internalizing behaviors; however, there were no statistically significant reductions in externalizing behaviors. In sum, though some findings from the current study suggest potential positive outcomes resulting from exposure to the SSIS SEL CIP under routine conditions (particularly for students experiencing some initial level of social difficulty), there are several inconsistencies with our initial predictions and results from prior efficacy trials featuring the original edition of the SSIS CIP. There are a number of important differences between the current and previous trials, though, that provide important context for situating the current findings.

Differences Between Current and Previous Trials

As noted previously, the current study was an effectiveness trial intended to examine student outcomes resulting from implementation of the SSIS SEL CIP by end users (teachers)

under routine conditions (typical levels and approaches to training, support, and implementation) in elementary classrooms. To achieve this goal, we recruited schools that were actively considering adoption of a universal SEL program in their primary grades. Although the project provided all program materials/resources to teachers free of charge (and on a delayed timeline for classrooms randomly assigned to business-as-usual comparison condition), the research team did *not* provide any professional development, coaching, feedback, or other forms of implementation support regarding program implementation. Instead, all supports were provided locally and decisions regarding implementation were made by grade-level teachers randomly assigned to the implementation condition. (Depending on the school, administrator and/or other support personnel may have also weighed in on training and implementation decisions.)

Although there were several notable similarities with the previous efficacy studies (e.g., classroom teachers were the interventionists who taught the SSIS CIP to their students) and the current trial, there were some important differences. Specifically, teachers in the efficacy trial were trained by research staff, expectations for implementation dosage (i.e., teaching approximately one SSIS CIP unit per week) were clearly communicated to teachers and administrators, and timely feedback was provided to teachers whose implementation (within lesson and/or overall progress within the curriculum) deviated from the expected protocol. With none of this external guidance, training, and support in place during the routine conditions of the current effectiveness trial, there was (not surprisingly) much more variability in implementation from an overall dosage perspective (number of units and lessons taught) as well as within-lesson modifications and adaptations. With regard to the former, the number of complete units (all 3 lessons per unit) taught by intervention teachers ranged from 0 to 11. Similarly, the overall number of lessons taught ranged from 7 to 36.

With regard to changes within lessons, although many teachers adhered closely to the lesson scripts, several reported (or were observed) making program changes, which were largely in the form of reductions (removing or condensing content or materials) and augmentations (adding material or content; Neugebauer et al., *in press*). Hunter et al. (2022) noted that changes and choices around implementation were made for a variety of reasons - both proactive (e.g., meeting perceived instructional needs of their students) and reactive (e.g., time constraints, unexpected developments such as fire drills). Thus, even though the mean number of lessons across the current sample was only somewhat lower than the dosage levels observed in the previous efficacy trials, there was much greater variability in implementation under the routine conditions of the current study, which in turn, may have resulted in the differences in findings. Such variability in implementation under routine conditions has been reported in other studies of universal SEL programs with efficacy evidence such as PATHS (Hennessey et al., 2020; Humphrey et al., 2016). Further, Wiglesworth et al. (2016) noted that greater effects were reported for most outcomes when studies were conducted under efficacy compared to effectiveness conditions.

Another potential consideration is what teachers in the business-as-usual comparison condition were doing with respect to social-emotional and social skill instruction in their classrooms. In the present study, we asked teachers in the control condition to specify on a weekly basis if they held any formal class meetings and/or other activities that primarily focused on social behavior at school. We also asked them if any of their meetings or activities focused on the same skills that the SSIS SEL CIP core lessons cover. On the weekly surveys completed during the treatment teachers' implementation period, comparison teachers reported holding an average of 3 class meetings per week (SD = 2) to discuss/teach positive classroom behaviors.

They also reported focusing on an average of 3 social skills that appear in the SSIS SEL CIP core units per week (*SD* = 3). While we are unable to directly compare the BAU practices in the current study and those in the previous SSIS CIP efficacy trials, SEL adoption, spending and implementation has grown substantially in the decade since those trials were completed (Bryant et al., 2021). Although the business-as-usual teachers in this sample did not report implementing any formal universal curricular programs focused on positive social behavior, it possible that growing awareness and emphasis of SEL practices and approaches generally facilitated increased informal opportunities for building students' SEL skills in the counterfactual classrooms relative to those in the previous study. It is also possible that adopting an intervention in some classrooms may cause schools to reallocate resources in a way that increases support or professional development for the other classrooms, possibly diluting an intervention effect (Jacob et al., 2019).

A third key difference between the current effectiveness trial and prior efficacy studies is that the prior efficacy trial featured the original version of the SSIS Classwide Intervention Program (Elliott & Gresham, 2007), and the current effectiveness trial featured the updated Social-Emotional Learning edition of the SSIS CIP (Elliott & Gresham, 2017). Though the original and updated SEL versions of the CIP are identical in regard to the instructional format (Tell, Show, Do…) and approaches (modeling, practice, feedback…), there are a few notable differences between these versions. For example, whereas the original version of the CIP included 10 skill units (30 lessons), the SEL edition used in the current trial included 13 additional advanced skill units (39 lessons) that were available for teachers to use in their classrooms. Although most of the skill units that teachers ultimately chose to teach in the current effectiveness trial were units that overlapped with those in the original CIP, approximately 37%

of teachers taught lessons from the new advanced skill units either in addition to, or in some cases in lieu of, some of the original skill units. Thus, though several classrooms demonstrated similar levels of dosage (i.e., lessons taught) to those in the previous efficacy trial, the skills taught still varied across at least some of those classrooms.

Beyond the additional content variability introduced by the availability of the advanced skill units in the CIP SEL edition, there are some differences between the materials of the current and original versions of the program that may have implications for how teachers use/implement the CIP SEL lessons. For example, the current version of the program provides digital resources to teachers (e.g., PowerPoints) that were not available in the original edition. In addition, though a printed Teachers' Implementation Guide is still included with the SEL edition, the instructional scripts were revised from the original 2008 version to make them easier for teachers to follow. Similarly, the SEL edition provides downloadable digital student response sheets that are streamlined relative to the printed student response booklets that were included in the original SSIS CIP. Though we do not have sufficient data within the current study to determine to what extent, if any, differences between the SEL and original versions of the SSIS CIP may have contributed to the outcome differences between the current study and previous efficacy trials, they are potentially important factors for examination in future studies and consideration for adopters of SSIS SEL CIP.

Limitations

There are a few key limitations within which the current findings must also be considered. First and perhaps foremost, the effectiveness trial began with the 2018-19 school year and was to enroll new cohorts of schools and classrooms each school year through the 2021-22 school year. Those plans changed dramatically during the second year of enrollment (2019-20

school year) when COVID-19 resulted in the closure of most U.S. schools in March of 2020. As a result, we had to stop our study shortly thereafter which resulted in the loss of 6 schools and approximately 20 first grade classrooms where baseline data were collected but teachers were unable to complete much (and in some classrooms any) implementation of the program. With the continuation of the pandemic into 2022, we were unable to resume working with new cohorts of schools and classrooms. Although the sample reported in this article is very consistent at the student, teacher, and school levels regarding the planned demographic diversity for our original target population, it is approximately one-third of the planned sample size for our Grade 1 effectiveness trial. Thus, the power to detect statistical significance and ability to determine more precise estimates of effects were lower than envisioned for this initial trial, and it is unknown if results would differ if the original planned sample size could have been obtained.

Beyond the sample limitations resulting from the pandemic, the nature of the effectiveness trial also created some challenges from a data collection perspective. For example, though we used multiple methods to assess fidelity, quality, and dosage throughout the implementation phase of the project, in some cases it was difficult to ascertain what changes teachers made, when, and for what reasons. Similarly, variability of implementation timelines due to local implementation decisions created some challenges for coordination of our assessment windows and data collection efforts. In addition, though the research team provided no implementation support, some of the necessary data collection practices may have affected implementation. For example, having teachers complete weekly surveys to document their program completion and adherence, or having field staff observe lesson delivery, may have caused teachers to adjust their practices. While tracking implementation is necessary as a

research procedure, we are not able to rule out its unintended effect on teachers' typical practices.

Finally, though our measurement battery has strong psychometric evidence to support its use, the only behavior measures that were collected for all participants were solely from the perspective of the teachers and may have been influenced by other factors (e.g., perceptions of the SSIS SEL CIP or the research project itself). Relatedly, resources only allowed us to collect observation data for a randomly selected subsample of participants, and each observation represented a sample of students' time in the classroom.

Future Directions

There are several important directions for future research considering the current findings. Given the target sample was unable to be obtained prior to the onset of the pandemic and the differences in mean outcomes between the effectiveness and prior efficacy trials, replication of the current study with larger samples would be helpful to further elucidate outcomes resulting from implementation of the SSIS SEL CIP. Such replication(s) would not only provide further insight regarding the effects of this universal program when implemented under routine classroom conditions but also the individual and system-level factors that influence local implementation decisions.

Multi-site effectiveness studies are crucial to elucidating potential challenges in implementation, variations in local approaches, and/or differences in impact when translating interventions into settings with real-world limitations on resources and supports (Center for Early Learning & Public Health, 2021; Curran et al., 2012; National Center for Special Education Research, 2016). Future studies with additional diverse student samples, especially those that incorporate hybrid designs that evaluate not only outcomes but also implementation in routine

practice, could help identify if certain implementation patterns or profiles exist under routine conditions.

As noted previously, although the SSIS CIP effectiveness and efficacy trials share a number of similarities (e.g., teachers as implementers, scripted materials intended for use with minimal training and support), there was significantly more variability in implementation – both dosage and instructional adherence – in the current trial. As such a key future direction is identifying if certain implementation patterns/profiles tend to emerge under routine conditions, and if so, do any consistently promote more positive outcomes for young students. Relatedly, if flexible approaches to implementation are necessary for universal programs to be responsive to local contextual and cultural considerations, results from current and future studies can help inform what core components (Wigelsworth et al., 2020) of the SSIS SEL CIP must be maintained to ensure that the program is as effective as possible when implemented under authentic conditions in primary classrooms. If such components or features can be identified, studies featuring adaptive designs (NASEM, 2022) could be undertaken to confirm their critical role in promoting positive prosocial outcomes for students experiencing the SSIS SEL CIP within their classroom.

Conclusions

Results of the current trial suggest that the main effects of the SSIS SEL CIP are neither statistically nor practically significant when the program is implemented under typical conditions in first grade classrooms within schools considering adoption of universal SEL programs and practices. Though several interactions were observed based on student pretest scores and race/ethnicity, only a few remained statistically significant post-BH correction for potential Type-1 error inflation. These interactions suggest some positive benefit for students exhibiting

lower initial levels of responsibility and observed positive social behavior and for White students in the communication and social engagement domains.

Although findings from an earlier efficacy trial with the first edition of the SSIS CIP in first grade classrooms yielded small positive (main) effects in several prosocial domains, many did not remain statistically significant post-BH correction and effect size confidence intervals ranged from negligible (and in some cases negative in valence) to large positive. In addition, all effect size confidence intervals from the current trial included 0 and consistently ranged from moderate negative to moderate positive. Though there is some overlap in findings with previous efficacy trials of the SSIS CIP in primary grades, there are a number of differences as well. Potential explanations for these differences are greater implementation variability (dosage, adaptations, and approach) and some changes in format of the current SSIS SEL CIP edition (PowerPoint lessons, online distribution of materials) relative to the earlier edition. Given we cannot definitively determine which of these, or perhaps other unmeasured, factors accounted the observed differences, we recommend that adopters of the SSIS SEL CIP maintain similar implementation practices to the original trial (DiPerna et al., 2018) and collect local data to evaluate student outcomes associated with implementation of the program. In addition, we encourage additional effectiveness studies of the SSIS SEL CIP and other universal programs to better understand their (a) actual implementation and (b) associated student outcomes when delivered under typical classroom and school conditions. Such studies will allow potential adopters to better understand common approaches to implementation and the critical components or features necessary to yield positive outcomes while allowing for flexible adaptations in response to local contextual considerations.

References

Aarons, G. A., Sklar, M., Mustanski, B., Benbow, N., & Brown, C. H. (2017). "Scaling-out" evidence-based interventions to new populations or new health care delivery systems. *Implementation Science, 12*(1), 1–13. https://doi.org/10.1186/s13012-017-0640-6

Al-Ubaydli, O., List, J. A., & Suskind, D. (2019). The science of using science: Towards an understanding of the threats to scaling experiments (No. 2019–73). https://doi.org/10.2139/ssrn.3391481

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

Briesch, A. M., Volpe, R. J., & Floyd, R. G. (2018). *School-based observation: A practical guide to assessing student behavior.* New York, NY: Guilford.

Bryant, G., Mainelli, A., Crowley, S., Glennen, C., & Edzie, K. (2021). *Finding your place 2021: Social emotional learning takes center stage in K-12*. Tyton Partners. https://tytonpartners.com/post-pandemic-schooling-emphasizing-social-emotional-learning/

Byrt, T., Bishop, J., & Carlin, J. B. (1993). Bias, prevalence and kappa. *Journal of Clinical Epidemiology, 46*(5), 423–429. https://doi.org/10.1016/0895-4356(93)90018-v

Curran, G. M., Bauer, M., Mittman, B., Pyne, J. M., & Stetler, C. (2012). Effectiveness-implementation hybrid designs: Combining elements of clinical effectiveness and implementation research to enhance public health impact. *Medical Care, 50*(3), 217–226. https://doi.org/10.1097/MLR.0b013e3182408812

Flay, B. R., Biglan, A., Boruch, R. F., Castro, F. G., Gottfredson, D., Kellam, S., Mościcki, E. K., Schinke, S., Valentine, J. C., & Ji, P. (2005). Standards of evidence: Criteria for efficacy, effectiveness and dissemination. *Prevention Science, 6*(3), 151–175. https://doi.org/10.1007/s11121-005-5553-y

Center for Early Learning & Public Health. (2021). *How a better understanding of scaling can enhance early childhood policy and improve lives.* TMW Center for Early Learning and Public Health, The University of Chicago. https://tmwcenter.uchicago.edu/resources/how-a-better-understanding-of-scaling-can-enhance-early-childhood-policy-and/

Conduct Problems Prevention Research group. (2010). The effects of a multiyear universal social-emotional learning program: The role of student and school characteristics. *Journal of Consulting and Clinical Psychology*, *78*(2), 156–168. https://doi.org/10.1037/a0018607

Dettmer, A. M., Taylor, K. A., & Chhin, C. S. (2017). *Responses to the request for public comment on IES-funded efficacy, replication, and effectiveness studies.* https://ies.ed.gov/whatsnew/pdf/Response_to_Public_Comment_on_Goals_3_and_4.pdf

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., 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), 1–16. https://doi.org/10.1037/edu0000191

Domitrovich, C. E., Gest, S. D., Jones, D., Gill, S., & DeRousie, R. M. S. (2010). Implementation quality: Lessons learned in the context of the Head Start REDI trial. *Early Childhood Research Quarterly, 25*(3), 284–298. https://doi.org/10.1016/j.ecresq.2010.04.001

D'Souza, K. (2021). What happens to children who missed kindergarten during Covid-19 crisis? *EdSource*, 5–11. https://edsource.org/2021/what-happens-to-children-who-missed-kindergarten-during-covid-19-crisis

Durlak, J. A. (2015). Studying program implementation is not easy but it is essential. *Prevention Science, 16*(8), 1123–1127. https://doi.org/10.1007/s11121-015-0606-3

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. https://doi.org/10.1007/s10464-008-9165-0

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

Elias, M. J. (2019). What if the doors of every schoolhouse opened to social-emotional learning tomorrow: Reflections on how to feasibly scale up high-quality SEL. *Educational Psychologist*, *54*(3), 233–245. https://doi.org/10.1080/00461520.2019.1636655

Elliott, S. N., & Gresham, F. M. (2007*). SSIS: Classwide Intervention Program*. Pearson.

Elliott, S. N., & Gresham, F. M. (2017). *SSIS SEL Edition Classwide Intervention Program manual*. Pearson.

Embry, D. D., & Biglan, A. (2008). Evidence-based kernels: Fundamental units of behavioral influence. *Clinical Child and Family Psychology Review*, *11(3),* 75-113. doi: 10.1007/s10567-008-0036-x

Evans, R., Murphy, S., & Scourfield, J. (2015). Implementation of a school-based social and emotional learning intervention: Understanding diffusion processes within complex systems. *Prevention Science, 16*(5), 754–764. https://doi.org/10.1007/s11121-015-0552-0

Greenberg, E. (2021). An equitable return to early education demands a smart approach to data. *Urban Wire: The Blog of the Urban Institute*. https://www.urban.org/urban-wire/equitable-return-early-education-demands-smart-approach-data

Greenberg, M. T. (2010). School-based prevention: Current status and future challenges. *Effective Education, 2*(1), 27–52. https://doi.org/10.1080/19415531003616862

Gresham, F., & Elliott, S. N. (2008*). Social Skills Improvement System rating scales.* Pearson Assessments.

Gresham, F. M., Elliott, S. N., Cook, C. R., Vance, M. J., & Kettler, R. (2010). Cross-informant agreement for ratings for social skill and problem behavior ratings: An investigation of the Social Skills Improvement System-Rating Scales. *Psychological Assessment, 22*(1), 157–166. https://doi.org/10.1037/a0018124

Gresham, F. M., & Elliott, S. N. *SIS SEL Edition Manual.* Pearson Assessments.

Gresham, F. M., Elliott, S. N., Metallo, S., Byrd, S., Wilson, E., & Cassidy, K. (2018). Cross-informant agreement of children's social-emotional skills: An investigation of ratings by teachers, parents, and students from a nationally representative sample. *Psychology in the Schools, 55*(2), 208–223. https://doi.org/10.1002/pits.22101

Hamre, B. K., & Pianta, R. C. (2005). Can instructional and emotional support in the first-grade classroom make a difference for children at risk of school failure? *Child Development, 76*(5), 949–967. https://doi.org/10.1111/j.1467-8624.2005.00889.x

Hamre, B. K., Pianta, R. C., Mashburn, A. J., & Downer, J. T. (2007). Building a science of classrooms: Application of the CLASS framework in over 4,000 early childhood and elementary classrooms. New York, NY: Foundation for Child Development.

Hennessey, A., & Humphrey, N. (2020). Can social and emotional learning improve children's academic progress? Findings from a randomised controlled trial of the Promoting Alternative Thinking Strategies (PATHS) curriculum. *European Journal of Psychology of Education, 35*(4), 751–774. https://doi.org/10.1007/s10212-019-00452-6

Herbers, J. E., Hayes, K. R., & Cutuli, J. J. (2021). Adaptive systems for student resilience in the context of COVID-19. *School Psychology*, *36*(5), 422–426. https://doi.org/10.1037/spq0000471

Horner, R. H., Sugai, G., & Fixsen, D. L. (2017). Implementing effective educational practices at scales of social importance. *Clinical Child and Family Psychology Review, 20*(1), 25–35. https://doi.org/10.1007/s10567-017-0224-7

Humphrey, N., Barlow, A., Wigelsworth, M., Lendrum, A., Pert, K., Joyce, C., Stephens, E., Wo, L., Squires, G., Woods, K., Calam, R., & Turner, A. (2016). A cluster randomized controlled trial of the Promoting Alternative Thinking Strategies (PATHS) curriculum. *Journal of School Psychology, 58*, 73–89. https://doi.org/10.1016/j.jsp.2016.07.002

Humphrey, N., Kalambouka, A., Wigelsworth, M., Lendrum, A., Deighton, J., & Wolpert, M. (2011). Measures of social and emotional skills for children and young people: A

systematic review. *Educational and Psychological Measurement, 71*(4), 617–637. https://doi.org/10.1177/0013164410382896

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), 147–154. https://doi.org/10.1037/spq0000232

Hunter, L. J., DiPerna, J. C., Hart, S. C., Neugebauer, S., & Lei, P. (2022). Examining Teacher Approaches to Implementation of a Classwide SEL Program. *School Psychology, 37*(4), 285–297. https://doi.org/10.1037/spq0000502

Jacob, R. T., Doolittle, F., Kemple, J., & Somers, M. A. (2019). A framework for learning from null results. *Educational Researcher, 48*(9), 580–589. https://doi.org/10.3102/0013189X19891955

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. https://doi.org/0279-6015

Jones, S. M., & Bouffard, S. M. (2012). Social and emotional learning in schools: From programs to strategies. *Social Policy Report, 26*(4), 1–33. https://doi.org/10.1002/j.2379-3988.2012.tb00073.x

Jones, D. E., Crowley, D., & Greenberg, M. (2017). *Improving social emotional skills in childhood enhances long-term well-being and economic outcomes [Issues Brief}*, *June*, 1–11.

Jones, D. E., Greenberg, M., & 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

Jones, J. M. (2021). The dual pandemics of COVID-19 and systemic racism: Navigating our path forward. *School Psychology*, *36*(5), 427–431.

Jones, S. M. & Doolittle, E. J. (2017). Social and emotional learning: Introducing the issue. *The Future of Children, 27*(1), 3-12. https://www.jstor.org/stable/44219018

Kamenetz, A., Treviño, M. A., & Bakeman, J. (2020, October 9). Enrollment is dropping in public schools around the country. *NPR*. https://www.npr.org/2020/10/09/920316481/enrollment-is-dropping-in-public-schools-around-the-country

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*(6), 463–477. https://doi.org/10.1016/j.jsp.2015.09.002

Malti, T., Ribeaud, D., & Eisner, M. P. (2011). The effectiveness of two universal preventive interventions in reducing children's externalizing behavior: A cluster randomized controlled trial. *Journal of Clinical Child and Adolescent Psychology, 40*(5), 677–692. https://doi.org/10.1080/15374416.2011.597084

McGraw Hill Education. (2021). *2021 Social and emotional learning report*. https://mheducation.com/sel-survey

Morgan, P. L., Farkas, G., Tufis, P. a., & Sperling, R. A. (2008). Are reading and behavior problems risk factors for each other? *Journal of Learning Disabilities, 41*(5), 417–436. https://doi.org/10.1177/0022219408321123

National Academies of Sciences, Engineering, and Medicine [NASEM] (2022). *The future of education research at IES: Advancing an equity-oriented science.* Washington, DC: The National Academies Press. https://doi.org/10.17226/26428

National Center for Special Education Research, Technical Working Group Meeting Summary. (2016, October 14). Building evidence: What comes after an efficacy study? https://files.eric.ed.gov/fulltext/ED580305.pdf

National Scientific Council on the Developing Child (2007). *The science of early childhood development.* https://developingchild.harvard.edu/resources/the-science-of-early-childhood-development-closing-the-gap-between-what-we-know-and-what-we-do/

Neugebauer, S. R., Sandilos, L. E., DiPerna, J. C., Hunter, L. J., & Hart, S. C. (in press). 41 teachers, 41 different ways: Exploring teacher implementation of a universal social-emotional learning program under routine conditions. *Elementary School Journal.*

O'Connor, E. E., Cappella, E., McCormick, M. P., & McClowry, S. G. (2014). An examination of the efficacy of INSIGHTS in enhancing the academic and behavioral development of children in early grades. *Journal of Educational Psychology, 106*(4), 1156–1169. https://doi.org/10.1037/a0036615

Osher, D., Kidron, Y., Brackett, M., Dymnicki, A., Jones, S., & Weissberg, R. P. (2016). Advancing the science and practice of social and emotional learning: Looking back and moving forward. *Review of research in education*, (Volume 40, 644–681). American Education Research Association. https://doi.org/10.3102/0091732X16673595

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

Rabiner, D. L., Godwin, J., & Dodge, K. A. (2016). Predicting academic achievement and attainment: The contribution of early academic skills, attention difficulties, and social competence. *School Psychology Review, 2*, 250–267. https://doi.org/10.17105/SPR45-2.250-267

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

Sandilos, L. E., Shervey, S. W., DiPerna, J. C., Lei, P., & Cheng, W. (2016). Structural validity of CLASS K-3 in primary grades: Testing alternative models. *School Psychology Quarterly, 32*(2), 226–239. https://doi.org/10.1037/spq0000155

Sanetti, L. M. H., & Collier-Meek, M. A. (2019). Increasing implementation science literacy to address the research-to-practice gap in school psychology. *Journal of School Psychology, 76*(July), 33–47. https://doi.org/10.1016/j.jsp.2019.07.008

Sim, J., & Wright, C. C. (2005). The kappa statistic in reliability studies: use, interpretation, and sample size requirements. *Physical Therapy, 85*(3), 257–268.

Sklad, M., Diekstra, R., De Ritter, M., Ben, J., & Gravesteijn, C. (2012). Effectiveness of school-based universal social, emotional, and behavioral programs: Do they enhance students' development in the area of skill, behavior, and adjustment? *Psychology in the Schools, 49*(9), 892–909. https://doi.org/10.1002/pits

Spivak, A. L., & Farran, D. C. (2012). First-grade teacher behaviors and children's prosocial actions in classrooms. *Early Education and Development*, *23*(5), 623–639. https://doi.org/10.1080/10409289.2011.566164

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

U.S. Department of Education, Institute of Education Sciences. (2022). *Search funded research grants and contracts.* https://ies.ed.gov/funding/grantsearch

U.S. Department of Education, Office of Special Education and Rehabilitative Services. (2021). *Supporting child and student social, emotional, behavioral, and mental health needs*. https://www2.ed.gov/documents/students/supporting-child-student-social-emotional-behavioral-mental-health.pdf

Volpe, R. J., & DiPerna, J. C. (2018). Cooperative Learning Observation Code for Kids-2. Unpublished observation code.

What Works Clearinghouse. (2022). *What Works Clearinghouse procedures and standards handbook*, *version 5.0.* U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance (NCEE). https://ies.ed.gov/ncee/wwc/Handbooks

Wigelsworth, M., Lendrum, A., Oldfield, J., Scott, A., ten Bokkel, I., Tate, K., & Emery, C. (2016). The impact of trial stage, developer involvement and international transferability on universal social and emotional learning programme outcomes: A meta-analysis. *Cambridge Journal of Education, 46*(3), 347–376. https://doi.org/10.1080/0305764X.2016.1195791

Wigelsworth, M., Mason, C., Verity, L., Qualter, P., & Humphrey, N. (2022). Making a case for core components: New frontiers in SEL theory, research, and practice. *School Psychology Review*, Advance online publication. https://doi.org/10.1080/2372966X.2021.2004863

Wollersheim Shervey, S., Sandilos, L. E., DiPerna, J. C., & Lei, P. (2017). Social validity of the Social Skills Improvement System-Classwide Intervention Program (SSIS-CIP) in the primary grades. *School Psychology Quarterly. 32*(3), 414–421. https://doi.org/10.1037/spq0000203

Wrabel, S., Hamilton, L., Whitaker, A., & Grant, S. (2018). *Investing in evidence-based social and emotional learning*. Rand Corporation. https://doi.org/10.7249/rr2739

Zieher, A. K., Cipriano, C., Meyer, J. L., & Strambler, M. J. (2021). Educators' implementation and use of social and emotional learning early in the COVID-19 pandemic. *School Psychology*, *36*(5), 388–397. https://doi.org/10.1037/spq0000461

Figure 1

Flow of Participants through the SSIS SEL CIP Effectiveness Trial

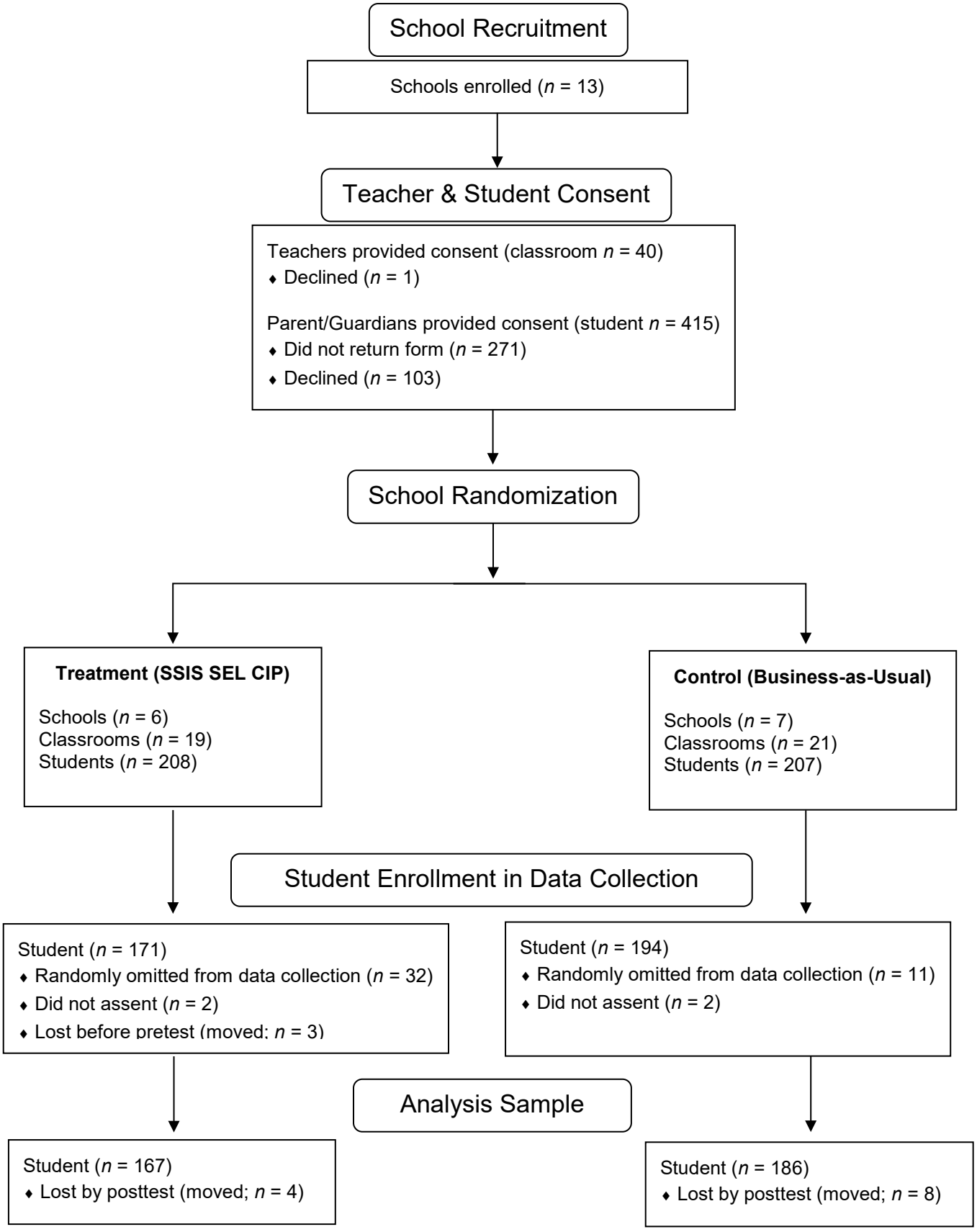

Figure 2

Interaction between Student-level Baseline and Experimental Condition on Responsibility

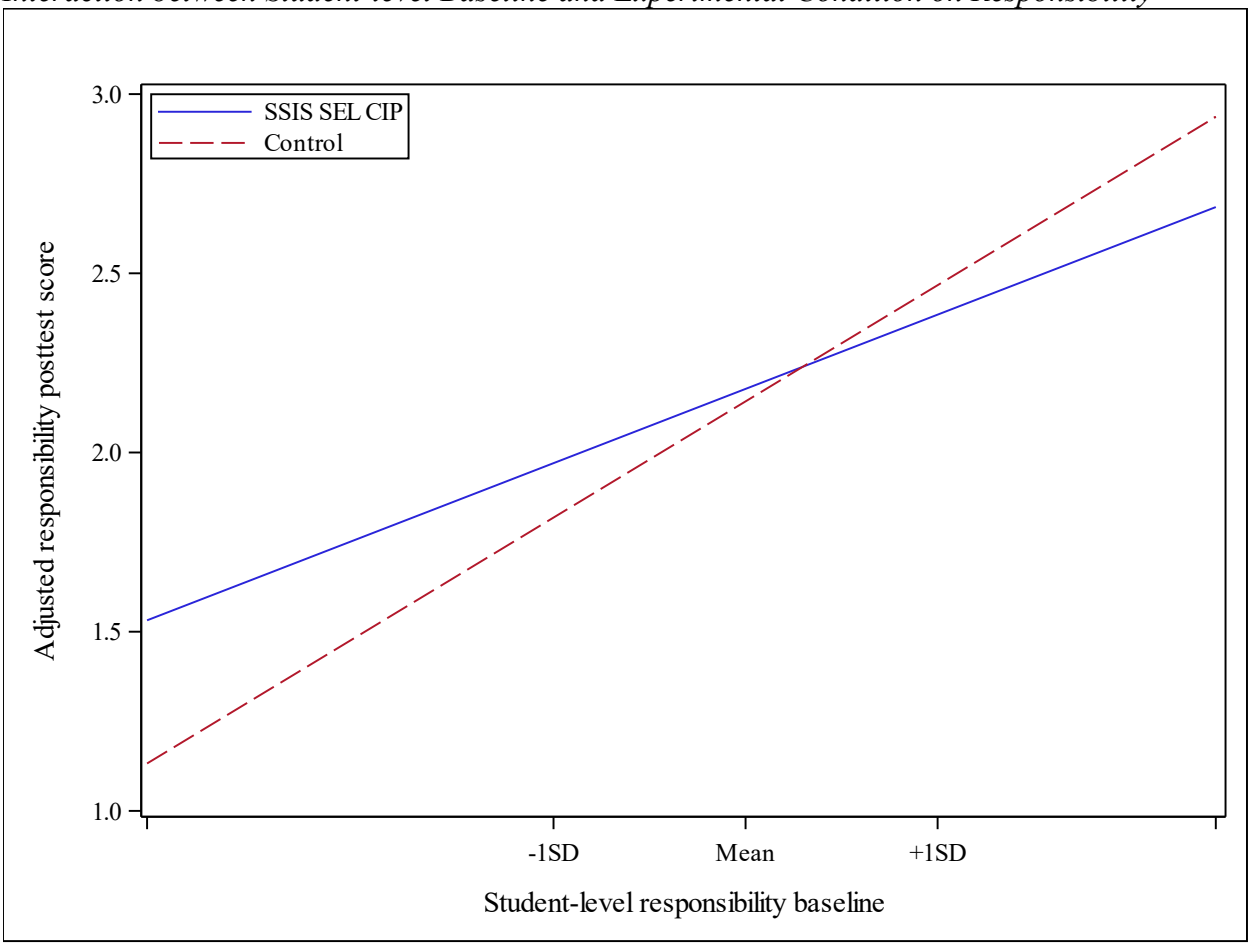

Figure 3

Interaction between Student-level Baseline and Experimental Condition on Positive Social

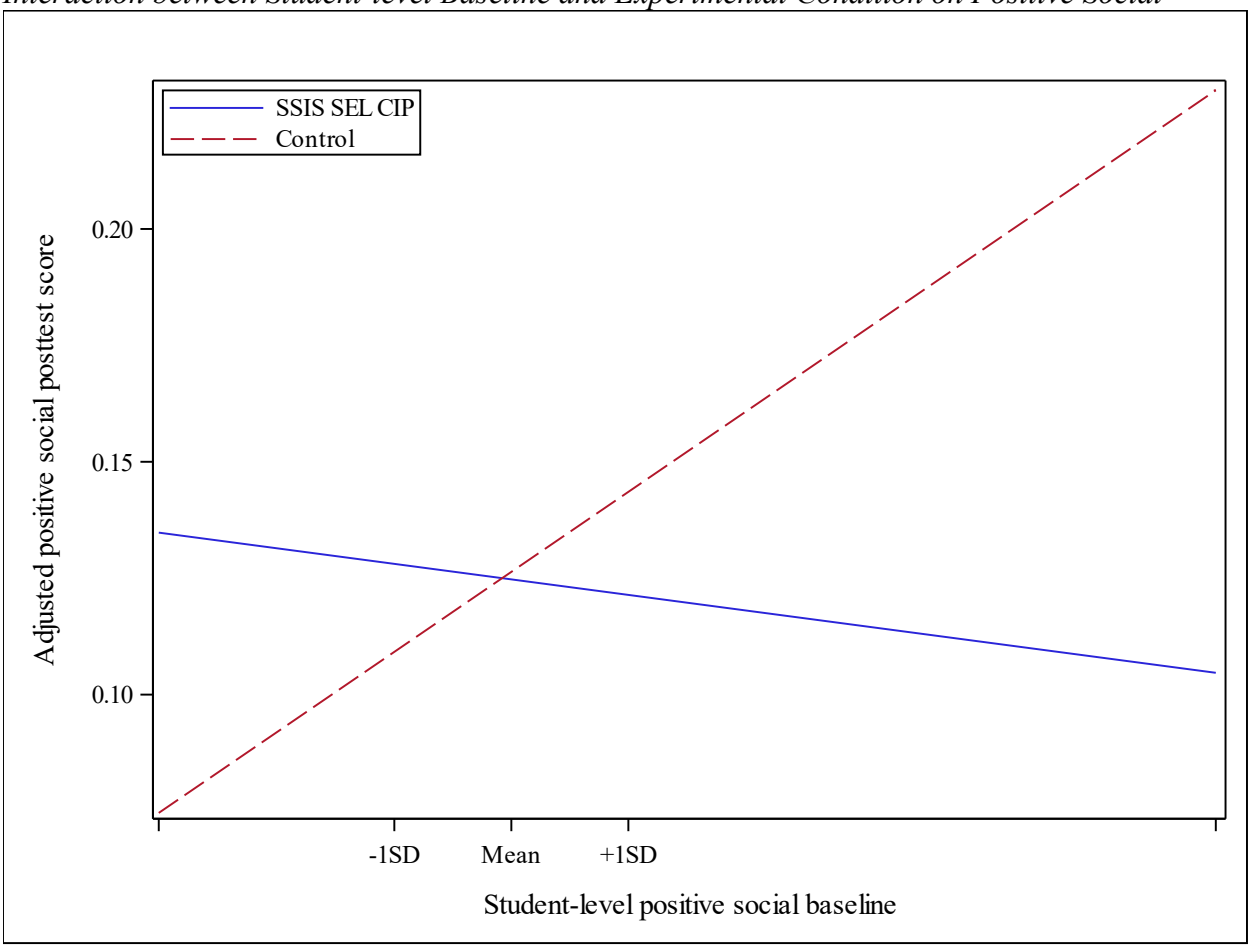

Figure 4

Interaction between Student Race and Experimental Condition on Communication

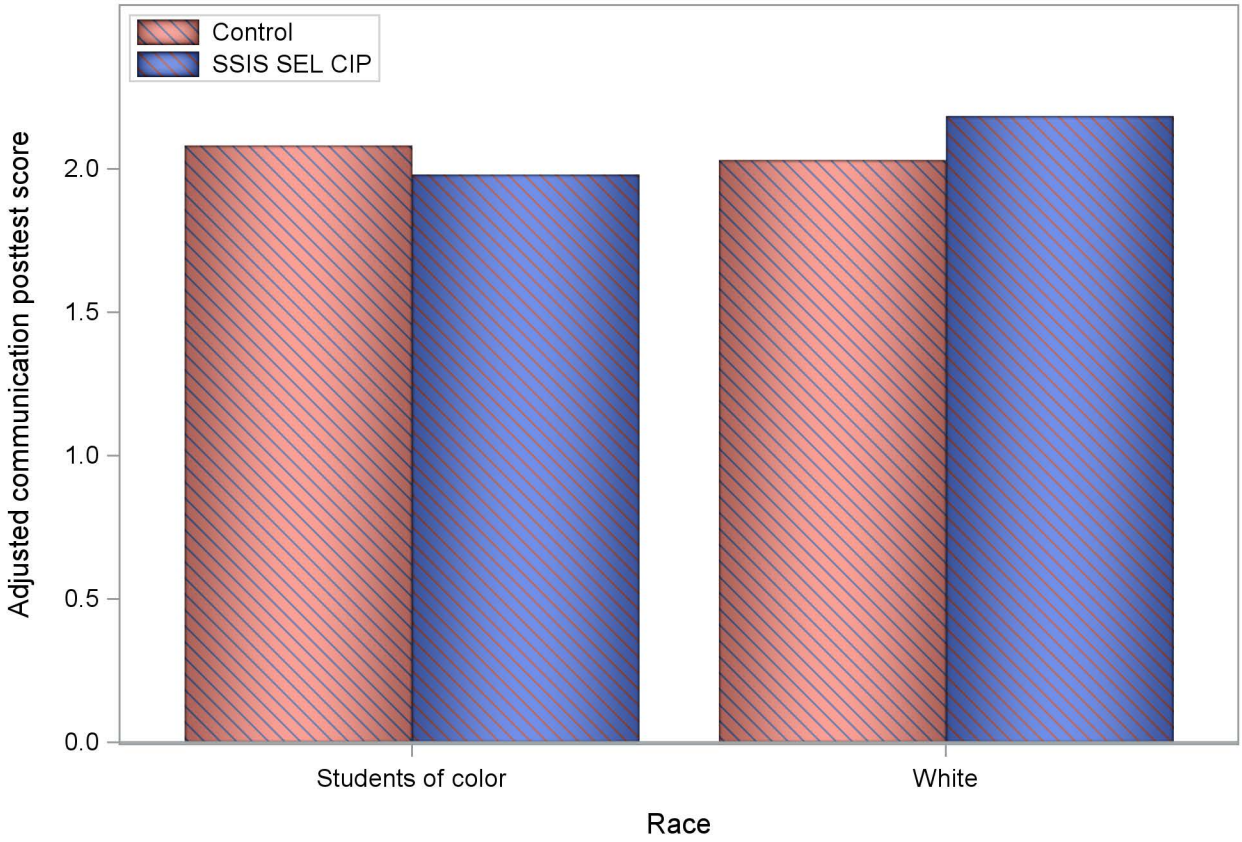

Figure 5

Interaction between Student Race and Experimental Condition on Social Engagement

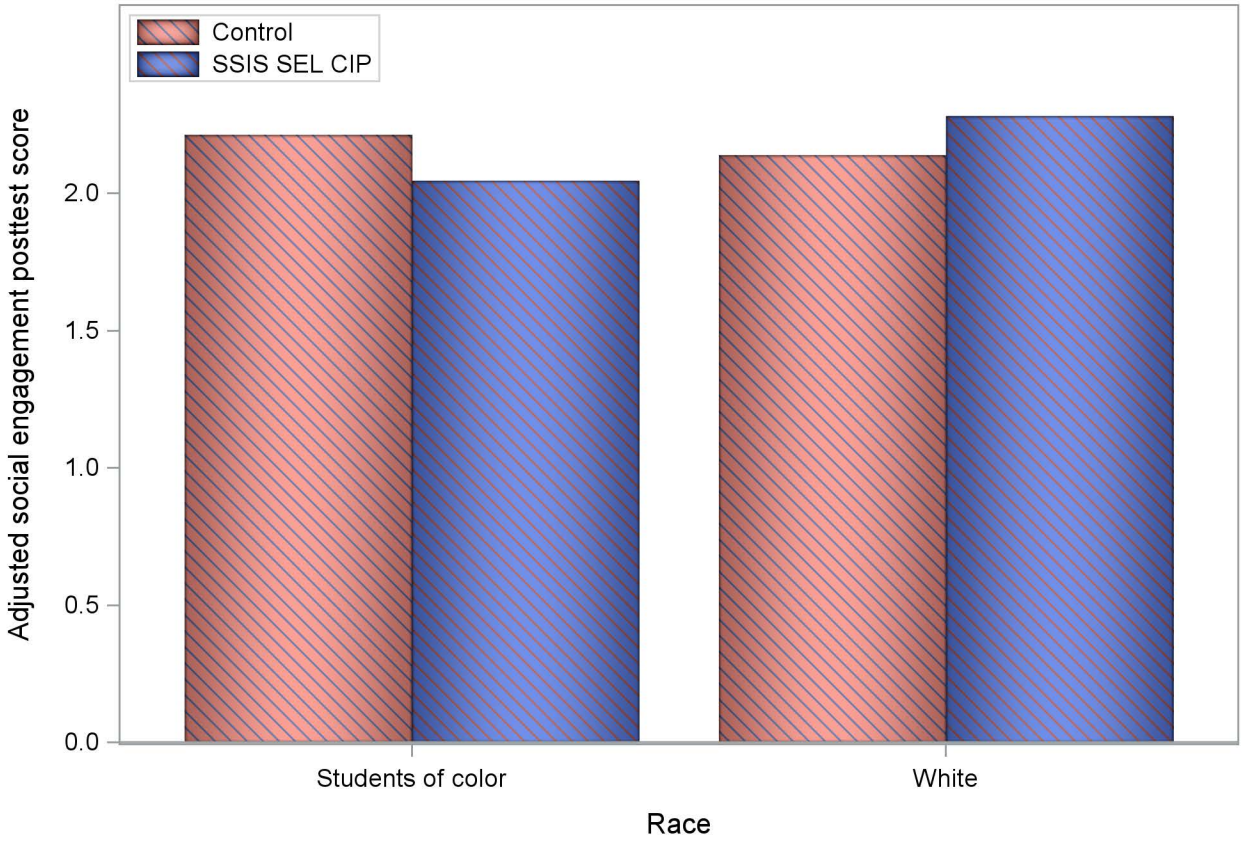

Figure 6

Interaction between Student-level Baseline and Experimental Condition on Problem Behaviors

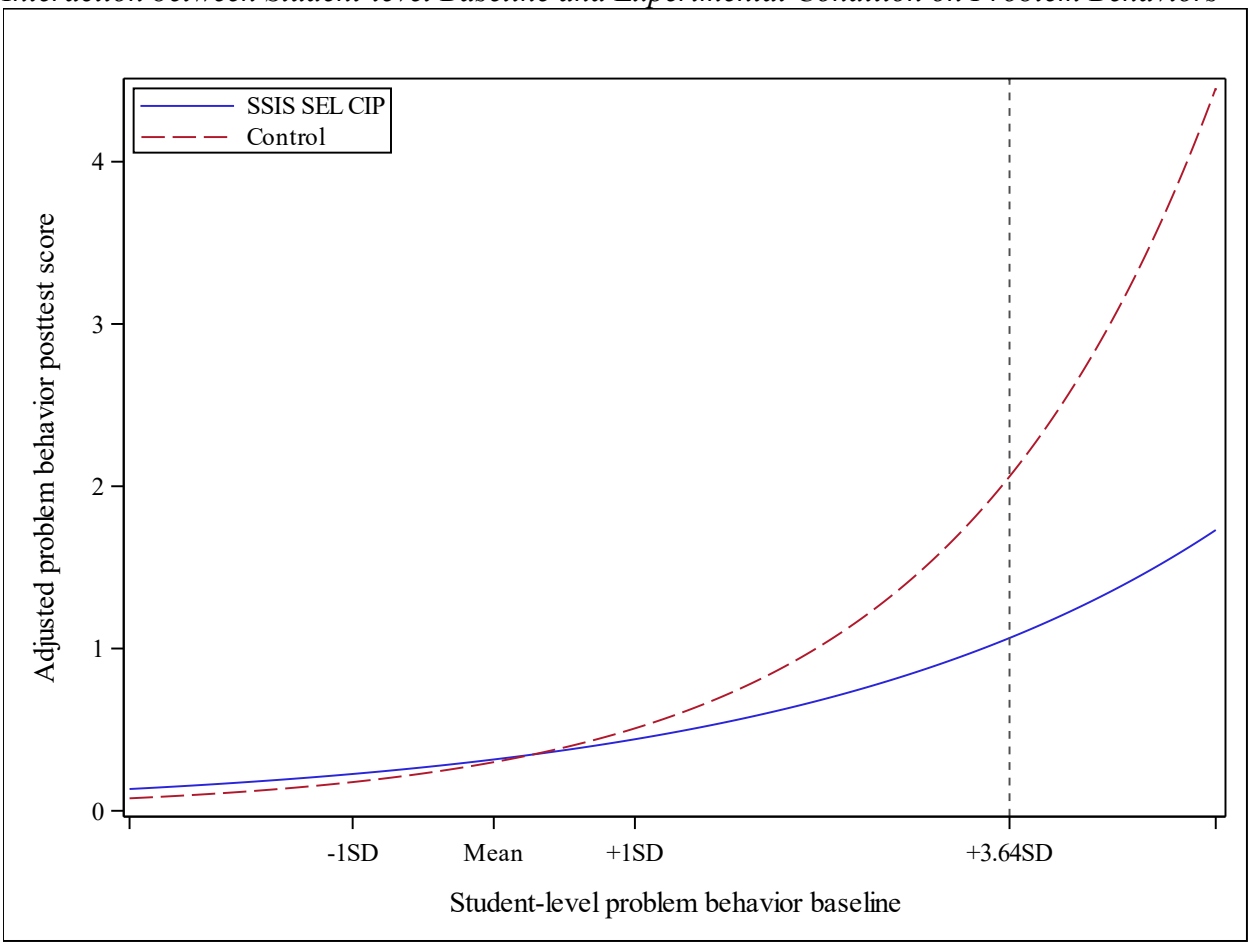

Figure 7

Interaction between Student-level Baseline and Experimental Condition on Externalizing

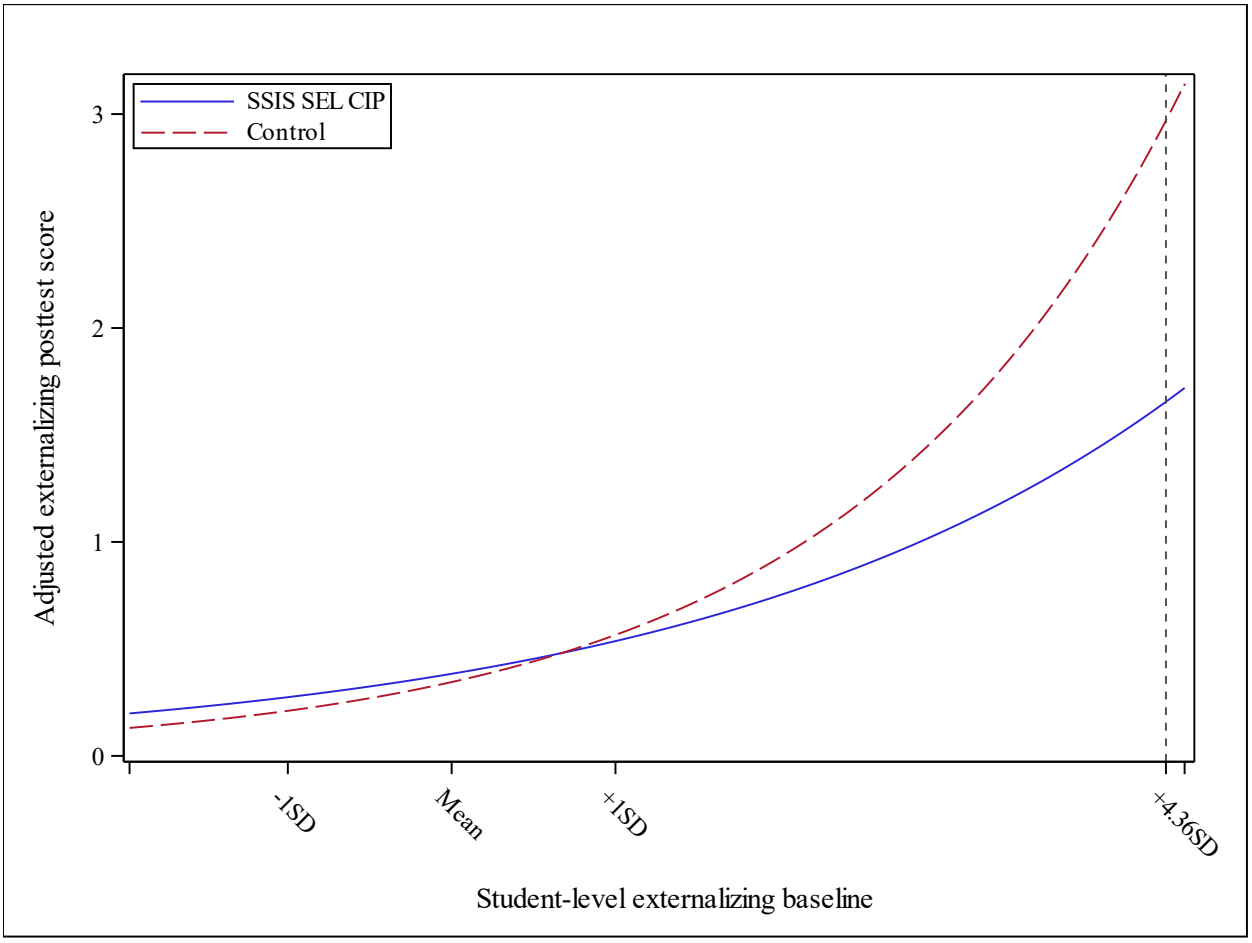

Figure 8

Interaction between Student-level Baseline and Experimental Condition on Bullying

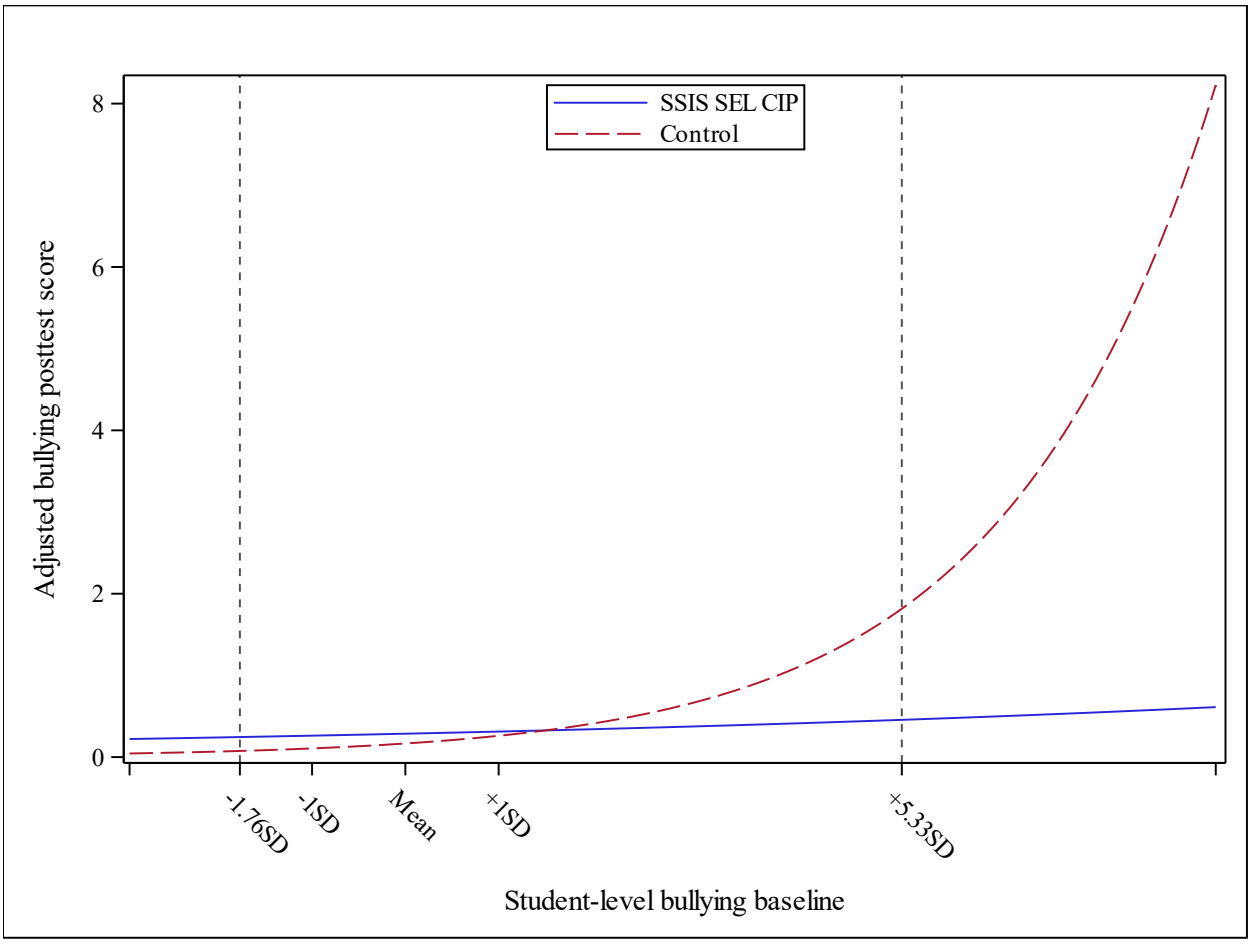

Figure 9

Interaction between Student-level Baseline and Experimental Condition on Hyperactive-inattentive

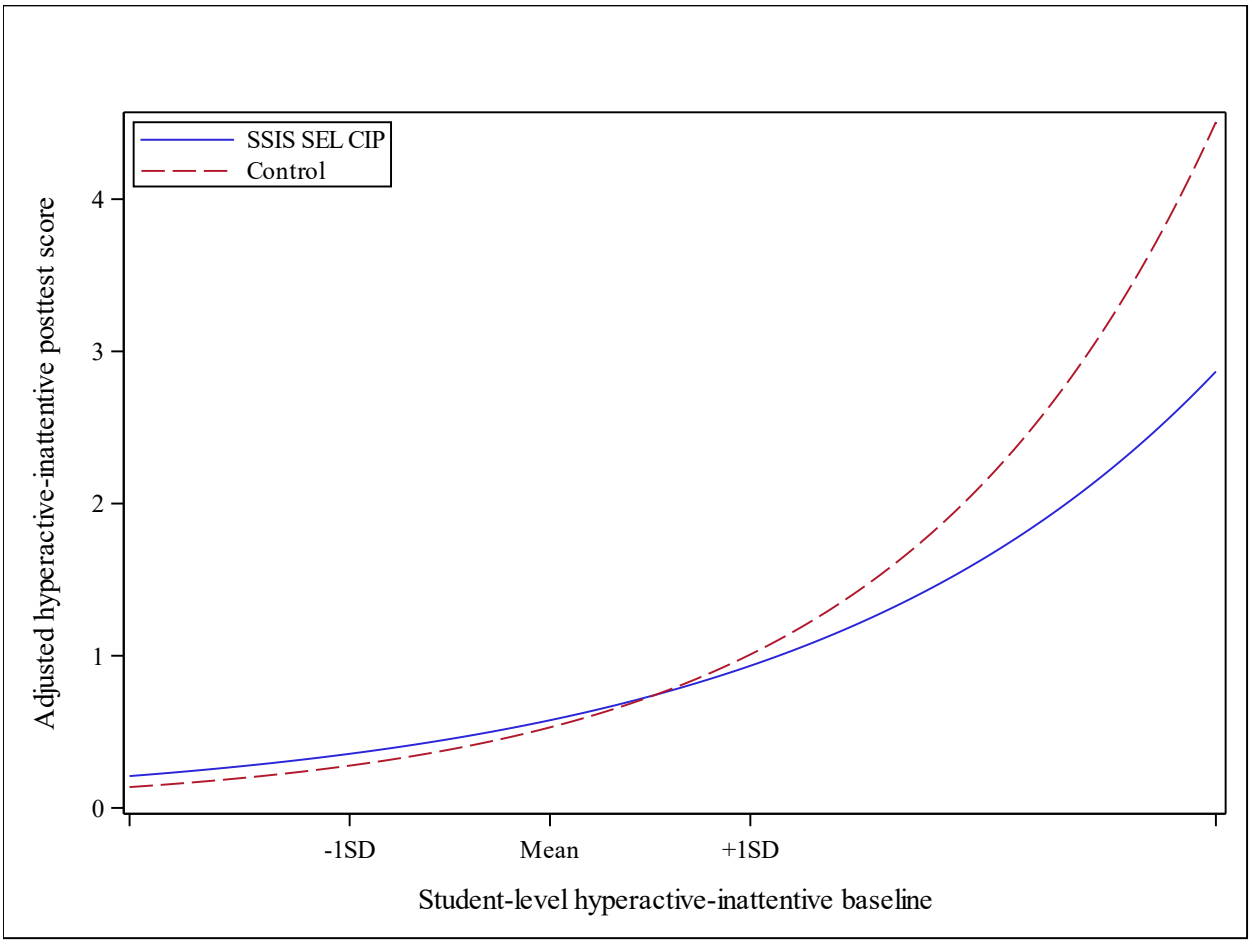

Figure 10

Interaction between Student-level Baseline and Experimental Condition on Instructional Interference