School Psychology Quarterly

Evaluation of Second Step on Early Elementary Students' Academic Outcomes: A Randomized Controlled Trial

Clayton R. Cook, Sabina Low, Joanne Buntain-Ricklefs, Kelly Whitaker, Michael D. Pullmann, and Jaclyn Lally

Online First Publication, May 24, 2018. http://dx.doi.org/10.1037/spq0000233

School Psychology Quarterly

1045-3830/18/$12.00
http://dx.doi.org/10.1037/spq0000233

Evaluation of *Second Step* on Early Elementary Students' Academic Outcomes: A Randomized Controlled Trial

Clayton R. Cook
University of Minnesota

Sabina Low
Arizona State University

Joanne Buntain-Ricklefs, Kelly Whitaker, Michael D. Pullmann, and Jaclyn Lally
University of Washington

Research has consistently linked social–emotional learning to important educational and life outcomes. Early elementary represents an opportune developmental period to proactively support children to acquire social–emotional skills that enable academic success. Using data from a large scale randomized controlled trial, the purpose of this study was to investigate the impact of the 4th edition of *Second Step* on early elementary students' academic-related outcomes. Participants were Kindergarten to 2nd grade students in 61 schools (310 teachers; 7,419 students) across six school districts in Washington State and Arizona. Multilevel models (Time × Condition) indicated the program had no positive main effect impact on academic outcomes. However, moderator analyses revealed that quality of implementation, specifically a measure of student engagement and dosage, was found to be associated with significant, albeit small, reading and classroom behavior outcomes. Findings from this study provide support for *Second Step* when implemented in the context of high engagement and higher dosage to have small but potentially meaningful collateral impact on early academic-related outcomes.

Impact and Implications

This study examined the collateral impact of a widely used social–emotional learning (SEL) program (*Second Step*) on early elementary children's academic outcomes. Findings emphasized the importance of specific dimensions of fidelity that may be associated with outcomes, as well as additional research that focused on developing a better understanding of the degree to which SEL programming impacts children's early academic performance.

Keywords: social emotional learning, social skills, fidelity of implementation, *Second Step*

In the school context, social, emotional, and behavioral problems present significant, immediate challenges to teaching and learning. Although recent federal and state mandates have prompted increased accountability of academic instruction and student outcomes (e.g., Common Core Standards, teacher evaluation), many school environments continue to be disrupted by student emotional and behavioral difficulties (e.g., bullying, aggression, social withdrawal, defiance; Walker, Ramsey, & Gresham, 2004). Indeed, educators continually rank emotional and behavioral problems among their top classroom concerns (Buscemi, Bennett, Thomas, & Deluca, 1996; Bushaw & Lopez, 2010; Langdon & Vesper, 2000). Students who exhibit these problems not only miss out on valuable instructional time themselves, but their behaviors can also interfere with teacher delivery of instructional content and inhibit classmate learning (Hinshaw, 1992; Walker et al., 2004).

Although schools are primarily charged with providing instruction to facilitate the academic achievement of students, there is growing recognition among those involved in education that student social and emotional well-being is instrumental to academic success (Schonert-Reichl, 2017). Owing to this, there is consensus among many researchers, policymakers, and practitioners that social–emotional learning (SEL) programs should be adopted and integrated with academic practices to promote school success (Brackett & Rivers, 2014). Advocates of SEL have pushed for greater balance between academic learning and social–emotional education to develop self-sufficient individuals who are adequately prepared for work and life (National Research Council, 2012). Consistent with this push, thousands of schools nationwide are adopting and implementing SEL programs to promote both academic and social–emotional out-

Editor's Note. Editor for this manuscript was Shane R. Jimerson.

Clayton R. Cook, Department of Educational Psychology, University of Minnesota; Sabina Low, T Denny Sanford School of Social and Family Dynamics, Arizona State University; Joanne Buntain-Ricklefs, Kelly Whitaker, and Michael D. Pullmann, Psychiatry and Behavioral Health Department, University of Washington; Jaclyn Lally, College of Education, University of Washington.

Correspondence concerning this article should be addressed to Clayton R. Cook, Department of Educational Psychology, University of Minnesota, Twin Cities, MN. E-mail: crcook@umn.edu

comes of children (Collaborative for Academic and Social Emotional Learning, 2014).

In general, SEL is a curricular approach to school-based universal prevention that consists of teaching students core social–emotional competencies related to identifying and regulating their emotions, setting and working to achieve positive goals, demonstrating empathy and understanding the perspectives of others, cultivating and sustaining positive relationships, making socially responsible decisions, and handling interpersonal conflicts constructively (Zins, Bloodworth, Weissberg, & Walberg, 2007). A recent meta-analysis of 213 studies examining the impact of SEL indicated that SEL curricula are not only associated with significant improvements in students' social–emotional skills, but they were associated with an average 11 percentile increase on academic achievement (Durlak, Weissberg, Dymnicki, Taylor, & Schellinger, 2011). Furthermore, research has shown that early social behavior strongly predicts academic achievement up to five years later, even after controlling for early academic achievement (Caprara, Barbaranelli, Pastorelli, Bandura, & Zimbardo, 2000; Malecki & Elliot, 2002). Although evidence supports the adoption and use of an SEL curriculum, not all programs are equally effective, and each curriculum must stand on its own empirical support.

Second Step Elementary

Given the recognized benefits of teaching students social–emotional skills, several SEL programs have been developed and adopted in the school setting. One of the widely implemented SEL programs for elementary students is *Second Step*, developed by Committee for Children ([CfC] 2016), which is a nonprofit organization based in Seattle. *Second Step* is a skills-focused SEL curriculum that emphasizes directly teaching student's skills that strengthen their ability to learn, demonstrate empathy and compassion for others, manage their negative emotions, and solve interpersonal problems. The *Second Step* logic model (see Figure 1) suggests that students who are provided direct SEL instruction will acquire social–emotional skills, opportunities to practice those skills, and receive reinforcement for exhibiting those skills. The *Second Step* theory of change also suggests that students are likely to experience a range of improved intermediate outcomes, which would result in a cascade of positive distal outcomes (CfC, 2016).

Previous experimental studies have found mixed support for the earlier versions of the *Second Step* program to produce positive child outcomes, which is consistent with other smaller or less rigorous studies that too have resulted in mixed findings. For example, Grossman et al. (1997) conducted a randomized control trial and found that physical aggression decreased among students in the *Second Step* classrooms, when compared to students in the control classrooms, and these positive findings were maintained at a 6-month follow-up. Other studies have demonstrated that students receiving *Second Step* lessons demonstrated increased social skills at posttest when compared with children in control classrooms (Holsen, Iversen, & Smith, 2009; Holsen, Smith, & Frey, 2008). Despite these positive findings, a recent school randomized trial (n = 12 schools) by Nebbergall (2009; 3rd Edition *Second Step*) found no positive or negative effects of *Second Step* on school achievement or positive behaviors. In the case of this study the control schools were, on average, found to be implementing a fairly high level of SEL programming, albeit not as formalized as the intervention schools, making it challenging to differentiate programming between intervention and control schools. These findings, nonetheless, introduce mixed findings regarding the efficacy of *Second Step*.

The most recent investigation of *Second Step* involved a large-scale randomized control trial (61 schools) examining the impact of the new 4th Edition of the program on social-behavioral outcomes over a 1-year period (Low, Cook, Smolkowski, & Buntain-Ricklefs, 2015). Hierarchical models revealed the program had positive main effects on teacher-reported social and behavioral indices, with effect sizes in the small range. The majority of significant findings were moderated effects, with eight out of 11 outcome variables indicating the intervention produced significant improvements in social–emotional competence and behavior for children who started the school year with skill deficits relative to their peers.

SEL and Early Elementary Academic Outcomes

Despite the empirical support for *Second Step* and SEL programming more broadly, there is less known about the impact of

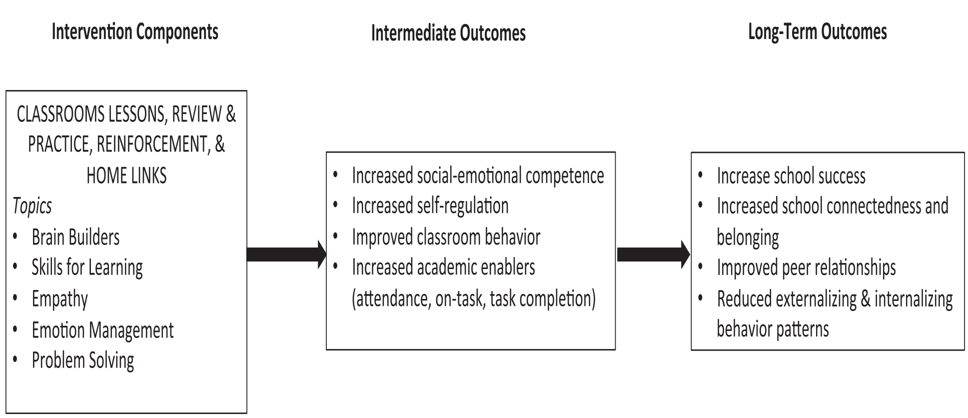

Figure 1. Logic model for the *Second Step*® program.

SEL programming on children's academic outcomes during specific developmental periods, particularly the early elementary school years (Rhoades, Warren, Domitrovich, & Greenberg, 2011). SEL programming for young children in the early elementary grades is particularly important considering the research showing that students who get off to a good start academically are significantly more likely to be successful in the later grades and beyond (Snow, Burns, & Griffin, 1999). SEL provides the basis for teaching young children self-regulatory skills that serve as enablers to lifelong learning (Heckman & Masterov, 2004). Moreover, the prevention of social and behavioral difficulties is more effective than later efforts to remediate more intensive problems (Greenberg et al., 2003). Early elementary represents an opportune developmental period to proactively support children to begin developing the social–emotional competence to prevent social and behavioral problems that interfere with learning, as well as enable them to profit from their early learning experiences.

Further, most SEL research supports the primary effects of SEL programs (i.e., positive impact of social and emotional skills and decreased problem behaviors). Although research has supported the impact of SEL programming, in general on students' academic outcomes (Durlak et al., 2011), no research has established a relationship between *Second Step* and academic achievement. Notwithstanding the prior research linking SEL programming to improved outcomes, there remain gaps in the literature and further validation of theory is needed. First, few large scale studies have been performed examining whether SEL programming results in academic outcomes for early elementary students. Second, little research on this developmental period has explicitly examined aspects of fidelity of implementation as it relates to SEL effects on academic outcomes. Third, few research studies examining SEL programs have utilized multilevel models that take into account school-, classroom-, and individual-level effects.

Purpose of the Current Study

Using data from the rigorous, large-scale randomized control trial of the fourth edition of *Second Step* discussed earlier, the primary purpose of this study was to evaluate the impact of *Second Step* on early elementary students' academic outcomes. A secondary aim was to examine classroom-level moderators of treatment effectiveness in order to better understand the conditions under which programs like *Second Step* may produce effects on academic outcomes (Flay et al., 2005). Specifically, we examined the influence of different aspects of fidelity of implementation on academic outcomes, which have been shown to influence program outcomes (Proctor & Brownson, 2012). Student outcome data included direct observations of students' on-task and disruptive behaviors in the classroom and curriculum-based measurement probes of academic performance (i.e., reading and mathematics).

Related to the secondary aim, numerous studies have underscored the importance of assessing fidelity of implementation during efficacy and effectiveness studies based on findings indicating that dimensions of fidelity (adherence, dosage, and competency/engagement) account for differential outcomes (Durlak & Dupre, 2008). Different dimensions of fidelity were incorporated into this investigation, with adherence (i.e., implementing core components as planned) and dosage (i.e., number of lessons) capturing traditional dimensions of fidelity and engagement (i.e., students' level of engagement) reflecting the quality or competency with which *Second Step* was delivered (Perepletchikova & Kazdin, 2005). Although it is well understood that fidelity affects the outcomes obtained in prevention programs (Durlak & Dupree, 2008), significantly less is known about the factors that influence quality implementation of school-based programs delivered by teachers (Owens et al., 2014). Thus, this study examined different dimensions of fidelity (adherence, engagement, dosage) and their relationship to changes in student academic-related outcomes.

Hypotheses. The specific hypotheses that guided this study were informed by prior literature on SEL programs regarding their collateral impact on short-term academic outcomes (Durlak et al., 2011). Drawing upon Figure 1, it was hypothesized that early elementary students who participated in *Second Step* would demonstrate relatively small improvements in (a) reading and math and (b) classroom behavior (on-task behavior and disruptive behavior). This hypothesis was based on prior research demonstrating smaller effects of universal prevention programs (Neil & Christensen, 2009), particularly studies examining impact on more distal outcomes.

We hypothesized that fidelity of implementation would moderate the impact of *Second Step* on academic outcomes. Specifically, we hypothesized that stronger fidelity in delivering *Second Step* as measured by a composite of different dimensions of fidelity (e.g., dosage, adherence, engagement) that could differentially be related to better academic-related outcomes. Moreover, we postulated that we would find a specific moderated effect with regard to lesson engagement as a proxy of teacher competency in delivering *Second Step*, considering the literature linking practitioner competency to fidelity of implementation (Perepletchikova & Kazdin, 2005; Sanetti & Kratochwill, 2009).

Method

Participants

This study included students in kindergarten through second grade enrolled in five school districts across the Puget Sound area of Washington and in one district in Mesa, Arizona. School districts ranged from rural to urban settings and were recruited in spring 2012 after approval from the institutional review boards (IRBs). School districts, teachers, students, and parents of the students consented to participate in accordance with IRB procedures.

Recruitment and retention. The Washington site was able to secure and maintain the participation of 41 schools across five school systems. On the basis of power analyses to secure the participation of a sufficient number of classrooms and students, on average, six randomly selected classrooms per school participated in data collection, though all classrooms in the intervention schools were provided the intervention. A total of 224 teachers agreed to participate and passive parental permission was obtained for 4,891 students, only 1.4% of parents declined. The Arizona site was able to secure and maintain participation from 20 schools from the Mesa School District. An average of five classrooms per school (minimum = 3; maximum = 7) participated in data collection, with a total of 97 teachers. Passive parental permission was obtained for 2,879 students, only 1% of parents declined. Across both Washington and Arizona sites, a total of five schools (8% of all recruited school) and 15 teachers (5% of all teachers) opted out of participation in this study.

Student- and teacher-level demographics and descriptive information are displayed in Tables 1 and 2, with statistical tests (*t* tests and crosstabulations with chi-square tests) comparing teachers in the *Second Step* condition with teachers in the control condition. The total child sample was 7,419, with 3,727 students in *Second Step* and 3,692 students in the control condition. There were more students in the *Second Step* condition who were in Kindergarten, and fewer who were in 1st grade. With regard to socioeconomic status, 50% and 78% of participating students in Washington and Arizona, respectively, received free and reduced lunch. Although there were some significant differences in the racial and ethnic breakdown of the students in the two sites (see Table 1), the total sample was relatively representative of the ethnicity (nationally 64% non-White) and socioeconomic (nationally 48% of students receive free and reduced lunch) distribution of school-age children in the United States (U.S. Census Bureau, 2011).

The total teacher sample (see Table 2) was 310, with 59 *Second Step* teachers and 151 control teachers. Teacher demographics were comparable across conditions, with some significant differences in race/ethnicity.

Procedures and Design

Overview. The study used a large-scale, matched, randomized-controlled design with 61 elementary schools randomly assigned within their district to either the early start (treatment; $n = 31$) or delayed start (control; $n = 30$) conditions (see Low, Cook, Smolkowski, & Buntain-Ricklefs, 2015 for CONSORT diagram). The delayed start condition did not receive *Second Step* during the time period of this study. Schools within Washington and Arizona were matched on free and reduced lunch and percent of non-White students for design purposes (Murray, 1998). There were no significant differences between treatment and control groups on baseline outcome measures (see results section). The present study includes data from the fall (T1) and spring (T2) assessments gathered in Year 1. The overall study represents an evaluation of the impact of implementing two consecutive years of *Second Step*.

Table 1
Child-Level Sample Descriptive Information at Fall Quarter

Variable	Control *n* (%)	Second Step *n* (%)	Total sample *n* (%)
Total students	3,692	3,727	7,419
Grade			
Kindergarten	1482 (40.1)	1653 (44.4)	3135 (42.3)
1	1991 (53.9)	1863 (50.0)	3854 (51.9)
2	219 (5.9)	211 (5.7)	430 (5.8)
Sex			
Male	1772 (48.0)	1788 (48.0)	3,560 (48.0)
Female	1657 (44.9)	1704 (45.7)	3,361 (45.3)
Missing	263 (7.1)	235 (6.3)	498 (6.7)
Race			
Asian	368 (10.0)	333 (8.9)	701(9.4)
Native Hawaiian or other Asian/Pacific Islander	25 (.7)	46 (1.2)	71 (1.0)
Black or African American	232 (6.3)	212 (5.7)	444 (6.0)
American Indian or Alaska Native	86 (2.3)	123 (3.3)	209 (2.8)
Caucasian/White, non-Hispanic	1137 (30.8)	1542 (41.4)	2679 (36.1)
More than one race	196 (5.3)	183 (4.9)	379 (5.1)
Hispanic	908 (24.6)	761 (20.4)	1669 (22.5)
Missing	740 (20.8)	527 (14.1)	1267 (17.1)
Student special education status			
Not in special education	2418 (65.5)	2524 (67.7)	4942 (66.6)
Special education	321 (8.7)	309 (8.3)	630 (8.5)
Missing	953 (25.8)	894 (24.0)	1847 (24.9)
Student English language learner (ELL) status			
Not an ELL	2075 (56.2)	2221 (59.6)	4296 (57.9)
ELL student	829 (22.5)	716 (19.2)	1545 (20.8)
Missing	788 (21.3)	790 (21.2)	1578 (21.3)
Age	6.2 (.7)	6.2 (.8)	6.2 (.8)
Number of school days missed	9.0 (7.7)	9.2 (7.7)	9.2 (7.8)
Fall percentage intervals on-task behavior	83.3 (20.1)	81.9 (20.6)	82.6 (20.4)
Fall percent intervals disruptive behavior	8.8 (14.3)	9.5 (15.4)	9.1 (14.9)
Fall oral reading fluency words reading correct per minute	24.2 (33.4)	22.8 (33.8)	23.63 (33.6)
Fall math percent correct	28.2 (27.2)	26.9 (27.7)	27.5 (27.5)
Spring percent intervals on-task behavior	80.1 (22.5)	79.7 (22.6)	79.9 (22.5)
Spring percentage intervals disruptive behavior	9.6 (16.5)	8.6 (14.7)	9.1 (15.6)
Spring oral reading fluency words reading correct per minute	48.5 (42.3)	48.8 (45.2)	48.7 (43.7)
Spring math percent correct	55.9 (32.2)	54.3 (33.4)	55.1 (32.8)

Note. For the following variables, *t* test or chi-square $p < .05$: grade, race, student English language learner status, Fall percentage intervals on-task behavior, Spring percentage intervals disruptive behavior.

Table 2
Fall Quarter Teacher-Level Sample Descriptive Information

Variable	Control n (%)	Early start n (%)	Total sample n (%)
Total teachers	151	159	310
Site			
ASU	48 (31.8)	48 (30.2)	96 (31.0)
UW	103 (68.2)	111 (69.8)	214 (69.0)
Sex			
Male	9 (6.0)	3 (1.9)	12 (3.9)
Female	142 (94.0)	156 (98.1)	298 (96.1)
Hispanic or Latino/a?			
No	142 (94.0)	149 (94.3)	291 (93.9)
Yes	9 (6.0)	9 (5.7)	18 (5.8)
Missing	0	1 (.7)	1 (.3)
Race			
Asian	6 (4.0)	3 (1.9)	9 (2.9)
Native Hawaiian or other Asian/Pacific Islander.	0	3 (1.9)	3 (1.0)
Black or African American	0	2 (1.3)	2 (.6)
American Indian or Alaska Native	1 (.7)	1 (.6)	2 (.6)
Caucasian/White	128 (84.8)	143 (92.3)	271 (87.4)
More than one race (please specify):	10 (6.6)	3 (1.9)	13 (4.2)
Other	6 (4.0)	0	6 (1.9)
Missing	0	4 (2.6)	4 (1.3)
Highest degree received			
Bachelor,s degree	48 (33.8)	64 (42.1)	115 (37.1)
Master's degree	87 (61.3)	85 (55.9)	185 (59.7)
Professional degree	6 (4.2)	3 (2.0)	9 (2.9)
Doctorate degree	1 (.7)	0	1 (.3)
Grade(s) taught			
Kindergarten	61 (40.4)	70 (44.0)	131 (42.3)
Kindergarten–first grade split	4 (2.6)	1 (.6)	5 (1.6)
First grade	75 (49.7)	79 (49.7)	154 (49.7)
First grade–second grade split	4 (2.8)	2 (1.3)	6 (1.9)
Second grade	7 (4.6)	7 (4.4)	14 (4.5)
Age	42.9 (11.9)	44.3 (12.8)	43.67 (12.38)
Missing	2	5	7
Number of years teaching	14.4 (9.4)	15.9 (10.5)	15.19 (10.04)

Note. For race, t test or chi-square $p < .05$. ASU = Arizona State University; UW = University of Washington.

Training participation. The *Second Step* curriculum (1 hr) and Proactive Classroom Management (PCM; 3 hr) were provided to participating early start schools: *Second Step* was consistent with standard support operations provided by Committee for Children and intended to increase motivation to implement the program, allow teachers to become familiar with the content, and provide specific examples of how to deliver program with fidelity. All early start schools participated in the training, and all kindergarten, first, and second grade teachers involved in data collection participated in the webinar, which we determined by attendance sheets collected by school personnel.

The PCM trainings are not standard practice in *Second Step* implementation, but were a response to district needs at the time of recruitment. A very brief overview of classroom strategies was presented to meet the needs of schools without providing a sufficiently strong dosage that one would anticipate having a strong impact on classroom behaviors (consistent with train and hope, Stokes & Baer, 1977). Specifically, PCM strategies were delivered either via DVD or in-person depending on the preference of the school, and focused on practices that would help support, reinforce, and facilitate the student active engagement in learning, including *Second Step* lessons, such as positive greetings at the door, attention signals, 5 to 1 ratio, and behavior specific praise (Simonsen, Fairbanks, Briesch, Myers, & Sugai, 2008). As part of the training, a focus was on connecting the use of the PCM strategies to promoting students' use of the social–emotional skills being taught to them. For greater description of the PCM training and the method utilized to support the implementation of *Second Step* see Low et al. (2015).

Compensation. Participating schools, teachers, and school liaisons were given a financial stipend for their involvement in the study. Early start schools were provided the curricula at no-cost, and delayed schools were scheduled to receive the free curricula at the end of data collection.

Measures

Data were collected at three time points during the academic year in the fall, winter and spring. However, for the purposes of this study only two data collection periods—fall and spring—were included in the analyses as the winter data collection period involved gathering only limited subset of data (e.g., direct observation of classroom behavior). Data collection was not blind, as trained graduate students were knowledgeable of the condition of each of the participating schools.

School demographic and archival data. We collected school-level data from publically available online sources (e.g., National Center for Education Statistics website, school district websites) on the type of school (e.g., public vs. private), number of students, racial/ethnic composition of students, and percentage of students receiving free or reduced-price lunch. These data were used as covariates in the multilevel models.

Behavioral observation. To record class-wide and individual student behavior, a behavioral observation system was developed based on the Behavioral Observation of Students in Schools (BOSS; Shapiro & Kratochwill, 2000). The BOSS has been shown to produce acceptable interobserver agreement (IOA) and concurrent and predictive validity with other measures (Volpe, DiPerna, Hintze, & Shapiro, 2005). The three behavioral coding categories consisted of on-task behavior, off-task behavior, and disruptive behavior. The present manuscript focused on two aspects of classroom behavior that are most closely tied to social–emotional competence, on-task behavior, defined as behaviors that were consistent with the current learning task or instructional directive (e.g., listening to instruction, talking to peers about academic topic, reading, writing, raising hand, etc.) and disruptive behavior, defined as behaviors not pertinent to the assigned activity/task that negatively impact the learning environment (e.g., blurting, leaving one's seat, distracting peers, and making noises with objects). Off-task was not included because it represents the inverse of on-task behavior.

Observations were conducted in all classrooms (early and delayed start) across both sites by trained graduate students during core academic instruction time in the fall, winter and spring, but only fall and spring data were used in this study. Each student was observed for 2 min total, divided into 10-s intervals. To obtain class-wide estimates of on-task behavior, observers were instructed to begin with an identified student in the front or back of the classroom and systematically move to the next student to the left after each interval. After the observers made their way through all students in the class, they repeated the same process until the observation time elapsed. A minimum of 12 intervals of data per student and roughly 300 total intervals per class per data collected period were obtained. Identifiable information were used as part of the observation to link data back to individual students. This observation system allowed for the calculation of class-wide and individual student estimates.

Prior to conducting the observations, graduate students were trained on the observation system. Before beginning baseline data collection, each student was required to reach at least 90% agreement during practice trials with an identified observer who served as the anchor measure. IOA data consisting of two observers conducting the observation at the same time on the same students were collected on roughly 20% of the observation sessions. IOA was calculated using the point-by-point method, which consists of calculating agreement for each and every interval. This method has been shown to be a more accurate estimate of the agreement between raters for direct observation systems with interval recording formats (Shapiro & Kratochwill, 2000). The results revealed that IOA averaged 88% (minimum = 72% and maximum = 100%), which was associated with a Kappa value of .71 and is considered to be an acceptable level of interrater reliability (Bailey & Burch, 2002; Viera & Garrett, 2005).

Curriculum-based measurement (CBM). To assess potential growth in academic performance as a result of *Second Step*, commercially available CBM probes from *Aimsweb* were administered: (a) oral reading fluency (words read correct per minute) and (b) math calculation (number of digits correct in a minute and percent correct). R-CBM or oral reading fluency probes represent a standardized, general outcome measure of reading performance that is highly sensitive to students' response to instruction (Fuchs & Fuchs, 1999). M-CBM or math computation probes have been shown to be a reliable and valid general outcome measure of overall mathematics computation (Thurber, Shinn, & Smolkowski, 2002). Students received grade-equivalent probes for both of these measures.

The R-CBM is a 1-min timed reading of a short passage, which measures oral reading fluency (ORF). The Reading CBM was administered one-to-one with the student. The trained graduate research assistants administered the R-CBM using a stopwatch, a clipboard, a wipe cloth, grade-leveled passage in a plastic sleeve, a dry erase pen, and a laptop or a recording sheet to record the student's words read correctly (WRC) per minute. The M-CBM assessed math computational skills (M-COMP). It was an 8-min test that was group administered. Graduate students used a timer and provided copies of M-COMP testing sheet for the entire class. The graduate research assistants followed the standardized administration directions for R-CBM and M-COMP from the *Aimsweb* website.

Fidelity of implementation. Teachers were asked to complete weekly implementation logs to record adherence to the program, as well as adaptations and student engagement. Adherence had two components: adherence to the key lesson components (five items; yes/no) and adaptations/modifications (four items on 4-point Likert scale; e.g., "To what extent did you leave out parts of the lesson"). Engagement had two components: ratings of the degree of student engagement (three items on a 4-point Likert scale; e.g., "To what extent were students following along with the lesson") and estimated percentage of students who were engaged in the lesson (0% to 100%). For purposes of comprehension, it is important to know that the fidelity engagement variable was used to capture students' participation in and responsiveness to the lessons, whereas the on-task behavior dependent variable was used to capture the degree to which students were behaviorally engaged during core instructional time. Teachers were also asked to keep a log on how many lessons they completed by the end of the year and reported this information to a school liaison as an indicator of dosage. We modeled the measure after recommendations from Sanetti and Kratochwill (2011) for developing a reliable and valid measure of fidelity of implementation which included operationalizing the core components and providing for repeated ongoing assessment of fidelity. Previous research has demonstrated that the fidelity measure is associated with significant differential effects, lending support for the validity of the self-report measure (Low et al., 2015, 2016).

Analyses

Multilevel models (MLM) were run as three-level linear models (school, teacher, and student) with robust standard errors and full information maximum likelihood, predicting each of the dependent variables with separate models, using the HLM7 program (Raudenbush, Bryk, Cheong, Congdon, & du Toit, 2011). Because schools were the level of randomization, condition was entered as a level-3 (school level) variable and predicted scores and standard errors are adjusted for nesting at the school and teacher levels. We used four major dependent variables: reading (R-CBM; words correct per minute); math (M-CBM; percentile correct); academic time engaged (BOSS-OT or on task behavior); and disruptive behavior (BOSS-DB). BOSS off-task behavior was not included because it was highly negatively correlated with BOSS-OT ($r > .9$). Because grade level was highly related to all dependent variables (DVs), we transformed all DVs into z scores by grade and semester. The fall score for each DV were entered as a covariate in each model to control for baseline differences. The four mixed models testing main effects were as follows:

$$
\begin{aligned}
Z\_DV.2_{ijk} = {} & \gamma_{000} + \gamma_{001} * \mathrm{CONDITION}_k + \gamma_{100} * Z\_DV.1_{ijk} \\
& + \gamma_{101} * Z\_DV.1_{ijk} * \mathrm{CONDITION}_k + r_{0jk} + r_{1jk} * Z\_DV.1_{ijk} + u_{00k} \\
& + u_{10k} * Z\_DV.1_{ijk} + e_{ijk}
\end{aligned}
$$

Where Z_DV = Grade-adjusted z score for each of the DVs (reading words per min, math percentage correct; BOSS Academic Time Engaged, BOSS Disruptive Behavior); CONDITION = Dummy code of Early Start or comparison, with Early Start = 1; .1 and .2 = Fall and spring scores, respectively; γ_{OOO} = intercept; e_{ijk} = level-1 variance/error; r_{Ojk} = level-2 variance for intercept; r_{1jk} = level-2 variance for spring score; $u_{O.k}$ = random variance. It should be noted that the $Z\_DV.1_{ijk}$*CONDITION$_k$ term was included as a covariate to control for baseline scores.

Additional moderator analyses limited to the Early Start group examined possible relationships between intervention fidelity and academic outcomes. The three fidelity subscales (adherence, generalization, and engagement) each had a different possible range. Therefore, we created z scores and created a total score by taking the average of all three subscales. Ten teachers (6.3%) were missing some or all fidelity data. Data appeared to be missing at random because there were no significant differences between those with complete data and those with incomplete data on demographics, grade taught, classroom academic scores at fall and spring, and scores on the My Class Inventory at fall and spring. To preserve statistical power, we used multiple imputation (MI) in SPSS version 21 to create five data sets with estimated scores replacing the missing data for these teachers considering our secondary aim of examining the relationship between different dimensions of fidelity on student academic-related outcomes. MI models were based on linear regression with all predictive existing data, constrained at the minimum and maximum of valid data (Graham, 2009; Raghunathan, Reiter, & Rubin, 2003). MI analyses pooled results from all five models during the creation of multilevel models similar to those above, but including each of the three fidelity subscales as covariates. All predictor variables were included in each MLM. Each model included all effects as randomly varying in an initial run. To maximize statistical power and model parsimony, we eliminated random variance terms that did not contribute to model fit. This was done by iteratively building models, fixing variance component with the highest significance value $p > .10$, and running the model again, until all remaining variance components were significant at $p < .10$ and models were considered final. Our final four moderator models were:

$$
\begin{aligned}
Z\_READ.2_{ijk} = {} & \gamma_{000} + \gamma_{010} * \mathrm{BSECSTEP}_{jk} + \gamma_{020} * \mathrm{ADHERE}_{jk} + \gamma_{030} * \mathrm{ENGAGE}_{kk} \\
& + \gamma_{040} * \mathrm{GENERAL}_{jk} + \gamma_{100} * ZORFPM.1_{ijk} + r_{ojk} \\
& + r_{1jk} * Z\_READ.1_{ijk} + u_{00k} + u_{01k} * \mathrm{BSECSTEP}_{jk} \\
& + u_{04k} * \mathrm{GENERAL}_{jk} + e_{ijk}
\end{aligned} \tag{1}
$$

$$
\begin{aligned}
Z\_MATH.2_{ijk} = {} & \gamma_{000} + \gamma_{010} * \mathrm{BSECSTEP}_{jk} + \gamma_{020} * \mathrm{ADHERE}_{jk} \\
& + \gamma_{030} * \mathrm{ENGAGE}_{jk} + \gamma_{040} * \mathrm{GENERAL}_{jk} + \gamma_{100} * Z\_MATH.1_{ijk} \\
& + r_{ojk} + r_{1jk} * Z\_MATH.1_{ijk} + u_{02k} * \mathrm{ADHERE}_{jk} \\
& + u_{10k} * ZMTHPR.1_{jk} + e_{ijk}
\end{aligned} \tag{2}
$$

$$
\begin{aligned}
Z\_AET.2_{ijk} = {} & \gamma_{000} + \gamma_{010} * \mathrm{BSECSTEP}_{jk} + \gamma_{020} * \mathrm{ADHERE}_{jk} + \gamma_{030} * \mathrm{ENGAGE}_{jk} \\
& + \gamma_{040} * \mathrm{GENERAL}_{jk} + \gamma_{100} * Z\_AET.1_{ijk} + r_{ojk} \\
& + u_{01k} * \mathrm{BSECSTEP}_{jk} + e_{ijk}
\end{aligned} \tag{3}
$$

$$
\begin{aligned}
Z\_DB.2_{ijk} = {} & \gamma_{000} + \gamma_{010} * \mathrm{BSECSTEP}_{jk} + \gamma_{020} * \mathrm{ADHERE}_{jk} + \gamma_{030} * \mathrm{ENGAGE}_{jk} \\
& + \gamma_{040} * \mathrm{GENERAL}_{jk} + \gamma_{100} * ZDB.1_{ijk} + r_{0jk} + r_{1jk} * Z\_DB.1_{ijk} \\
& + u_{01k} * \mathrm{BSECSTEP}_{jk} + e_{ijk}
\end{aligned} \tag{4}
$$

Where Z_READ = Grade-adjusted z score for reading words per minute; Z_MATH = Grade-adjusted z score for math percent correct; Z_AET = Grade-adjusted BOSS Academic Time Engaged z score; Z_DB = Grade-adjusted BOSS Disruptive Behavior z score; BSECSTEP = number of sessions of Second Step delivered (grand-mean centered); ADHERE = Adherence score; ENGAGE = Engagement score; GENERAL = Generalization score; .1 and .2 = Fall and spring scores, respectively; γ_{OOO} = intercept; e_{ijk} = Level 1 variance/error; r_{Ojk} = Level 2 variance for intercept; r_{1jk} = Level 2 variance for spring score; $u_{O.k}$ = random variance.

Results

Participant Descriptives, Mobility, and Missing Data

There were 714 students (9.6%) absent during fall data collection. By the spring, 635 students (8.1%) moved out of the district, to a different school, or to a different classroom and left the study; 22 (.3%) moved to a different classroom or a different school in the same condition and remained in the study; 38 (.5%) left the study for unknown reasons; 67 (.9%) left the study because the teacher declined; and 626 (8.4%) were absent during spring data collection. Student attrition was unrelated to condition ($\chi^2 = .38$, $p = .54$), gender ($\chi^2 = .31$, $p = .65$), and gender ($\chi^2 = .77$, $p = .31$). Data was missing on at least one of the four dependent variables for 10.5% to 11.1% of the sample in the fall and 16.1% to 18.0% in the spring. Although this may seem high, relative to other school-based studies conducted with ethnically and socioeconomically diverse students, the percentage of missing data is relatively low (e.g., Gillham et al., 2007). Previous research has shown that the mean rate is 25% (Biglan et al., 1991).

Descriptives for untransformed spring outcome scores are displayed in Table 1. Independent t tests found that the control group had more disruptive behavior, $t(5952.34) = 2.469$, $p = .014$; no differences were found for on-task behavior, reading, or math scores.

Fidelity of Implementation

Of all indicators of implementation, dosage varied the most within schools. The average number of lessons completed across sites was 17.42 ($SD = 3.72$, range = 7 to 25). The average number of lessons completed across sites was 17.42 ($SD = 3.72$, range = 7 to 25). The school-level (unconditional) intraclass correlation (ICC) for two-level models (teachers nested within schools), calculated as the Level 2 variance divided by the total model variance, was .32, indicating that 32% of the variance in the number of sessions delivered was accounted for at the school level; hence, the number of lessons completed was more similar among classrooms in each individual school than across all schools. Most teachers made only a few adaptations, but a few made a number of modifications ($M = 1.92$, $SD = 1.28$, range = 0.00 to 6.55, ICC = .02). Adherence scores ranged from .7 to 2.0 and averaged 1.77 ($SD = .25$); Engagement ranged from −3.5 to 19.4 and averaged 5.9 ($SD = 1.2$); Generalization ranged from −15.8 to 51.7 and averaged 8.7 ($SD = 3.3$).

Main Effect Analysis

Table 3 displays the most salient statistics from the four main effect models, including the coefficients for the school-level z scores at spring semester and the fixed effect adjustment related to the early start condition. None of the basic models revealed a significant relationship between delivery of the *Second Step* program and academic-related outcomes. For instance, to interpret Table 3, the average control school words correct per minute z score in spring semester did not significantly differ from zero (coefficient = −.012, $p = .705$), and the *Second Step* group had nonsignificant scores that were .065 standard deviations higher than the control group (coefficient = .065, $p = .161$). The other statistics in Table 3 are interpreted in a similar manner.

Moderator Analyses

Table 4 shows the results from four multilevel models of school-level z-scores at spring semester, controlling for fall semester z score, predicted by all three fidelity subscales and the total number of lessons delivered. Degrees of freedom vary widely due to allowing terms to vary randomly (which resulted in smaller degrees of freedom) or fixing variables.

Table 3
MLM Fixed Effects From Four Basic Three-Level Models of the Relationship Between Second Step and Spring Semester School-Level Z-Scored Academic Outcomes

Fixed effects	Coefficient	*SE*	*t*	*df*	*p*
Words correct per minute	–.012	.032	–.381	59	.705
Second Step	.065	.046	1.419	59	.161
Math percent correct	–.013	.037	–.358	59	.722
Second Step	.040	.049	.825	59	.413
Academic time engaged	.015	.051	.287	59	.775
Second Step	–.019	.071	–.276	59	.784
Disruptive behavior	.025	.046	.544	59	.588
Second Step	–.062	.056	–1.097	59	.277

Note. Each model also included a covariate slope controlling for student's Fall scores. MLM = multilevel models.

There was a small but significant relationship between spring reading scores and engagement such that one standard deviation higher engagement z score was associated with a spring reading score that was .126 standard deviation higher ($p = .008$); spring reading scores were also highly correlated with fall reading score (coefficient = .858, $p < .001$). No other fidelity subscale or number of lessons delivered significantly predicted spring reading score. There was a significant relationship between spring math score and the number of lessons delivered, such that each additional lesson was related to math scores that were .016 standard deviations higher ($p = .048$); spring math score was also correlated with fall math score (coefficient = .607, $p < .001$), but spring math score was not significantly related to any fidelity subscale. Spring academic time engaged (on-task behavior) was significantly correlated with the number of lessons delivered, such that each additional lesson was related to a .023 standard deviations higher academic time engaged score ($p = .026$); spring academic time engaged was also correlated with fall academic time engaged (coefficient = .111, $p < .001$), but was not significantly related to any other fidelity subscale. Spring disruptive behavior was significantly moderated by the engagement subscale such that a one standard deviation higher engagement z score was related to a −.132 standard deviation lower disruptive behavior z score ($p = .046$); spring disruptive was also correlated with fall disruptive behavior (coefficient = .107, $p < .001$). There was some evidence, though it was not statistically significant, that the number of lessons may have been related to a lower disruptive behavior score (coefficient = −.016, $p = .063$). Spring disruptive behavior was not related to any other fidelity subscale.

Discussion

There is a need for continued research examining the impact SEL programs, like *Second Step*, as universal supports that promote better academic outcomes for early elementary students who are in a critical developmental period of acquiring core academic skills that will facilitate later learning. Such research is important considering that students who begin the elementary school years with academic deficits are significantly at-risk for additional academic difficulties and longer-term negative outcomes (Morgan, Farkas, & Wu, 2011; Stanovich, 1986). In light of this as well as the absence of research examining the academic outcomes associated with *Second Step Elementary*, the present study used data from a large-scale randomized controlled trial to examine its impact on the academic outcomes of early elementary students.

Results from this study suggest that *Second Step* was not associated with significant main effects on the academic outcomes measured in this study. Specifically, intervention and control classrooms did not significantly differ on CBM probes of oral reading fluency and mathematics computation and classroom on-task behavior. Although contradictory to our hypotheses, interpretation of these findings are understandable considering that the randomized controlled trial represented several features more characteristic of an effectiveness study (e.g., training delivered in a feasible manner and limited researcher implemented corrections for fidelity problems) than a strict efficacy study with high control over fidelity by researchers. As with all intervention research, interpretations of effects or the lack thereof require some indication that the intervention was implemented as it was designed (Kazdin, 2003).

Table 4
MLM Fixed Effects of Fidelity to Intervention on Spring Semester (SS) School-Level Z-Scored Academic Outcomes

Fixed effects	Coefficient	*SE*	*t*	*df*	*p*
Words correct per minute *z* score	.041	.031	1.315	30	.198
Number of SS lessons	–.002	.005	–.366	30	.717
Adherence *z* score	–.012	.018	–.634	63	.528
Engagement *z* score	.126	.045	2.791	43	.008
Generalization *z* score	–.028	.041	–.682	30	.501
Fall words correct per minute score (slope)	.858	.024	35.927	63	<.001
Math percent correct *z* score	.024	.033	.732	93	.466
Number of SS lessons *z* score	.016	.008	2.006	93	.048
Adherence *z* score	–.038	.039	–.956	30	.347
Engagement *z* score	.050	.059	.851	93	.397
Generalization *z* score	–.062	.041	–1.515	93	.133
Fall math *z* score (slope)	.607	.028	21.411	30	<.001
Academic time engaged *z* score	–.011	.050	–.212	124	.833
Number of SS lessons	.023	.010	2.337	30	.026
Adherence *z* score	–.004	.024	–.155	124	.877
Engagement *z* score	.077	.094	.824	124	.412
Generalization *z* score	.022	.057	.380	124	.704
Fall academic time engaged *z* score (slope)	.111	.017	6.596	2601	<.001
Disruptive behavior *z* score	–.016	.041	–.404	123	.687
Number of SS lessons	–.014	.007	–1.931	30	.063
Adherence *z* score	–.001	.024	–.025	123	.980
Engagement *z* score	–.130	.065	–2.017	123	.046
Generalization *z* score	–.046	.055	–.823	123	.412
Fall disruptive behavior *z* score (slope)	.107	.019	5.619	123	<.001

Note. Number of SS lessons is grand mean centered.

There are several putative explanations for why *Second Step* was not associated with main effects on academic outcomes. First, developmentally SEL programming may have more of a delayed impact on academic performance. Thus, comparing students who have and have not received SEL programming over longer periods of time will be important to examine whether the main effects emerge later on when the instruction transitions from learning core academic skills to applying the learned academic skills in the context of content learning (Vitaro, Brendgen, & Tremblay, 2001). Second, considering the ethnically and socioeconomically diverse sample included in this study, it is possible that *Second Step* leads to academic improvements for some students and not others. Third, *Second Step* and other SEL programs may have more beneficial effects for early elementary students with typical or above academic skills and thus social–emotional skills enable them to better profit from their learning experiences. Whereas students who are low with regard to academic skills (i.e., students who are more likely to attend urban or poorer rural school districts) may improve in social–emotional skills but need supports that target remediating academic deficits in order to demonstrate change on academic outcomes. Indeed, the sample of students included in this study was lower than the average for their same age peers on the measures of oral reading fluency and math computation. Last, there are likely contextual variables of schools and classrooms that mitigate the impact of *Second Step* on academic outcomes. Indeed, several studies examining universal prevention programs have demonstrated moderated effects. Consistent with this, as part of this study we conducted moderator analyses in order to examine the specific contextual factors that moderate the impact of *Second Step* on specific academic outcomes of early elementary students.

The nonsignificant main effects were followed up with specific moderator analyses within the *Second Step* group to examine whether fidelity of implementation differentially impacted student academic outcomes. Although the average fidelity ratings from teachers were relatively high (e.g., average of 80% fro adherence), there was significant variability between teachers, indicating that teachers differed in the quality with which they delivered *Second Step*. Results revealed that two fidelity metrics were related to differential academic effects: lesson engagement and dosage. Lesson engagement was related to significant effects with regard to gains in oral reading fluency and reductions in disruptive classroom behavior. It is important to note that the engagement measure was gathered at the teacher level, which suggests that teachers who rated their students as being more engaged as whole during the *Second Step* lessons were associated with students who performed higher on the reading probes and exhibited fewer disruptive behaviors. Dosage, as measured by the number of lesson delivered, was also found to have a significant moderating on student on-task behavior and math computation. Overall, these findings suggest that relevant academic-related outcomes may be achieved when teachers deliver the lessons in an engaging way and implement more of lessons throughout the year.

When considering these findings in addition to those from the previous study from this randomized controlled trial, *Second Step* has demonstrated significant small effects across multiple methods (teacher report, direct observation, academic probes) and multiple sources (teacher, student, observer) of data. It is important to note, however, most of the significant findings obtained represent moderated effects by either individual- (e.g., baseline status of the student) or classroom-level (e.g., fidelity of implementation) factors. The findings herein suggest that the 4th Edition of the *Second*

Step when combined with a brief PCM training appears to exhibit small collateral effects on certain academic-related outcomes for early elementary students. These findings suggest that SEL programs, such as *Second Step*, need to be implemented with sufficient dosage and in a way that engages students in lessons, as well as potentially combined with other universal supports (e.g., PCM) that seek to better manage behavior and prevent problems that interfere with learning (Domitrovich et al., 2010).

Limitations and Future Directions

Like all studies, this study has limitations that readers should be aware of and that pinpoint directions for future work. First, we focused on classroom-level variables when conducting the moderation analyses. Future studies should examine individual- and school-level moderator variables to better understand with whom and under what conditions *Second Step* differentially impacts academic outcomes. Second, this study focused solely on the year one data, limiting our ability to formally examine and interpret mediation, and mechanisms of change. Third, this study did not investigate specific mechanisms by which social–emotional competence leads to improved academic outcomes. Some researchers have suggested that social–emotional competence leads to greater self-regulation of attention and emotions, which enables students to be more receptive to academic instruction (Rhoades et al., 2011). Fourth, although teacher competency, as measured by student engagement, accounted for significant gains in reading and classroom behavior outcomes, it is difficult to disentangle the effects to determine whether the positive findings are due to general teacher competency or specific competency related to the implementation of the SEL program. Fifth, it was not possible to keep the observers blind to study condition considering resources, magnitude, and duration of this project, which presents is a threat to internal validity of the observational findings. Although keeping the observers blind would have provided greater internal validity, training and follow-up support were provided to observers to maintain their objectivity and prevent drift during the observations. Last, the conclusions reached in this study cannot necessarily extend beyond early elementary populations; further longitudinal work is needed to determine sustainability of these findings over time. Last, given the inclusion of the PCM training with the *Second Step*, it is difficult to disentangle what was responsible for the effects. Follow up studies linking improvements in specific social–emotional competence to improved academic performance will be necessary to isolate the contribution of SEL instruction to academic outcomes.

Implications

Consistent with prior research, the results of this study support the importance of fidelity of implementation as a key ingredient to facilitate the impact of SEL programming. This research suggests that teacher competency in particular, as measured by the degree of student engagement in lesson delivery, may serve as a potent aspect of the fidelity of SEL programming. Researchers have argued that even if an implementer of an intervention is adherent, they can deliver the intervention incompetently, which undermines engagement in the intervention and ultimately the likelihood of achieving desired outcomes (Perepletchikova, Treat, & Kazdin, 2007). Moreover, findings suggest that number of lessons delivered (i.e., dosage) is significantly related to the collateral impact of *Second Step* on academic outcomes. What is unclear from this research is the ideal number of lessons needed to produce relevant improvements in student academic outcomes. Collectively, the results support the need to utilize an implementation science perspective that examines various factors that facilitate the adoption and effective use of SEL curricula with optimal dosage. Implementation enhancement efforts such as high quality coaching, administrative accountability, and coaching support are likely necessary ingredients to ensure not only high quality adherence and adequate dosage but also competent delivery of lessons to ensure student engagement (Powell et al., 2015). Future research should continue to investigate the optimal fidelity profiles that lead to desired student outcomes.

The findings from this study also have implications for the importance of conducting moderator analyses to determine with whom and under what conditions students demonstrate a positive response to a SEL curriculum. A moderator is a variable that alters the direction or strength of the relation between a predictor and an outcome (Baron & Kenny, 1986; Holmbeck, 1997). Thus, a moderator effect represents an interaction whereby one variable depends on the level of another. Studies that examine individual- and contextual-level moderators are critical to developing a sophisticated understanding of the impact of SEL programs in general and individual curriculum specifically, and are representative of a more mature field of scientific inquiry. Indeed, SEL researchers have uncovered a variety of individual- and contextual-level variables that moderate the effectiveness of SEL programs (Bierman et al., 2010; Durlak et al., 2011).

Conclusion

This study offers some insight into the effectiveness of *Second Step* to promote positive outcomes. Perhaps more questions have surfaced given this research than have been answered, including what SEL skills lead to improved academic outcomes and what are the factors that contribute to teachers who implement lessons in an engaging way and are committed to implementing more of the lessons? Despite the questions that remain unanswered, this study adds to the growing literature on *Second Step*, which is one of the most widely adopted and implemented SEL programs in the United States. Researchers should continue to explore the impact of *Second Step* and how to improve implementation to enhance student outcomes.

References

Bailey, J. S., & Burch, M. R. (2002). *Research methods in applied behavior analysis*. Thousand Oaks, CA: Sage. http://dx.doi.org/10.4135/9781412985710

Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51,* 1173–1182. http://dx.doi.org/10.1037/0022-3514.51.6.1173

Bierman, K. L., Coie, J. D., Dodge, K. A., Greenberg, M. T., Lochman, J. E., McMahon, R. J., . . . the 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,* 156–168. http://dx.doi.org/10.1037/a0018607

Biglan, A., Hood, D., Brozovsky, P., Ochs, L., Ary, D., & Black, C. (1991). Subject attrition in prevention research. In C. B. Leukefeld & W. J. Bukoski (Eds.), *Drug abuse prevention intervention research: Methodological issues* (pp. 213–234). Rockville, MD: National Institute on Drug Abuse.

Brackett, M. A., & Rivers, S. E. (2014). An emotionally intelligent approach to bullying prevention. *Education Week, 33,* 40–33.

Buscemi, L., Bennett, T., Thomas, D., & Deluca, D. A. (1996). Head start challenges and training needs. *Journal of Early Intervention, 20,* 1–13. http://dx.doi.org/10.1177/105381519602000101

Bushaw, W. J., & Lopez, S. J. (2010). A time for change: The 42nd annual Phi Delta Kappa/Gallup Poll of the public's attitudes toward the public schools. *Phi Delta Kappan, 92,* 8–26. http://dx.doi.org/10.1177/003172171009200103

Caprara, G. V., Barbaranelli, C., Pastorelli, C., Bandura, A., & Zimbardo, P. G. (2000). Prosocial foundations of children's academic achievement. *Psychological Science, 11,* 302–306. http://dx.doi.org/10.1111/1467-9280.00260

Collaborative for Academic and Social Emotional Learning. (2014). Program Rating Guide. Retrieved from https://casel.org/guide/ratings/elementary/

Committee for Children (CfC). (2016). *Second Step social-emotional programming*. Seattle, WA: Author.

Domitrovich, C. E., Bradshaw, C. P., Greenberg, M. T., Embry, D., Poduska, J. M., & Ialongo, N. S. (2010). Integrated models of school-based prevention: Logic and theory. *Psychology in the Schools, 47,* 71–88.

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,* 327–350.

Durlak, J. A., Weissberg, R. P., Dymnicki, A. B., Taylor, R. D., & Schellinger, K. B. (2011). The impact of enhancing students' social and emotional learning: A meta-analysis of school-based universal interventions. *Child Development, 82,* 405–432. http://dx.doi.org/10.1111/j.1467-8624.2010.01564.x

Faraone, S. V., Biederman, J., Weber, W., & Russell, R. L. (1998). Psychiatric, neuropsychological, and psychosocial features of *DSM–IV* subtypes of attention-deficit/hyperactivity disorder: Results from a clinically referred sample. *Journal of the American Academy of Child & Adolescent Psychiatry, 37,* 185–193. http://dx.doi.org/10.1097/00004583-199802000-00011

Flay, B. R., Biglan, A., Boruch, R. F., Castro, F. G., Gottfredson, D., Kellam, S., . . . Jil, P. (2005). Standards of evidence: Criteria for efficacy, effectiveness and dissemination. *Prevention Science, 6,* 151–175.

Fuchs, L. S., & Fuchs, D. (1999). Monitoring student progress toward the development of reading competence: A review of three forms of classroom-based assessment. *School Psychology Review, 28,* 659–671.

Gillham, J. E., Reivich, K. J., Freres, D. R., Chaplin, T. M., Shatté, A. J., Samuels, B., . . . Seligman, M. E. (2007). School-based prevention of depressive symptoms: A randomized controlled study of the effectiveness and specificity of the Penn Resiliency Program. *Journal of Consulting and Clinical Psychology, 75,* 9–19. http://dx.doi.org/10.1037/0022-006X.75.1.9

Graham, H. (2009). *Understanding health inequalities*. New York, NY: McGraw-Hill International.

Greenberg, M. T., Weissberg, R. P., O'Brien, M. U., Zins, J. E., Fredericks, L., Resnik, H., & Elias, M. J. (2003). Enhancing school-based prevention and youth development through coordinated social, emotional, and academic learning. *American Psychologist, 58*(6–7), 466–474. http://dx.doi.org/10.1037/0003-066X.58.6-7.466

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

Heckman, J. J., & Masterov, D. V. (2004). *The productivity argument for investing in young children (Working Paper 5). Invest in Kids Working Group, Committee for Economic Development* (Unpublished manuscript). Chicago, IL: University of Chicago.

Hinshaw, S. P. (1992). Academic underachievement, attention deficits, and aggression: Comorbidity and implications for intervention. *Journal of Consulting and Clinical Psychology, 60,* 893–903. http://dx.doi.org/10.1037/0022-006X.60.6.893

Holmbeck, G. N. (1997). Toward terminological, conceptual, and statistical clarity in the study of mediators and moderators: Examples from the child-clinical and pediatric psychology literatures. *Journal of Consulting & Clinical Psychology, 65,* 599–610.

Holsen, I., Iversen, A. C., & Smith, B. H. (2009). Universal social competence promotion programme in school: Does it work for children with low socio-economic background? *Advances in School Mental Health Promotion, 2,* 51–60. http://dx.doi.org/10.1080/1754730X.2009.9715704

Holsen, I., Smith, B. H., & Frey, K. S. (2008). Outcomes of the social competence program *Second Step* in Norwegian Elementary Schools. *School Psychology International, 29,* 71–88. http://dx.doi.org/10.1177/0143034307088504

Jones, S. M., Brown, J. L., & Lawrence Aber, J. (2011). Two-year impacts of a universal school-based social–emotional and literacy intervention: An experiment in translational developmental research. *Child Development, 82,* 533–554. http://dx.doi.org/10.1111/j.1467-8624.2010.01560.x

Kazdin, A. E. (2003). *Research design in clinical psychology*. Needham Heights, MA: Allyn & Bacon.

Langdon, C. A., & Vesper, N. (2000). The sixth Phi Delta Kappa poll of teachers' attitudes toward the public schools. *Phi Delta Kappan, 81,* 607–611.

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

Low, S., Smolkowski, K., & Cook, C. (2016). What constitutes high-quality implementation of SEL programs? A latent class analysis of Second Step implementation. *Prevention Science, 17,* 981–991. http://dx.doi.org/10.1007/s11121-016-0670-3

Malecki, C. K., & Elliot, S. N. (2002). Children's social behaviors as predictors of academic achievement: A longitudinal analysis. *School Psychology Quarterly, 17,* 1–25. http://dx.doi.org/10.1521/scpq.17.1.1.19902

Marshall, R. M., Hynd, G. W., Handwerk, M. J., & Hall, J. (1997). Academic underachievement in ADHD subtypes. *Journal of Learning Disabilities, 30,* 635–642. http://dx.doi.org/10.1177/002221949703000607

Morgan, P. L., Farkas, G., & Wu, Q. (2011). Kindergarten children's growth trajectories in reading and mathematics: Who falls increasingly behind? *Journal of Learning Disabilities, 44,* 472–488. http://dx.doi.org/10.1177/0022219411414010

Murray, D. M. (1998). *Design and analysis of group-randomized trials*. New York, NY: Oxford University Press.

National Research Council. (2012). *The National Academies—Where the Nation turns for independent, expert advice*. Retrieved from http://www.nationalacademies.org/nrc/

Nebbergall, A. (2009). *An experimental evaluation of the effects of school-based universal prevention program on parent and teacher ratings of student behavior* (Doctoral dissertation). University of Maryland, College Park, Maryland.

Neil, A. L., & Christensen, H. (2009). Efficacy and effectiveness of school-based prevention and early intervention programs for anxiety.

Clinical Psychology Review, 29, 208–215. http://dx.doi.org/10.1016/j.cpr.2009.01.002

Owens, J. S., Lyon, A. R., Brandt, N. E., Warner, C. M., Nadeem, E., Spiel, C., & Wagner, M. (2014). Implementation science in school mental health: Key constructs in a developing research agenda. *School mental health, 6,* 99–111.

Perepletchikova, F., & Kazdin, A. E. (2005). Treatment integrity and therapeutic change: Issues and research recommendations. *Clinical Psychology: Science and Practice, 12,* 365–383. http://dx.doi.org/10.1093/clipsy.bpi045

Perepletchikova, F., Treat, T. A., & Kazdin, A. E. (2007). Treatment integrity in psychotherapy research: Analysis of the studies and examination of the associated factors. *Journal of Consulting and Clinical Psychology, 75,* 829–841. http://dx.doi.org/10.1037/0022-006X.75.6.829

Powell, B. J., Waltz, T. J., Chinman, M. J., Damschroder, L. J., Smith, J. L., Matthieu, M. M., . . . Kirchner, J. E. (2015). A refined compilation of implementation strategies: Results from the Expert Recommendations for Implementing Change (ERIC) project. *Implementation Science, 10,* 21. http://dx.doi.org/10.1186/s13012-015-0209-1

Proctor, E. K., & Brownson, R. C. (2012). Measurement issues in dissemination and implementation research. *Dissemination and implementation research in health: Translating science to practice*. New York, NY: Oxford University Press.

Raghunathan, E., Reiter, J. P., & Rubin, D. B. (2003). Multiple imputation for statistical disclosure limitation. *Journal of Official Statistics, 19,* 1–16.

Raudenbush, S. W., Bryk, A., Cheong, Y. F., Congdon, R., & du Toit, M. (2011). HLM statistical software (Version 6) [Computer software]. Lincolnwood, IL: Scientific Software International, Inc.

Rhoades, B. L., Warren, H. K., Domitrovich, C. E., & Greenberg, M. T. (2011). Examining the link between preschool social–emotional competence and first grade academic achievement: The role of attention skills. *Early Childhood Research Quarterly, 26,* 182–191. http://dx.doi.org/10.1016/j.ecresq.2010.07.003

Sanetti, L. M. H., & Kratochwill, T. R. (2009). Toward developing a science of treatment integrity: Introduction to the special series. *School Psychology Review, 38,* 445.

Sanetti, L. M. H., & Kratochwill, T. R. (2011). An evaluation of the treatment integrity planning protocol and two schedules of treatment integrity self-report: Impact on implementation and report accuracy. *Journal of Educational & Psychological Consultation, 21,* 284–308. http://dx.doi.org/10.1080/10474412.2011.620927

Schonert-Reichl, K. A. (2017). Social and emotional learning and teachers. *The Future of Children, 27,* 137–155.

Shapiro, E. S., & Kratochwill, T. R. (Eds.). (2000). *Behavioral assessment in schools: Theory, research, and clinical foundations*. New York, NY: Guilford Press.

Simonsen, B., Fairbanks, S., Briesch, A., Myers, D., & Sugai, G. (2008). A review of evidence-based practices in classroom management: Considerations for research to practice. *Education & Treatment of Children, 31,* 351–380. http://dx.doi.org/10.1353/etc.0.0007

Snow, C. E., Burns, M. S., & Griffin, P. (1999). *Language and literacy environments in preschools*. (ERIC Document Reproduction Service No. ED426818).

Sprague, J., Walker, H., Golly, A., White, K., Myers, D. R., & Shannon, T. (2001). Translating research into effective practice: The effects of a universal staff and student intervention on indicators of discipline and school safety. *Education & Treatment of Children, 24,* 495–511.

Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly, 24,* 360–407. http://dx.doi.org/10.1598/RRQ.21.4.1

Stokes, T. F., & Baer, D. M. (1977). An implicit technology of generalization. *Journal of Applied Behavior Analysis, 10,* 349–367. http://dx.doi.org/10.1901/jaba.1977.10-349

Sutherland, K. S., & Wehby, J. H. (2001). Exploring the relationship between increased opportunities to respond to academic requests and the academic and behavioral outcomes of students with EBD. *Remedial and Special Education, 22,* 113–121. http://dx.doi.org/10.1177/074193250102200205

Thurber, R. S., Shinn, M. R., & Smolkowski, K. (2002). What is measured in mathematics tests? Construct validity of curriculum-based mathematics measures. *School Psychology Review, 31,* 498–513.

U.S. Census Bureau. (2011). *2011 Data release*. Retrieved from http://www.census.gov/acs/www/data_documentation/2011_release/

Viera, A. J., & Garrett, J. M. (2005). Understanding interobserver agreement: The kappa statistic. *Family Medicine, 37,* 360–363.

Vitaro, F., Brendgen, M., & Tremblay, R. E. (2001). Preventive intervention: Assessing its effects on the trajectories of delinquency and testing for mediational processes. *Applied Developmental Science, 5,* 201–213. http://dx.doi.org/10.1207/S1532480XADS0504_02

Volpe, R. J., DiPerna, J. C., Hintze, J. M., & Shapiro, E. S. (2005). Observing students in classroom settings: A review of seven coding schemes. *School Psychology Review, 34,* 454–473.

Walker, H. M., Ramsey, E., & Gresham, F. M. (2004). *Antisocial behavior in school: Evidence-based practices*. Belmont, CA: Wadsworth Publishing Company.

Zins, J. E., Bloodworth, M. R., Weissberg, R. P., & Walberg, H. J. (2007). The scientific base linking social and emotional learning to school success. *Journal of Educational & Psychological Consultation, 17*(2–3), 191–210. http://dx.doi.org/10.1080/10474410701413145

Received September 25, 2015
Revision received September 21, 2017
Accepted September 26, 2017 ■