Journal of Applied Developmental Psychology 62 (2019) 145–159

Contents lists available at ScienceDirect

Journal of Applied Developmental Psychology

journal homepage: www.elsevier.com/locate/jappdp

A randomized efficacy trial of the second step early learning (SSEL) curriculum

Carole C. Upshur[a,b,*], Melodie Wenz-Gross[a,c], Christopher Rhoads[d], Miriam Heyman[a,1], Yeonsoo Yoo[a,2], Gail Sawosik[a,3]

[a] Department of Family Medicine and Community Health, University of Massachusetts Medical School, 55 Lake Ave North, Worcester, MA 01655, United States
[b] Department of Quantitative Health Sciences, University of Massachusetts Medical School, 55 Lake Ave North, Worcester, MA 01655, United States
[c] Department of Psychiatry, University of Massachusetts Medical School, 55 Lake Ave North, Worcester, MA 01655, United States
[d] Neag School of Education, Dept of Measurement, Evaluation and Assessment, University of Connecticut, 249 Glenbrook Road Unit 3064, Storrs, CT 06269, United States

ARTICLE INFO

Keywords:
Early childhood
Preschool
Head start
Executive functioning
Social-emotional skills
Classroom quality

ABSTRACT

A classroom randomized efficacy trial conducted over four years in 7 community-based preschool and 6 Head Start programs investigated effects of the Second Step Early Learning (SSEL) curriculum on end of preschool executive functioning (EF) and social-emotional (SE) skills in low-income children. Outcomes are reported for $n = 770$ four-year-olds independently assessed for EF and SE by study staff in fall and spring of the pre-kindergarten year. Main outcomes were analyzed using two, three- level hierarchical linear models, one each for EF and SE skills. A significant effect (effect size of 0.15) for EF and a nonsignificant effect for SE were found. Secondary analyses found no significant differences on pre-academic skills. SSEL appears to have a meaningful impact on at-risk children's EF skills that supports its continued dissemination.

Introduction

Important skills for both school success and adult functioning develop during critical periods in early childhood. These include executive functioning (EF) skills such as inhibitory control, working memory, and attention (Bierman & Torres, 2016; Blair & Diamond, 2008; McClelland et al., 2007; Morrison, Ponitz, & McClelland, 2010), as well as social/emotional (SE) skills such as understanding and identifying emotions, showing empathy, regulating behavior, and establishing positive peer interactions (Carlo, Knight, Eisenberg, & Rotenberg, 1991; Crick & Dodge, 1994; Eisenberg, Spinrad, & Eggum, 2010). Central to development of such skills is the quality of the early environment, including adequate nutrition, health, and especially reciprocal and nurturing interactions with caregivers, including early childhood educators (Bierman & Torres, 2016; Bruce, Gunnar, Pears, & Fisher, 2013; National Scientific Council on the Developing Child, 2004). Deficits in the early environment such as stress due to poor socioeconomic conditions can significantly impede children's development in these areas by affecting underlying brain architecture and neurological maturation (Blair & Raver, 2015; Denham et al., 2012; Raver, 2004; Shonkoff, 2017; Yoshsikawa, Aber, & Beardslee, 2012). Thus, young children from lower SES backgrounds or those exposed to other stressors (e.g. parental depression) often miss opportunities to learn and practice prosocial and executive functioning skills prior to school entry which may underlie academic achievement gaps, resulting in less social competence, engagement in learning, and more externalizing problems than more advantaged peers (Ackerman, Brown, & Izard, 2004; Denham et al., 2012; Kaiser, Hancock, Cai, Foster, & Hester, 2000; Schultz, Izard, Ackerman, & Youngstrom, 2001; Sektnan, McClelland, Acock, & Morrison, 2010).

Association of EF and SE with school readiness and academic outcomes

Early achievement gaps for low-income children are well

* Corresponding author at: Department of Family Medicine and Community Health, University of Massachusetts Medical School, 55 Lake Ave North, Worcester, MA 01655, United States.
E-mail addresses: Carole.upshur@umassmed.edu (C.C. Upshur), melodie.wenzgross@umassmed.edu (M. Wenz-Gross), Christopher.rhoads@uconn.edu (C. Rhoads), Miriam.heyman@umassmed.edu (M. Heyman), yeonsoo.yoo@umassmed.edu (Y. Yoo), gail.sawosik@umassmed.edu (G. Sawosik).
[1] Present address: Ruderman Family Foundation, 2150 Washington St. Suite 300, Newton, MA 02462, United States;
[2] Present address: Fuller Theological Seminar, 135 Oakland Ave, Pasadena, CA 91182, United States;
[3] Present address: Reach Out and Read, PO Box 462, Westford, MA 01886, gail.sawosik@reachoutandread.org

https://doi.org/10.1016/j.appdev.2019.02.008
Received 23 March 2018; Received in revised form 16 February 2019; Accepted 26 February 2019

documented and a meta-analysis of several decades of preschool interventions to improve school readiness and academic outcomes for these children demonstrated modest effect sizes on cognitive and achievement scores (Duncan & Magnuson, 2013). Fairly consistently the gains disappear within a few years, although some longitudinal studies have found longer term global effects of preschool attendance on adolescent and adult functioning such as higher high school completion rates, lower delinquency rates, etc. (Duncan & Magnuson, 2013). These modest outcomes and lack of long-term academic effects from most compensatory preschool programs like Head Start, have led to current thinking that it is more important to bolster underlying neurocognitive self-regulation skills in at risk children, rather than to focus solely on preacademic skills such as letter recognition or counting (Blair & Raver, 2015).

Evidence that preschool SE and EF affect school readiness and adjustment is growing. For example, preschool emotion knowledge was found to predict kindergarten teacher's ratings of academic success (Denham, Wyatt, Bassett, Echeverria, & Knox, 2009), as well as kindergarten assessed literacy and numeracy (Torres, Domitrovich, & Bierman, 2015), and academic and attention skills in first grade (Rhoades, Warren, Domitrovich, & Greenberg, 2011). In addition, higher levels of EF or preschool growth in EF have been found to be associated with Head Start children's math and reading scores in kindergarten (Welsh, Nix, Blair, Bierman, & Nelson, 2010). The attention component of EF seems to be particularly powerful in predicting longitudinal academic outcomes. Duncan et al. (2007) found that attention skills at age 5 and 6 predicted math and reading scores in adolescence, and parent ratings of child attention at age 4 were found to be associated with post-high school reading and math skills as well as college graduation rates (McClelland, Acock, Piccinin, Rhea, & Stallings, 2013). In a longitudinal study of the effect of pre-kindergarten EF skills, as well as learning-related behaviors (e.g., following directions, and classroom rules and routines, and engaging in learning tasks), on Head Start children, Sasser, Bierman, and Heinrichs (2015) found EF directly predicted level and growth through third grade in math and teacher-rated academic engagement, while the learning-related behaviors predicted reading skills, and teacher-rated aggression and social skills.

Theoretical link of SE and EF to academic outcomes

While SE and EF skills are interrelated and develop transactionally in early childhood (Blair, 2002; Ursache & Blair, 2012), they each may contribute unique dimensions to self-regulation that improve children's capacity for learning. Lack of the SE skill of identifying and regulating emotion interferes with ability to follow classroom rules and routines (Campbell, Shaw, & Gilliom, 2000; Eisenberg et al., 2010; Liu, 2004). A related EF skill that requires children to inhibit a prepotent response (inhibitory control) is required for children to wait their turn or avoid lashing out at a peer (Blair, 2002). Inhibitory control, along with emotion knowledge has been found to be associated with prosocial behavior and social problem solving in 3- and 4-year-olds (Denham, Bassett, Zinsser, & Wyatt, 2014). Aggression and hyperactivity in early childhood have also been associated with poor EF skills (Hughes & Ensor, 2007, 2011). Nesbitt, Farran, and Fuhs (2015) investigated how EF skills may be linked to academic success and found that children with better attention, inhibition, and working memory were able to engage more in learning related behaviors such as following multi-step instructions, cooperating with teachers and peers, and being less disruptive. Both SE and EF skills are based in neurological development of brain functions in the limbic and prefrontal cortex, which can be disrupted or overwhelmed by excess arousal and stress in early childhood without compensatory interventions (Fox, Calkins, & Bell, 1994; Shonkoff, 2017). However, recent research in early brain development shows that intervention can be effective in addressing and ameliorating such deficits due to early brain plasticity (Boyce, 2016; Shonkoff, 2017; Takesian & Hensch, 2013). For example, an 8-week circle time intervention that taught increasingly complicated stop and go games using oral and visual cues, significantly improved preschool children's performance on an EF task for those children who started the school year with lower scores, and was significantly associated with end of year gains in a pre-literacy measure (Tominey & McClelland, 2011).

Universal classroom interventions designed to increase preschool Children's SE and EF

Building on increasing understanding of young children's neurocognitive development, and the ways in which poverty and other stressors can disrupt the development of basic attentional and self-regulatory processes which can interfere with learning, a number of classroom interventions have been developed and tested in the last two decades. Examples of these include the PATHS (Promoting Alternative Thinking Strategies) curriculum, the Incredible Years/Dinosaur School curriculum, the Chicago School Readiness Program, and Tools of the Mind. The PATHS intervention consists of once or twice weekly classroom lessons delivered by classroom teachers that focus on positive classroom routines, emotion knowledge and expressing feelings, self-control strategies (e.g. self-talk and calming down), and social problem solving. In a classroom randomized study of Head Start children, PATHS was found to have significant effect on several directly-assessed SE measures (e.g., emotion knowledge, anger attribution) but no significant effect on EF measures such as inhibitory control, attention, or problem solving (Domitrovich, Cortes, & Greenberg, 2007). Another version of PATHS implemented in a randomized controlled trial of Head Start programs in conjunction with a literacy curriculum (Head Start REDI-Research-Based Developmentally Informed) versus "usual curricula," found significant effects on some literacy outcomes, as well as directly assessed emotion recognition, increased prosocial responses, and decreased aggression (Bierman et al., 2008). A marginally significant effect was found on one directly assessed EF measure in the REDI study (Bierman, Nix, Greenberg, Blair, & Domitrovich, 2008), and longitudinal outcomes following participants into third grade documented better trajectories for intervention children on teacher-rated social and learning behaviors (Nix et al., 2016), and children with low EF in preschool did better by third grade on a composite of directly assessed EF skills and academic competence (Sasser, Bierman, Heinrichs, & Nix, 2017).

The Incredible Years/Dinosaur School (IY) intervention is based on clinical work with young children with conduct problems and provides extensive teacher training on positive classroom management paired with scripted classroom lessons delivered two times per week (Webster-Stratton, Reid, & Stoolmiller, 2008). A study of preschool/Head Start through first grade classrooms compared usual curricula to IY and found significant impact on teacher's use of positive discipline and overall improved classroom atmosphere, but no effects on child directly-observed behavior problems or classroom disengagement, although children with higher baseline problems seemed to improve most (Webster-Stratton et al., 2008). The IY materials for teacher training were implemented along with in-class mental health coaches in the Chicago School Readiness Program (CSRP) designed to improve the quality of classrooms and teacher-interactions with children facing adverse risks, and to improve school readiness skills (Raver et al., 2009). Using site randomized Head Start classrooms, CSRP found significant teacher-reported decreases in internalizing and externalizing child behavior problems, but a non-significant trend to reduced disruptive behaviors using direct child observation. In examining EF skills, CSRP found significant gains for intervention children on directly assessed EF tasks that required memory and set shifting, but not inhibition, as well as significant gains in literacy and premath skills (Raver et al., 2011).

Finally, the Tools of the Mind curriculum was developed to promote the development specifically of EF skills (e.g., inhibitory control,

working memory, and cognitive flexibility/problem solving) as a classroom-based teacher delivered intervention (Diamond, Barnett, Thomas, & Munro, 2007). It is based on cognitive development theory that suggests that social learning, imaginative play, and use of language contribute to children's self-regulation. Teachers are provided a set of daily activities that scaffold play scenarios to reinforce collaborative behavior, planning and self-regulation of children (such as using self-talk to remember rules and follow directions). While early studies among urban preschool children compared to other curricula found significant gains in directly assessed EF skills (Diamond et al., 2007) subsequent studies have found mixed results across different age groups. For example, Solomon et al. (2018) found no main effects for Tools of the Mind implemented in a classroom randomized study of Canadian preschools compared to usual district curricula.

Taken together, these studies verify the importance of development of early self-regulatory behaviors to children's ability to learn and achieve in school. They also suggest that work remains to understand how to optimally promote the various aspects of social/emotional and executive functioning skills in the context of preschool classrooms serving low-income children who are particularly at risk of lagging in such development.

The current cluster-randomized control study evaluated a newer curriculum (the Second Step Early Learning Curriculum, or SSEL, Committee for Children, 2011a) with similar aims to improve social/emotional skills and self-regulation in at risk preschoolers, in two cohorts of Head Start and community preschool programs serving low-income children. It was designed in response to a call from the Institute of Education Sciences for increased investigations of interventions focused on social/emotional learning in preschool. The University of Massachusetts Medical School Institutional Review Board approved the study for human subjects' participation.

SSEL curriculum

SSEL is a commercially available early learning kit (http://www.cfchildren.org/second-step/early-learning) targeted to 4–5 year olds, although it can also be used in mixed age 3–5 year old classrooms. The Committee for Children estimated in 2016 that close to 8000 classrooms nationally were using the curriculum (email communication, Committee for Children, April 2018). SSEL is based on an extensive review of the developmental literature in several areas, including, the development of social emotional competence and self-regulation, and how these contribute to school readiness. SSEL uniquely integrates activities and instruction in emotion recognition, empathy, and social problem solving with self-regulation techniques such as self-talk and learning to calm down, along with daily 'Brain Games" that that require starting and stopping activities based on various oral or visual cues. The intervention components of SSEL thus span many of the approaches of prior interventions that have had promising results such as PATHS (Domitrovich et al., 2007), and Red Light, Purple Light (Tominey & McClelland, 2011). The logic model for SSEL describes the combination of the weekly theme curriculum topics along with ongoing teacher reinforcement, the Brain Builder games, and providing information to parents for reinforcement at home, will increase short-term/proximal outcomes of children's self-regulation and social/emotional competence, reduce aggression, improve peer relations, and improve on-task and classroom behavior, which will in turn result in distal/long-term improved school readiness, success, and engagement. SSEL is not designed as a substitute for literacy, math, or science activities, but rather addresses the underlying social and cognitive processes necessary for successful learning that will improve children's overall behavioral and academic success (Committee for Children, 2011b).

SSEL has scripted, five day-a-week, brief large and small group lessons with 28 weekly themes, along with suggested extension and generalization activities. Day 1 of each week introduces the weekly theme using puppets; Day 2 uses the picture on the curriculum card to describe a theme-related situation and how to solve it; Days 3 and 4 are reinforcement days that involve small or large group practice activities; and Day 5 involves reading a book that addresses the weekly theme, such as recognizing emotions, or playing fairly. The kit comes with large, colorful weekly lesson cards designed to show children a situation reflecting the weekly theme, with the teacher's script and instructions on the back; a CD of songs to be played and sung daily, with words that reinforce the weekly theme; puppets that are used to talk out the lessons; posters showing different social-emotional skills; small cards with children's faces showing different emotions; and detailed instructions for Brain Builder games to be played daily that are intended to help children practice attention, working memory, and inhibition. Strategies for reinforcing EF skills are given such as asking children to engage in 'think time' before raising their hand, asking for the group to show nonverbal agreement (e.g., pat your head) to engage them when one child or the teacher is giving an answer, and using random calling in group activities to bring children back to focus and reinforce those paying attention but sitting quietly. In addition, there are suggestions to teachers for strategies to reinforce the specific skills throughout the day by asking children to *think ahead* about using the skills taught in upcoming activities, providing *ongoing reinforcement* when children demonstrate the skills, and asking children to *think back* and recall when they or someone else demonstrated a skill, as well as art, literacy, math, and STEM extension activities that incorporate the learning strategies of the curriculum. There are five units:1) six lessons covering skills for learning such as listening, paying attention, using self-talk to remember directions, and asking for help; 2) six lessons on empathy such as identifying feelings in self and others, learning how others feel, and demonstrating caring and helpfulness towards others; 3) six lessons on managing emotions such as identifying strong emotions and calming down; 4) seven lessons on friendship skills such as how to join a group, inviting others to play, fair ways to play, and techniques for calming down and solving problems; and 5) three lessons to review skills in preparation for transition to kindergarten. In addition to the materials for teachers, the curriculum kit provides a weekly hand out, that can be copied and distributed to parents, covering the weekly theme and activities that could be carried out at home to reinforce the theme.

Study rationale and hypotheses

SSEL was picked for evaluation because of its adequately detailed curricular materials and instructions for teachers such that it can be replicated (Schindler, Fisher & Shonkoff, 2017), its reasonable cost, and its focus on improving SE and EF skills through a universal classroom curriculum. Studies of at-risk preschool children enrolled in a precursor curriculum called Second Step, found reductions in classroom disruptiveness, higher quality classroom climate, and teacher interactions, as well as evidence for increased child emotion knowledge (McMahon, Washburn, Felix, Yakin, & Childrey, 2000; Upshur, Wenz-Gross, & Reed, 2013). SSEL incorporated more recent knowledge about the development of self-regulation in young children, changed the format from weekly to daily integration into other preschool activities, and added to already well-developed emotion knowledge and regulation units, activities and units addressing learning-related behaviors, working memory, and inhibitory control. Preliminary outcomes from Cohort 1 of the current study found significant differences between the intervention and control groups in executive functioning skills, and marginally significant differences in social/emotional skills (Upshur, Heyman, & Wenz-Gross, 2017). Another manuscript from the current study examined intervention contributions to children's kindergarten readiness scores at follow-up from preschool. Using structural equation modeling (SEM) analyses, no direct effect of SSEL was found on kindergarten readiness, but we found a relation between gains in EF skills and gains in preacademic skills and on-task behavior in preschool, and these in turn were associated with better kindergarten readiness scores (Wenz-Gross, Yoo, Upshur, & Gambino, 2018). No other published

studies currently exist on this curriculum.

This manuscript incorporates data from all four years of the study and explores the following main hypotheses related to the proximal outcomes for social/emotional learning and self-regulation described by the SSEL's logic model (Committee for Children, 2011a). Our two main hypotheses were:

1) Children receiving the SSEL curriculum in preschool compared to children in similar classrooms not receiving SSEL, will have higher end of year EF skills controlling for beginning of year EF, SE, general cognition and demographics.
2) Children receiving the SSEL curriculum in preschool compared to children in classrooms not receiving SSEL, will have higher end of year SE skills controlling for beginning of year SE, EF, general cognition, and demographics.

In addition to investigating the primary impacts on SE and EF, we explored SSEL's secondary effects on academic school readiness skills by examining prereading and math skills, at end of preschool. We also explored classroom quality, since improving classroom climate and teacher interactions had been found in an earlier version of the curriculum (Upshur et al., 2013). Our two secondary hypotheses were:

1) Children receiving the SSEL curriculum in preschool compared to children in similar classrooms not receiving SSEL, will have higher end of year preacademic skills, controlling for beginning of year preacademic skills, general cognition, and demographics.
2) Classrooms where teachers were trained to deliver the SSEL curriculum will be rated as higher quality at the end of the preschool year, controlling for beginning of preschool year quality ratings.

Method

Participants

Sample children

A total of 770 children comprise the final analysis sample with complete pretest, demographic, and outcome data. They were assessed for SE and EF outcomes in one of the four years of the study when they were age 4 or slated to enter kindergarten the following year. Of these, 383 were assessed only in the first year of each cohort, and 207 only in the second year of each cohort, after one year of preschool. An additional group of 181 were assessed in the second year of each cohort but had been present in the study the prior year, as three year-olds, and thus had two years of preschool, but only one year of directly assessed outcomes. Table 1 provides baseline characteristics. Children's average age was 53.0 months (SD = 4.0), and they were evenly divided by sex. The sample was diverse with about one quarter African American, and about two-fifths each Anglo American and Hispanic American. The Peabody Picture Vocabulary Test (PPVT) standard score (M = 96.9, SD = 15.2) was slightly below the population mean. About half the sample's primary parent had only a high school education or less, about one-quarter lived with a married parent, and slightly more than one quarter had incomes below $10,000 a year, with more than half of sample families reporting incomes of less than $20,000 annually.

Sample teachers

A total of 187 teachers in both conditions participated in the study over 4 years. They were mostly female (only three were male); 78% were Anglo American, 14% Hispanic American, and 6% African American. Teacher's average age was 37 years (SD = 12.5; range = 19 to 66), mean childcare experience was 13 years (SD = 9.0; range = 0 to 42 years), and most teachers (55%) had a college degree or higher, with another 32% having an Associates degree, and 13% having only a high school diploma. Fewer than half of teachers (42%) were married, while two thirds had children of their own. There were no differences in any of the collected teacher characteristics between intervention and control teachers.

Table 1
Baseline characteristics of children and families.

	Whole sample (n = 770)	Intervention (n = 393)	Control (n = 377)	P value
Child age in months Mean (SD)	53.0 (3.98)	53.2 (3.91)	52.7 (4.03)	0.23
Child sex (%)				
Male	50.8	50.6	50.9	0.93
No. child in household				
Mean (SD)	2.36 (1.10)	2.28 (1.09)	2.44 (1.11)	0.06
Child ethnicity (%)				
Anglo-American	42.3	44.8	39.8	0.46
African-American	26.0	24.7	27.3	0.53
Hispanic-American	40.6	36.6	44.8	0.06
Asian-American	2.1	2.5	1.6	0.38
Other	2.9	2.3	3.4	0.34
Not provided	8.2	9.9	6.4	0.09
PPVT standard score Mean (SD)	96.90 (15.2)	98.09 (14.9)	95.66 (15.4)	0.10
Parent education (%)				
< High school	12.1	12.0	12.2	0.93
High school	33.4	34.6	32.1	0.55
Some college	42.7	41.7	43.8	0.61
College graduate	11.8	11.7	11.9	0.90
Family income (%)				
< $10,000	26.6	27.0	26.2	0.75
$10,000–$19,999	27.1	26.5	27.9	0.67
$20,000–$29,999	24.0	23.4	24.7	0.70
$30,000–$39,999	10.1	9.7	10.6	0.66
$40,000–$49,999	5.1	5.1	5.0	0.95
$50,000+	7.0	8.4	5.6	0.32
Marital Status (%)				
Married	28.4	29.3	27.6	0.61

Note. For continuous variables p-values are from two level HLMs with cohort and condition as the independent variables. For categorical variables p-values are from two-level logistic models with an indicator variable for the category in question as the outcome variable and cohort and condition as the independent variables.

Sample classrooms

Initially sixty-seven classrooms in 13 sites were randomly assigned within site to deliver SSEL or usual curricula, for a period of two years each during the school year, 34 classrooms in 7 sites in Cohort 1 and 33 classrooms in 6 sites in Cohort 2. The sites were first randomly assigned to either cohort 1 or 2, and then random assignment of classrooms within site was conducted. Six sites were Head Start programs and 7 were community preschool programs that enrolled a large proportion of low-income and at-risk children. Both types of preschools participated in each cohort, but the sites were completely different in each cohort. Teacher turnover, enrollment expansions and contractions, and administrative switching of teacher assignments resulted in classroom changes within each cohort between the two years of participation. Because most classrooms were mixed age, but only 4-year olds were assessed, classrooms where there were no or only one 4-year old present with complete data were dropped for the current analysis, resulting in two fewer intervention classrooms in Year 4 compared to Year 3.Two additional control classrooms were added in Year 2, that had been assigned as controls in Year 1, but did not have sufficient 4 year-old enrollment in Year 1 to remain in the classroom sample for that year. In Year 2 an additional intervention classroom was added due to a trained teacher being assigned to a new room, while the other trained teacher in the team remained as the lead teacher in the prior intervention classroom. We therefore have treated the classrooms as different each

year for purposes of child outcome analyses, resulting in 13 control and 15 intervention classrooms in Year 1; 15 control and 16 intervention classrooms in Year 2; 14 control and 15 intervention classrooms in Year 3; and 14 control and 13 intervention classrooms in Year 4, or a total of 56 control and 59 intervention classrooms across four years. The community preschools were all full day, year-round programs, while the Head Start classrooms consisted of some half day classrooms, and some extended day classrooms that met for 8 or 9 months a year. Because of these significant program differences, randomization was conducted within site/building where classrooms were similar in structure (e.g. program type, length of day etc.). Once study participation was completed for each cohort, the control classrooms received curriculum kits and a group training to use the curriculum. Site administrators and intervention teachers were also coached on ways to support spread of the curriculum to the control classrooms. All sites were accredited by the National Association for the Education of Young Children (https://www.naeyc.org/accreditation) and held state licenses for child care/early childhood education.

Curriculum training and implementation

For each cohort, curriculum kits and the related story books were purchased and distributed to all intervention classrooms in early November of Years 1 and 3, immediately after random assignment and completion of child baseline assessments. Intervention teachers were asked to add SSEL to their existing curricular requirements, which for both the community and Head Start classrooms was Creative Curriculum (Teaching Strategies, LLC, 2002–2012). Because Committee for Children cross-walked the SSEL lesson themes with the Creative Curriculum and Head Start frameworks (US Department of Health and Human Services, Administration for Children and Families, 2011) and demonstrated substantial overlap, implementing SSEL fulfilled many of the existing requirements in the participating classrooms, but provided more detailed instructional activities for the teachers to address existing SE/EF goals. Control teachers continued to implement Creative Curriculum without the additional SSEL activities. SSEL does not provide specific instruction on pre-math or pre-literacy skills and all classrooms continued with existing instructional strategies to incorporate these pre-academic skills into daily activities. All teachers and administrators were given an orientation to the study before initiation and assignment and asked not to share SSEL with control teachers. While some contamination could have occurred, sites were periodically asked about this and the feedback we received indicated contamination was not a problem.

Monthly, cross-site, 2-h large group teacher meetings were held in the first year for intervention teachers (7 in total), and bi-monthly meetings in the second year (5 in total), of each cohort to support implementation, address questions, and facilitate teacher's sharing of successes and challenges. Primary materials for these sessions were provided by Committee for Children to accompany the curriculum kits and are available on line to purchasers of the kits (e.g. power point slides outlining the logic model and themes, and videos demonstrating implementation of the teaching strategies and Brain Games etc.). These sessions were led by the PI and Co-PI, with assistance of other study staff and postdoctoral fellows and supplementary materials and activities were added to materials provided by Committee for Children, such as role- playing scenarios, small group discussion topics, and research summaries and videos (training slides and materials are available upon request from the authors).

Teachers were paid for these overtime sessions, and were given dinner, and continuing education credits. Anonymous feedback sheets were collected after each session to help prepare topics for the following session. The group sessions included a review and discussion of the curricular goals and how to implement activities for the various units of the curriculum, information on the underlying research on early child development that supported the curricular activities, video clips provided by Committee for Children or found on line, and small group discussions to problem solve on delivering aspects of the curriculum, or time for the teaching teams to plan extension activities for subsequent weeks.

After the first session in Years 1/3 which provided an overview of SSEL and gave instructions about getting started, each group session started with a check in about how far each teaching team was in delivering the curriculum and any specific challenges or successes, and included activities or topics based on feedback from the prior session (e.g. what teachers were confused about or wanted more help with), or activities to enhance implementation based on the classroom observations conducted by the core research staff the prior month. The initial sessions in Years 2/4 also summarized the prior year's successes and challenges and presented preliminary data about outcomes. Videos were also presented of some current classroom teachers successfully implementing the curriculum (parent permission for filming was secured).

The intervention classrooms were also visited monthly by a core study team member (PI, Co-PI, Field Coordinators, or post-doctoral fellows) trained with knowledge about SSEL goals and implementation activities, from the fall through May of the academic year (7 times in year 1 and 8 times in year 2). At these visits the classroom teachers were observed for 45–60 min delivering a specific SSEL lesson, and other classroom activities, and given on-the-spot coaching advice, and within a few days, a detailed written feedback memo on their implementation quality. A study-designed fidelity rating was also completed at the coaching/observation sessions to measure implementation rates, and to guide future coaching and group training sessions, but teachers were not made aware of their ratings. On-going discussions and interrater reliability were conducted periodically among the group of observer/coaches to assure all team members were providing consistent coaching and reliable implementation ratings.

Parent sessions were also held at each site, one or two a year, run jointly by intervention teachers and core study staff, to explain the curriculum and target skills, and encourage use of the weekly parent handouts to implement complementary activities at home. These were not well attended and other data about potential family involvement was beyond the scope of the current study. We did verify that the parent handouts were distributed weekly by the classroom teachers.

The study curriculum training was in addition to usual teacher training routinely provided to all preschool teachers by each site. While we did not track control teacher training, following state regulations, sites had monthly teacher trainings and other workshops that provided required continuing education credits covering issues such as child safety, review of curricula, and special topics such as positive discipline, creative arts projects etc. In the last two years of the study as new state licensing requirements were implemented establishing new standards for learning and social emotional activities for all preschools (Massachusetts Department of Early Education and Care, 2015), additional trainings on social emotional development and activities were provided at all sites. In several sites additional individual social emotional coaching was provided to control classroom teachers. Thus, while control classroom teachers did not benefit from specific instruction in the SSEL curriculum, particularly in the second cohort, many received additional training on implementing social/emotional supports in the classroom.

Study enrollment and data collection

At the beginning of each cohort, teachers were provided an explanation of the study by study staff, typically at site-based teacher meetings, and consented to providing brief background information (e.g. years teaching, education, etc.). Subsequently as new teachers were hired, the same consent was administered. Each fall site staff who had completed human subject protections training enrolled children by administering a university-approved consent form to parents when they

registered their children for preschool, or at drop off or pick up early in the school year. All families of preschool aged children (3 to 5 years old) were invited to participate if the parent could complete the consent form in either Spanish or English. Overall consent rates across the 4 study years ranged from 88 to 91% of all families enrolled at the site. Reasons for lack of consent included parent refusals, the family's language was other than English or Spanish, and the child was in state custody.

After enrollment, children who were 4 years old, or were identified as planning to enter kindergarten the following year, were individually assessed in the fall (September-early November) and again in the spring (late March–May) by study Research Assistants (RAs) who were blind to study condition and hypotheses. RAs were retired educators, or local college or graduate students majoring in psychology, education or related fields and had some previous research experience. They completed 12 h of group training on the assessment measures, and 3–6 h of in vivo training administering measures to children in classrooms under supervision of a senior study staff member. Assessments were all conducted in the child care classroom or adjacent space and consisted of two, 30–45-min sessions, with measures given in a set order to maintain children's interest. The tasks included measures of EF and SE, as well as basic pre-academic skills such as letter recognition and counting and were all administered in English. A small number of children each year who were determined by classroom teachers to not have adequate receptive or expressive English language skills were not maintained in the sample.

The timing of assessments was consistent with other classroom RCTs where data collection necessarily takes several weeks in the fall and spring and some fall assessments may occur after the curriculum has started or, conversely, occur in the spring before the entire curriculum is delivered (see Bierman et al., 2008; Domitrovich et al., 2007; Webster-Stratton et al., 2008). In Years 1 and 3, baseline child assessments were completed before the SSEL curriculum was introduced, but about half of end of year assessments were started before the curriculum was fully completed, resulting in approximately 5 months of curriculum implementation between assessments. In Years 2 and 4 when already trained teachers were participating, the curriculum was started at the beginning of the school year, before most child assessments were completed, and some spring assessments were again conducted before all weekly lessons were completed, representing a slightly longer period of intervention (6 months) for most children in the sample. Thus, intervention outcomes are based on children receiving 70–80% of the 25 weekly core intervention lessons on average. However, both intervention and control children in each site were assessed during the same weeks both fall and spring, so the assessment schedule should not bias outcomes.

Measures

Curriculum implementation and fidelity

Teachers completed weekly curriculum reports documenting how much of each weekly theme was completed (response choices were 25%, 50%, 75%, or 100%), how many days the Brain Games were played, and the extent to which they modified the curriculum (0 = *never*, 1 = *occasionally*, 2 = *often*, 3 = *always*) using a form provided in the curriculum kit (Committee for Children, 2011a). In addition, fidelity ratings were completed once a month by study staff on observations of a weekly lesson using a study-developed form that rated classrooms with 31 items on a 5-point scale. Fourteen questions addressed frequency of using specific skills on a scale of 1 = *no implementation*, 2 = *attempted implementation*, 3 = *appropriate implementation once or twice*, 4 = *frequent appropriate implementation*, and 5 = *frequent and effective implementation*. The remainder of the questions were more qualitative and were also rated on a 5-point scale of 1 = *not at all*, to 5 = *very much* (for example, “teacher seems familiar with the activity,” “teacher makes the activity fun,” “teacher listens attentively when children speak”). The items included preparation, delivery, use of teaching strategies, engaging children, managing children's behavior, using reinforcing activities, collaboration of the teaching team, and children's engagement and observed use of skills. Interrater reliability was assessed on 10–15% of all observations in all four years and results for these observations indicated substantial agreement within one point: 93% in Year 1, 94% in Year 2, 94% in Year 3, and 95% in Year 4.

Independently documented curriculum activities in intervention and control classrooms

Two different independently assessed measures of curriculum implementation were collected from both intervention and control classrooms in order to better document the extent of EF and SE activities in the two conditions. Classroom lesson plans for one week each month (October–May) were collected each year and coded by study RAs blind to condition and study hypotheses. Coders identified and counted any activity listed that related to either SE or EF skills (e.g., reading books about feelings or friendship, playing memory or matching games) and an average number of total activities per week was calculated (combining SE and EF). Interrater reliability was conducted on 20% of the plans each year and ranged from 81%–89%.

The second assessment, the Social/Emotional & Executive Functioning Activities Scale (SEEF), was a study-developed observational scale, conducted by RAs blind to study condition and hypotheses, designed to detect the rates of types of EF and SE activities observed in a 2-h visit, inclusive of a group activity and free play, to each classroom at midyear. The scale was developed in Year 2 and implemented with randomly chosen classrooms in Years 2, and the full sample in Year 4. The scale was based on the “Adapted Teacher Style Rating Scale” developed for the Head Start Cares study (Mattera, Lloyd, Fishman, & Bangser, 2013), and mirrored in general terms, the core EF and SE skills and instructional methods promoted in the SSEL curriculum. Thirteen items focused on teacher's formal or informal instruction in three areas: 1) SE skills (e.g., identifying feelings and perspective taking, understanding strong emotions and how to calm down, and friendship skills); 2) EF skills (ways to listen and focus attention, ways to remember and follow directions, and playing games like Simon Says to build memory, attention and impulse control); and 3) on methods for reinforcing children's learning, and improving engagement and self-regulation at group time (e.g., encouraging children to think back on skills learned previously when a social problem situation arises, and promoting self-regulation by encouraging children to think before responding in group). Items were rated from 1 = never observed to 5 = frequently observed (alpha reliability across items = 0.85). In addition, 4 items rated overall classroom conditions–children's attentiveness, disruptiveness, prosocial behavior, and emotion regulation (rated from 1 = very little to 5 = very much). In Year 2, 15 classrooms were observed (8 intervention and 7 control). Year 4, 13 intervention and 14 control classrooms were observed. Interrater reliability, scored as within one-point agreement on each item, was conducted on 26% of the observations and was 95%.

Classroom quality

The Classroom Assessment Scoring System Pre-K (CLASS Pre—K, Pianta, LaParo, & Hamre, 2008), is a widely used observation system for education research that taps three domains: Emotional Support; Classroom Organization; and Instructional Support. Alpha reliabilities for the preschool version (0.79 to 0.91), and test-retest reliability (0.73–0.85) are high. Each dimension is rated on a 7-point scale and averaged across 20-min observation cycles. Completing at least four observational cycles offers a representative sampling (Pianta et al., 2008). Study RAs blind to condition and study hypotheses were trained and certified to administer CLASS observations and conducted them in fall and spring in 5 control and 9 intervention classrooms in Year 2, and then in 13 intervention and 14 control classrooms in Year 4.

Observations were scheduled for two different days in one week per recommendation of Kane and Staiger (2012) to ensure reliability and validity, since variation in teacher attendance, composition of the classroom, or special activities on a particular day were felt to sometimes dramatically affect classroom climate. Observers completed three cycles per day, for a total of 6 cycles per classroom each fall and spring. Scores were then averaged across all the cycles for each classroom. Interrater reliabilities, scored as within one-point agreement on each item, were conducted for 20% of study observations and were 98.5% within one point across the two different years of observations.

Covariates

Demographics

Data on child and family characteristics were collected at enrollment from families. This included child's sex, age and ethnicity, and family marital status, income, number of children in the family, and maternal education level.

Cognitive ability

As part of the individual assessment given to 4 year-olds, the Peabody Picture Vocabulary Test, 4th edition, (PPVT-4, Dunn & Dunn, 2007) was administered. It consists of a presentation of 4 pictures, and the assessor asking the child to point to the picture representing a specific word or concept (e.g. pencil, swimming). Alpha and split-half reliability of the PPVT are above 0.90 and test-retest reliability is 0.93 (Dunn & Dunn, 2007). The PPVT measures receptive verbal ability and correlates highly with measures of general cognition (Gullo & McLoughlin, 1982; Taylor, 1979). Standard scores were used in analyses as a cognitive control variable.

Executive functioning skills

Head-Toes-Knees-Shoulders (HTKS)

The HTKS task (McClelland et al., 2007) measures several dimensions of EF, including attention, inhibitory control, and working memory. Children are asked to perform actions contrary to what the examiner performs in three sets of 10 trials of increasing complexity. First children are asked to touch their heads when the examiner says to touch toes and touch their toes when the examiner says to touch their heads. Then the second set of trials uses the same rules, but the examiner adds shoulders and knees–asking them to touch their shoulders when she says touch their knees, and vice versa. In the third set of 10 trials the command to touch your head is paired with touching knees, and touching toes is paired with touching shoulders and vice versa. After practice trials with repeated instructions for each of the three sets of commands, the scoring is 0, 1 or 2 for wrong, self-corrected, or correct responses. A total score across all trials (range 0–60) is used as an indication of EF skill. Alpha reliability for our sample was 0.84.

Backward Digit Span

This is a task of working memory (Davis & Pratt, 1996) and requires a child to recite numbers in backwards order from what is told to them. For our study children were shown a puppet and told the puppet likes to say things backwards. The examiner would say “1, 2” and then have the puppet say “2, 1”. Children were then told to say the numbers the examiner said but backwards. The trials start with two digits and progress to five digits with three attempts at each digit length. The highest level of correctly repeating a string of numbers backward is scored with a range of 1–5.

Social/emotional skills

Emotion Matching Task (EMT)

The short form of this scale (24 items) was administered to measure emotion knowledge. It was designed for use with low-income and ethnically diverse populations and correlates with other measures of emotion knowledge. The alpha reliability of the short version is 0.72–0.74 (Seidenfeld, Johnson, Cavadel, & Izard, 2014). It uses photographs of children's faces with various expressions showing happiness, sadness, anger, fear and surprise (Izard, Haskins, Schultz, Trentacosta, & King, 2003). Four different tasks are involved: 1) matching expressions of the same emotion; 2) matching emotions with situational cues, e.g. “show me the one who just got a nice toy”; 3) naming the emotion in different pictures; and 4) using emotion labels stated by the examiner to match to a picture of the same emotion. The total score was used in this study to indicate emotion knowledge.

Challenging Situations Task (CST)

Social problem-solving skills were assessed using the CST. Children are presented six drawings of a problem situation (e.g. a child knocking over someone's blocks, a child being hit, a ball being taken away etc.) and then after each problem situation, four drawings that show choices for how they would react to the situation that reflect a prosocial choice, avoidance, crying, or aggression. The examiner describes the situation in the first card and then shows each of the four choice cards, labels the choices and asks the child to say or point to what they would do in that situation. Prosocial responses in this task have been related to measures of emotion knowledge (Denham, Bouril, & Belouad, 1994), and have been shown to increase after a social/emotional skills intervention (Bierman, et al., 2008). The number of prosocial responses out of the six trials was used as one of the SE outcomes in this study.

Measures of preacademic skills

Four tests from the Woodcock-Johnson Tests of Achievement III (WJ III, Woodcock, McGrew, & Mather, 2001/2007) were used to measure preschool preacademic skills. Pre-math skills were measured using the Applied Problems subscale, and the Letter-Word Identification subscale measured pre-reading skills. Two additional subscales measured oral language skills: 1) Understanding Directions, measuring listening comprehension, and 2) Story Recall, measuring oral expression. Understanding Directions requires the child to listen to examiner directions and then wait until the end of the directions when the examiner says “go” to point to various details in a picture in the correct order, e.g., “Point to the truck, and then the bird. Go.” Story Recall requires the child to listen to a story and repeat it back to the examiner. Points are totaled for number of story elements articulated. The WJIII is widely used test of ability and achievement that has been normed on a large and diverse sample from age 2 to adults. The subtests chosen for this study were appropriate for age 4 and older, and measured key school readiness abilities (McGrew, Schrank, & Woodcock, 2007). Reported alpha reliability of these scales ranged from 0.99 for Letter-Word Identification, to 0.77 for Story Recall. Standard scores were used in analyses.

Data analysis

Each cohort of teachers was followed for two years in large part due to the hypothesis that implementation of the core components of the Second Step curriculum would be stronger in the second year of implementation. However, the two years of implementation create some methodological difficulties. Classrooms (and therefore teachers) were randomly assigned to condition prior to the first year of implementation for a given cohort, and students had already been assigned to classrooms at the time of randomization. Thus, randomization ensures that estimated average impacts on children measured in year 1 of a given cohort are unbiased. However, there is the possibility that the placement of children in classrooms during year 2 of a cohort was influenced by teachers' assignments to the treatment or control condition. Because of this possibility, we interpret the estimates we report here as average causal effects of assigning a teacher to implement SSEL on his/her

students' social emotional and executive functioning capabilities, acknowledging that part of that effect may be due to changes in the composition of the students attending the class.

Preliminary analyses

In order to check whether there may have been compositional difference between students in treatment and control classrooms, and also to understand whether there may have been any biasing effect of sample attrition, we examined differences between treatment and control students with respect to a variety of covariates measured at baseline (namely the four EF and SE outcome measures, the PPVT, and child and family demographics; see Appendix tables A and B for further detail). The attrition rates were very similar in the treatment and control group. Both groups had just over 15% of students with valid pretest data who were missing post-test data.

When making comparisons with respect to baseline differences we followed guidance issued by the Institute of Education Sciences' What Works Clearinghouse (U.S. Department of Education, 2017a; U.S. Department of Education, 2017b). This guidance suggests computing a standardized difference between the groups in effect size units. Differences > 0.25 units are considered large enough to undermine causal inferences. Differences between 0.05 and 0.25 units are acceptable provided that the relevant covariates are adjusted for in statistical models. Differences < 0.05 do not require adjustment (although adjustment is acceptable). Effect sizes for continuous covariates were computed using hierarchical linear models with students nested in classrooms. The estimated mean differences were divided by the estimated total standard deviation (the square root of the sum of the student and classroom level variance components). Effect sizes for dichotomous covariates were computed using two level logistic regression models. Following What Works Clearinghouse (WWC) guidance the estimated difference in logits was divided by 1.65 to make the scaling of effect sizes for dichotomous covariates similar to the scaling for continuous covariates.

Across the 16 demographic and pretest variables examined, the effect size estimates ranged from −0.17 to 0.17. 40% had absolute value < 0.05. None of the differences was statistically significant. Due to many effect size estimates > 0.05 we report treatment effect estimates from statistical models that include adjustments for all demographic and pretest variables. In addition to examining baseline equivalence, our preliminary analyses included the computation of individual-level bivariate correlations between the key SE and EF measures in the fall and the spring, and bivariate correlations between these measures, fall PPVT scores, and child and family demographic characteristics.

Additionally, we did analyses to understand if there were differences on pretest measures between students assessed during year 1 of a given cohort and those assessed during year 2 of a cohort, and to understand whether there was an interaction between this relation and treatment. In most cases there was no meaningful difference. We did find small differences favoring students assessed in the second year for the EMT and CST pretest measures. The effect size for the EMT differences was 0.26 with *p*-value of 0.03 and the effect size for the CST difference was 0.20 with a *p*-value of 0.08. However, none of the interaction terms were statistically significant, nor were the effect sizes large. Thus, these fluctuations in pretest scores across time do not pose a threat to the internal validity of the study, since they appear to operate equally across intervention and control groups.

Main outcome analyses

In order to estimate omnibus effects of SSEL on both EF and SE, our main analytic models were three level hierarchical linear models with measures nested within students nested within classrooms. Separate models were fit for EF and SE outcomes. Fitting these multivariate multi-level models (Hox, 2010; pp. 188–204) with measures at level 1 (so each child at level 2 has two level 1 measurements associated with him/her) is a somewhat novel approach to data analysis. We adopt it here because it allows us to obtain an omnibus treatment effect estimate (ES) in each domain of interest while avoiding the loss of statistical power associated with multiple comparison corrections. Specifically, dummy variables at level 1 of the model are included for each measure, but these variables are not interacted with the treatment indicator variable that enters the model at level 3. Thus, the model provides a single treatment effect estimate (and associated standard error and *p*-value) that inherently accounts for the correlation of measurements taken on the same student.

Additionally, the outcomes measures were standardized (converted to *z* scores) before analyses were run using the sample of 770 children with complete outcome, pretest and demographic data. This was done to assure that neither of the two outcome measures used to measure EF and SE, respectively, would unduly influence the fit of the model. Models were fit using the HLM software version 7.01 (Raudenbush, Bryk, & Congdon, 2013). These two multi-level models allowed us to test the two main hypotheses: 1) children in the intervention condition compared to children in the control condition, controlling for pretest baseline scores, baseline PPVT, and demographics, would show greater EF skills at the end of preschool (taking into account both EF measures); and 2) children in the intervention condition compared to children in the control condition, controlling for pretest baseline scores, baseline PPVT, and demographics, would show greater SE skills at the end of preschool (taking into account both SE measures). Statistical models included baseline measures of all four main outcome indicators, baseline PPVT, parental income (coded to the midpoint of each ordinal category), child gender, child age in months, number of children in household, a dummy code for whether the caretaker is married, ethnicity (dummy coded in to 4 categories as reported in Tables 5 and 6) and parental education (dummy coded in to 4 categories as reported in Tables 5 and 6), as well as dummy codes for year of participation and if the child had previously attended the site as a 3 year-old.

Results

Teacher training, Attendance, and Turnover

Teacher attendance was counted if at least one member of the teaching team in each intervention classroom attended. In Years 1 and 3 where 7 sessions were offered, the mean attendance was 6.6 (*SD* = 1.21, range 3–7) and 6.2 (*SD* = 0.45, range 4–7) respectively. In Years 2 and 4 where 5 group sessions were offered, mean attendance was 4.75 (*SD* = 1.07, range 4–5) and 4.0 (*SD* = 1.53, range 0–5) respectively. (Note in Year 4, one classroom team was unable to attend any of the evening sessions due to childcare and college class conflicts. However, they participated more than half the sessions in Year 3, and participated in all of the on-site observations and coaching sessions in both years). Overall 87.5% of the classrooms in Year 1and 81.25% in Year 2 were represented at all of the sessions, but only 52.9% were represented in Year 3, and 53.8% in Year 4. However, in year 4, due to substantial teacher turnover from the prior year, an extra two-hour training was provided for all new teachers and this was attended by 100% of all teacher teams.

Monthly observations and coaching sessions were conducted November–May in Years 1 and 3 and October–May in Years 2 and 4. Because these observation and coaching sessions were only conducted if a trained teacher in the classroom was present and delivered a specific SSEL lesson, there was 100% participation of classrooms in the coaching sessions, but we did not track if all members of the teaching team were always present the day the observation and in-person coaching was delivered. However, the written feedback was always addressed to all teaching team members.

While it was intended that teacher teams participate in the study for two years, turnover could not be prevented. In Year 2, 69.4% of trained intervention teachers remained from Year 1, and in Year 4 only 46.3%

of intervention teachers remained from Year 3. In addition, during the course of the school year we also found changes in composition of intervention classroom teacher teams, with 1 to 5 classrooms out of the sample each year affected by teachers leaving and being replaced during the course of the year. In order for a classroom to remain in the intervention group during or across years, at least one trained teacher had to be present and delivering the curriculum throughout the year, and this condition was met despite this turnover, except in Year 4 when 2 intervention classrooms were dropped because of lack of trained teachers.

Curriculum implementation and fidelity

Mean reported implementation of curriculum activities each week by teachers was 90.0% (*SD* = 9.0%, range = 67%–100%). Teachers reported playing the Brain Games an average of 3.9 days/week (*SD* = 0.72, range = 1.9–5.0). Most teachers reported making "occasional" modifications to the curriculum. The median reported extent of modifications was 1.4 (*SD* = 0.41, range = 0.90–2.5), or somewhat more than "occasionally." Across the four years of the study, 5 (8.5%) of the 59 intervention classrooms did not complete the 25 core weekly themes within the school year (2 of these completed 23 and 3 completed 24), while 35 (59.3%) completed all 28 weekly themes (including the last 3 kindergarten/review themes).

The mean total fidelity rating assigned by core staff during observations across all classrooms was 3.4 (*SD* = 0.51, range = 2.32–4.50). In Year 1 the mean fidelity rating was 3.61 (*SD* = 0.58); in Year 2 it was 3.44 (*SD* = 0.43); in Year 3 in was 3.11 (*SD* = 0.39); and in Year 4 it was 3.37 (*SD* = 0.52). About a quarter of the classrooms (*n* = 13; 22%) across the four years did not achieve an overall average score of 3.0, considered adequate implementation, but only 3 of these were below a 2.5, while 5 achieved an overall score of 2.9 or above. There were 8 classrooms with an overall mean score above 4.0.

Independently documented curriculum activities

Table 2 shows the results for the SEEF measure subscales and coded lesson plans between intervention and control classrooms in years 2 and 4 of the study. There were significant differences between intervention and control classrooms on the total number of both SE and EF-related activities per week coded on lesson plans (Means 11.44 and 3.54 respectively; $p < .001$; *ES* = 2.2). The SEEF observations also revealed significant differences ($p = .001$; *ES* = 1.02) in the total score of teacher-directed activities between intervention (*M* = 2.99, *SD* = 0.61) and control classrooms (*M* = 2.41, *SD* = 0.52). However, examining the subscales, those for Empathy and Friendship Skills were not significantly different, while Skills for Learning ($p = .01$), Emotion Management ($p = .009$), and Instructional Methods ($p < .001$) showed significant differences. Thus, the two scales measuring EF type activities both favored the intervention classrooms, but only one of the three subscales measuring SE activities, the Emotion Management subscale, favored the intervention classrooms. This subscale taps teaching of self-regulation skills when children are upset. These results seem quite consistent with our knowledge that control classrooms increasingly implemented SE activities due to changing state regulations.

Preliminary analyses of the SE and EF outcomes

Table 3 shows correlations among key SE and EF outcome measures and child and family demographics in both fall and spring and between fall and spring. Correlations between the same measure in fall and spring (e.g., fall CST Prosocial and spring CST Prosocial) were moderately strong (0.36–0.58), while within time point, the relation among the 4 main outcome measures was smaller (0.11–0.48), with the Backward Digit Span EF measure the least associated with the two SE measures (0.11 with CST both fall and spring, and 0.16 and 0.07 with the EMT fall and spring respectively), and more closely associated with the other EF measure, HTKS (0.36 in fall and 0.31 in spring). Correlations among the outcome measures and child and family demographics were small in the fall (0.01–0.12) except for child age and PPVT. In the spring, in addition to child age and PPVT, child race, parent education, and family income were somewhat more highly correlated with the outcomes, especially the two EF variables (0.11–0.19).

Table 4 shows means and SDs for the four outcome measures in fall and spring and the effect size (ES) at each time point, using a multilevel variant of Cohen's D (Hedges, 2007). There were small ES favoring intervention children in the fall, except for Backward Digit, but in the spring the ES increased considerably for the two EF measures. The ES for the two SE measures remained small and similar. These data help to contextualize understanding of the final models. While small, the differences at baseline are accounted for in the final models by including baseline scores as covariates.

Overall outcome analyses

Effects on executive functioning skills

Table 5 shows the results of the three-level model predicting end of preschool EF skills. It can be seen that condition significantly predicted EF skills ($\beta = 1.45$, $p < .001$), controlling for baseline SE and EF, baseline general cognition (PPVT), year of participation in the study, if the child had 2 years of preschool, and demographic variables. The effect size for EF is 0.15. While baseline SE and EF skills, baseline general cognition, child age, child ethnicity, and parent education also

Table 2
SEEF subscales and coded lesson plans: Group differences in year 2 of each cohort.

Subscale	Control	Intervention	*p* value	Effect Size
	M (*SD*)	*M* (*SD*)		
SEEF[a]				
Skills for learning	2.30 (0.51)	2.79 (0.68)	0.01	0.82
Empathy	2.50 (0.77)	2.71 (1.18)	0.48	0.21
Emotion management	2.22 (0.80)	2.98 (1.04)	0.01	0.82
Friendship skills & problem-solving	3.04 (0.98)	3.43 (1.08)	0.22	0.38
Instructional method	2.24 (0.71)	3.07 (0.60)	< 0.001	1.26
Overall mean score	2.41 (0.52)	2.99 (0.61)	0.001	1.02
Lesson Plans[b]				
Mean number of EF/SE[c] activities per week	3.54 (2.77)	11.44 (4.26)	< 0.001	2.2
Total sum of EF/SE activities across all coded weeks	28.35 (22.15)	91.52 (34.04)	< 0.001	2.2

[a] Social Emotional and Executive Functioning Scale. Sample includes *n* = 23 control classrooms and *n* = 21 intervention classrooms.
[b] Lesson Plan sample includes *n* = 31 control classrooms and *n* = 29 intervention classrooms.
[c] EF/SE = executive functioning or social emotional.

Table 3
Individual level correlations between key variables and child/family characteristics[a].

	CST-Pro[b]	EMT[c]	HTKS[d]	Back-ward digit	CST-Pro	EMT	HTKS	Back-ward digit	Condition[e]	Parental education	Family income	Fall PPVT	Child race[f]	Parent marital status[g]	Child age	Child sex[h]
	Fall	Fall	Fall	Fall	Spring	Spring	Spring	Spring								
Fall																
CST-Pro	–	–	–	–	0.34	0.11	0.22	0.21	0.04	0.07	0.03	0.21	0.01	0.01	0.19	< 0.01
EMT	0.19	–	–	–	0.17	0.48	0.35	0.23	0.06	0.02	0.06	0.41	< 0.01	0.02	0.29	0.08
HTKS	0.22	0.30	–	–	0.23	0.16	0.58	0.48	0.06	0.10	0.10	0.41	0.12	−0.05	0.29	0.04
Backward digit	0.09	0.16	0.39	–	0.11	0.06	0.30	0.45	0.01	0.08	0.07	0.23	0.09	−0.05	0.17	< −0.01
Spring																
CST-Pro	–	–	–	–	–	–	–	–	0.03	0.05	0.04	0.16	0.12	0.03	0.20	< 0.01
EMT	–	–	–	–	0.11	–	–	–	0.06	0.01	0.03	0.20	−0.06	0.05	0.15	0.09
HTKS	–	–	–	–	0.22	0.21	–	–	0.14	0.11	0.16	0.40	0.19	−0.02	0.32	0.04
Backward digit	–	–	–	–	0.23	0.11	0.49	–	0.12	0.19	0.14	0.37	0.16	< −0.01	0.26	0.02

[a] p-values are omitted due to their lack of validity in the context of individual level correlations with nested data.
[b] Challenging Situations Task-Prosocial.
[c] Emotion Matching Task.
[d] Head-Toes-Knees-Shoulders.
[e] 1 = treatment, 0 = control.
[f] 1 = Anglo-American, 0 = other.
[g] 1 = married, 0 = other.
[h] 1 = male, 2 = female.

Table 4
Descriptive statistics for EF and SE measures by intervention and control[a].

Measure	Whole sample	Control	Intervention	Effect size[b]
	M (*SD*)	*M* (*SD*)	*M* (*SD*)	
Fall				
CST-Pro[c]	2.31 (1.65)	2.24 (1.66)	2.37 (1.64)	0.07
EMT[d]	21.07 (5.00)	20.79 (5.13)	21.35 (4.87)	0.10
HTKS[e]	8.91 (13.19)	7.91 (12.40)	9.88 (13.86)	0.15
Backward digit	1.13 (0.42)	1.26 (0.57)	1.13 (0.41)	< 0.01
Spring				
CST-Pro	2.75 (1.69)	2.71 (1.69)	2.79 (1.70)	0.04
EMT	24.27 (3.53)	24.06(3.62)	24.48 (3.44)	0.12
HTKS	18.16 (17.03)	15.78 (16.37)	20.44(17.35)	0.25
Backward digit	1.33 (0.62)	1.26 (0.57)	1.41 (0.66)	0.21

[a] Whole sample: $n = 770$; Intervention: $n = 393$; Control: $n = 377$.
[b] The effect size was calculated by running multilevel models with each measure as the outcome and condition (intervention / control) as the only predictor. The effect size, a multi-level variant of Cohen's D, is equal to the regression coefficient for condition divided by the square root of the total variance computed from a two-level hierarchical linear model with condition as the only predictor as per Hedges (2007).
[c] Challenging Situations Task-Prosocial.
[d] Emotion Matching Task.
[e] Head-Toes-Knees-Shoulders Task.

were significantly associated with end of year EF skills, the model demonstrates that intervention children's spring EF skills were still significantly higher than children in the control condition.

Effects on social/emotional skills

Table 6 shows the results of the three-level model predicting end of preschool SE skills. In this model all the same covariates were utilized as for the EF model, and we found no significant effect of intervention condition on the end of preschool SE skills. The estimated effect size was very small, only 0.02. The only significant predictors in this model were baseline SE skills and child age. In contrast to the EF model, where baseline SE skills had a significant association with end of year EF skills, baseline EF skills were surprisingly not significantly associated with end of year SE skills.

Secondary hypotheses

Pre-academic outcomes

Table 7 shows the means and SDs for the pre-academic measures used in the study. It can be seen that for the Applied Problems, Understanding Directions, and Story Recall measures, all skills that should be associated with EF, intervention children had slightly higher mean scores in the spring while control children had slightly higher mean scores on Letter Word. Two level models run separately for each pre-academic outcome, but not adjusting for pre-test or covariates, found no significant differences between intervention and control children so no further analyses were conducted.

Classroom quality outcomes

A total of 41 different classrooms were observed during fall and spring in Years 2 ($n = 14$) and 4 ($n = 27$) using the CLASS measure. Repeated measures analyses of variance (ANOVA) were run with group as the independent variable and each of the CLASS subscales as the depending variable combining classrooms from the two years. Table 8 indicates the mean fall and spring and change scores by condition. The study was not well powered for this analysis, but we viewed exploration of potential effects on classroom quality as important. The effect size estimates demonstrate small to moderate improvements across the scales favoring the intervention classrooms, although none were statistically significant. It can be seen that on Emotional Support and

Table 5
Multilevel model predicting executive functioning skills at end of preschool.

Fixed effects	Estimate	Standard error
Intercept	−3.940***	0.361
Time 1 EF1 (Backward digit)[a]	0.419***	0.056
Time 1 EF2 (HTKS)[b]	0.023***	0.002
Time 1 SE1 (CST-Pro)[c]	0.033*	0.014
Time 1 SE2 (EMT)[d]	0.012*	0.005
Time 1 PPVT[e] (centered)	0.009***	0.002
Child sex[f]	−0.036	0.045
Child age Time 1	0.033***	0.006
Parent income (in $1000 s)	0.002	0.001
Education: high school[g]	0.131**	0.047
Education: some college[g]	0.174***	0.048
Education: college grad[g]	0.243**	0.076
Married[h]	0.007	0.043
No. of children in house	0.028	0.022
White[i]	0.130**	0.045
African-American[i]	−0.030	0.051
Hispanic- American[i]	−0.029	0.055
Assessed Y1 of cohort	0.060	0.052
Participated 2 years	0.063	0.057
Condition[j]	0.145***	0.045

	Covariance parameters (ICC)	Covariance parameters: Unconditional model (ICC)
Residual	0.516	0.516
Intercept (student)	0.087	0.375
Intercept (classroom)	0.003	0.090
Fit Statistics		
−2 Log Likelihood	3582.0	4116.9

[a] Executive Functioning skill 1: Backward Digit Span Task.
[b] Executive Functioning skill 2:Head-Toes-Knees-Shoulders Task.
[c] Social Emotional skill 1: Challenging Situations Task-Prosocial.
[d] Social Emotional skill 2: Emotion Matching Task.
[e] Peabody Picture Vocabulary Test Standard Score- representing general cognition.
[f] 1 = male, 0 = female.
[g] Education reference category = less than high school grad.
[h] 1 = married, 0 = not married.
[i] race reference category = Asian and other.
[j] 1 = intervention, 0 = control.
* $p < 0.05$.
** $p < 0.01$.
*** $p < 0.001$.

Table 6
Multilevel model predicting social/emotional skills at end of preschool.

Fixed effects	Estimate	Standard error
Intercept	−2.11***	0.402
Time 1 EF1 (Backward Digit)[a]	0.003	0.062
Time 1 EF2 (HTKS)[b]	0.004	0.002
Time 1 SE1 (CST-Pro)[c]	0.091***	0.015
Time 1 SE2 (EMT)[d]	0.049***	0.006
Time 1 PPVT[e] (centered)	0.002	0.002
Child sex[f]	−0.074	0.048
Child age Time 1	0.015**, *, ***	0.007
Parent income (in $1000 s)	−0.001	0.002
Education: high school[g]	0.073	0.080
Education: some college[g]	0.012	0.078
Education: college grad[g]	0.057	0.104
Married[h]	0.094	0.057
No. of children in house	0.002	0.022
White[i]	0.032	0.054
African-American[i]	0.022	0.063
Hispanic- American[i]	0.049	0.056
Assessed Y1 of cohort	−0.053	0.060
Participated 2 years	0.028	0.068
Condition[j]	0.021	0.050

	Covariance parameters (ICC)	Covariance parameters: Unconditional model (ICC)
Residual	0.845	0.895
Intercept (student)	0.002 (0.00)	0.098 (0.10)
Intercept (classroom)	0.006 (0.01)	0.003 (0.00)
Fit Statistics		
−2 Log Likelihood	4124.7	4356.8

[a] Executive Functioning skill 1: Backward Digit Span Task.
[b] Executive Functioning skill 2:Head-Toes-Knees-Shoulders Task.
[c] Social Emotional skill 1: Challenging Situations Task-Prosocial.
[d] Social Emotional skill 2: Emotion Matching Task.
[e] Peabody Picture Vocabulary Test Standard Score- representing general cognition.
[f] 1 = male, 0 = female.
[g] Education reference category = less than high school grad.
[h] 1 = married, 0 = not married.
[i] race reference category = Asian and other.
[j] 1 = intervention, 0 = control.
* $p < 0.05$.
** $p < 0.01$.
*** $p < 0.001$.

Classroom Organization, intervention classrooms were rated somewhat better at the beginning of the school year, while the control classrooms were rated slightly better on the Instructional Support scale. In the spring, however, intervention classrooms improved across all three subscales while control classrooms had slightly lower scores on each subscale than in fall.

Discussion

The purpose of the current study was to conduct a classroom randomized control efficacy trial of the SSEL curriculum in preschools serving low-income children, with a focus on investigating the primary impact on EF and SE skills, and potential secondary impact on pre-academic skills and classroom quality. We found, consistent with a prior analysis of this trial and with the curriculum logic model, using only the first two years of data (Upshur et al., 2017), a significant effect on intervention children's growth in EF skills, directly assessed by two different tasks, Backward Digit Span (Davis & Pratt, 1996), and Head-Toes-Knees-Shoulders (McClelland et al., 2007). There was no detected effect on children's SE skills, although a marginal trend was found in the prior sample (Upshur et al., 2017). We attribute the drop off of the SE outcome as potentially due to new state regulations that required all preschool programs to implement more SE-type learning activities (Massachusetts Department of Early Education and Care, 2015), and this was born out by classroom observations by independent observers unaware of study condition that found no significant differences between intervention and control classrooms in two subscales assessing teachers delivering instruction on emotion learning and social problem solving skills.

Table 7
Descriptive statistics for academic outcomes by intervention and control.

Measure	Whole sample	Control	Intervention
	M (SD)	M (SD)	M (SD)
Fall			
Letter Word	95.77(12.79)	95.82(12.00)	95.73(13.52)
Applied Problems	100.87(11.25)	100.52(10.39)	101.21(12.03)
Understanding Directions	88.53(17.52)	88.08(17.72)	88.96(17.34)
Story Recall	86.60(17.87)	85.90(17.50)	87.27(18.22)
Spring			
Letter Word	96.48(13.00)	97.28(11.91)	95.71(13.93)
Applied Problems	101.84(11.33)	101.19(10.65)	102.46(11.92)
Understanding Directions	88.06(17.05)	86.77(17.09)	89.29(16.94)
Story Recall	91.79(20.06)	90.90(19.81)	92.65(20.29)

Note: Whole sample: $n = 770$; Control: $n = 377$; Intervention: $n = 393$.

Table 8
Change in CLASS subscales by condition.

CLASS subscale	Fall mean (*SD*)	Spring mean (*SD*)	Change score mean (*SD*)	Change score Cohen's *d*
Emotional support				0.44
Intervention	5.98 (0.90)	6.19 (0.68)	0.21 (0.70)	
Control	5.94 (0.78)	5.81 (1.04)	−0.13 (0.84)	
Classroom organization			0.20 (0.65)	0.27
Intervention	5.55 (0.93)	5.75 (0.67)	−0.04 (1.11)	
Control	5.49 (0.89)	5.45 (1.09)		
Instructional support				0.38
Intervention	2.66 (0.94)	3.10 (1.01)	0.44 (1.09)	
Control	2.87 (1.11)	2.85 (1.15)	0.02 (1.34)	

Note: Whole sample: $n = 41$; Intervention: $n = 22$; Control: $n = 19$.

The SSEL curriculum is one of several universal preschool interventions with similar goals to improve the self-regulation of at-risk preschool children in order to impact classroom behavior and preacademic skills, and subsequent kindergarten and school readiness. Most of the interventions have measured both SE and EF skills, but there is a mixed pattern of results. Interventions like the PATHS curriculum, (Domitrovich et al., 2007; Bierman, Nix, et al., 2008; Bierman, et al., 2008) have found more consistent significant effects on children's directly assessed emotion recognition, and social problem solving (*ES* = 0.28–0.40, as well as parent, observer, or teacher rated social skills (*ES* = 0.26–0.50), while no or only marginal effects on EF skills. Similarly, evaluations of the Tools of the Mind curriculum in preschool children has found early positive effects on directly measured EF skills (Diamond et al., 2007) but no effects in more recent cluster-randomized studies with larger samples (Solomon et al., 2018). However, a cluster randomized kindergarten study of Tools of the Mind found a significant effect (*ES* = 0.14) on a directly assessed measure of EF (Backward Digit Span) used in the current study, as well as other significant effects on reaction time and processing speed for EF tasks (*ES* = 0.08–0.12), as well as significant effects on overall math outcomes (*ES* = 0.13), effect sizes in line with the current study (Blair & Raver, 2014). In contrast, the CSRP intervention focused on improvements in teacher skills and classroom quality, found larger significant effects on two directly assessed tasks of executive functioning (*ES* = 0.37), as well as significant direct gains in multiple areas of academic skills, including early literacy, math, and vocabulary (Raver et al., 2011).

Although not a large body of evidence, the core relevance of EF skills such as working memory, inhibition, and attention, to both emotional self-regulation and cognitive development seems to be emerging as a powerful link to children's early school success. For example, Bierman and Nix, et al., 2008 found EF to both moderate and mediate gains in literacy skills as well as social/emotional skills (Bierman, Nix, et al., 2008). EF similarly mediated intervention effects on math, literacy and task behavior, which in turn predicted kindergarten screening scores, in another paper published on the current study sample (Wenz-Gross et al., 2018). These studies point to underlying developmental processes involved with EF as key dimensions required for learning. As noted by Blair and Raver (2015), development of self-regulation (encompassing both EF and SE as measured in intervention and laboratory studies) is a complex neurobiological process that is dependent on both initial individual differences and the impact of environmental exposures, including the quality of social contexts provided by parents and classrooms. They advocate that improving preschool education and school readiness requires promotion of self-regulatory processes in the classroom that will support both individual and group learning, so that activities directed at developing specific academic skills are maximally successful. The SSEL curriculum findings so far suggest it can provide one such approach.

Disappointingly, we found no direct impact of the SSEL curriculum on children's growth in preacademic skills, although this was not explored as a primary hypothesis since the curriculum logic model indicated school success as a long-term outcome. However, using the same sample, an earlier manuscript explored the impact of SSEL participation on kindergarten readiness scores (administered at entry to kindergarten by public schools several months after SSEL participation; Wenz-Gross et al., 2018). While no direct effect of SSEL participation or end of year EF skills on kindergarten entry scores was found, SEM models showed end of year EF skills predicated end of year pre-math, pre-literacy and on-task behavior skills which in turn predicted kindergarten entry scores. Thus, SSEL appears to successfully impact attention, working memory, inhibition, and on-task behavior that promote academic learning consistent with the SSEL logic model. Parallel with other classroom intervention studies (Raver et al., 2011; Sasser et al., 2015; Sasser et al., 2017, this suggests that underlying regulatory skills are important developmental processes that need to be fostered in addition to direct instruction on pre-academics. In addition, while Wenz-Gross et al. (2018) did not find that SSEL directly impacted gains in SE, consistent with the current study, SE was related to on-task behavior and was directly and indirectly (through pre-math skills) associated with kindergarten screening scores. In sum, SSEL as well as other classroom interventions that address children's emotional control, social skills, and executive functioning appear to improve children's learning and school performance through multiple pathways.

Other early childhood classroom interventions have also addressed the overall classroom environment and teacher skills in order to improve children's behavior and learning. While CSRP and IY (Raver et al., 2009; Webster-Stratton et al., 2008) directly addressed and measured teacher skills and classroom environment, the SSEL logic model does not specify classroom quality outcomes. Nevertheless, in a prior study of SSEL's precursor curriculum, significant differences in quality were found for intervention classrooms (Upshur et al., 2013), thus we chose to explore the secondary impacts on classroom quality. While unfortunately underpowered for this purpose, we did find small to moderately strong effect sizes (although not statistically significant) supporting the effect of the curriculum on higher quality classroom dimensions using a well-developed, independently administered, classroom observation tool (CLASS, Pre—K, Pianta et al., 2008). Our *ES* differences (0.27–0.44) were not as strong as some other preschool interventions designed to improve classroom quality, however, our teacher training hours (14 h in Years 1/3 and 10 h in Years 2/4) and in-classroom support (approximately an hour per month) were more modest, and therefore more sustainable under conditions of wider dissemination. For example, the CSRP intervention included 30 h of teacher workshops, weekly in-classroom mental health consultant coaching averaging 4.5 h, teacher stress-reduction workshops, and 10 weeks of individual mental health consultant intervention to 3–4 children in each class with higher rates of behavioral issues (Raver et al., 2008; Raver et al., 2009). The *ES* for four subscales of an earlier version of the CLASS assessment measure used in the CSRP study ranged from 0.53–0.89 (Raver et al., 2008). Similarly, the Incredible Years/Dinosaur intervention provided 28 h of group teacher training, and a study staff member to co-lead all the classroom lessons with

teachers. Coding of teacher sensitivity and classroom management behavior conducted by independent observers for the Head Start teachers in the IY study found *ES* from 0.51 for warmth, to 1.24 for effective discipline (Webster-Stratton et al., 2008). This group of studies show that improving teacher skills plays a role in improving child outcomes and suggest that formal cost-effectiveness studies are warranted to tease out the level of optimal teacher training and implementation support required to boost children's skills.

The SSEL intervention also meets criteria described by McClelland, Tominey, Schmitt, and Duncan (2017) for successful social/emotional learning interventions (SEL) in preschool. While SSEL was not one of the reviewed interventions due to the dearth of published studies so far, it includes the three components of the various successful programs described: 1) a teacher training component to build teacher skills to model appropriate behavior; 2) direct instructional activities of increasing difficulty built into daily activities; and 3) engagement of children's families. While the current study did not collect data on family participation, this is an important area for future study since the curriculum kits do provide extensive materials to engage families. McClelland et al. (2017) also indicate that the cost of materials, time needed for teacher development and feasibility of implementation of SEL interventions need to be considered in terms of overall effectiveness. We feel the SSEL curriculum has potential for further dissemination due to the reasonable costs, ease of implementation, and modest teacher training and supervision burden to achieve adequate fidelity, which in turn seem to produce meaningful changes in children's EF skills.

The current study was designed to estimate the overall efficacy of the SSEL curriculum in meeting its designed goals to improve at-risk preschool children's EF and SE skills as proximal outcomes in order to improve long-term school readiness. As pointed out by many in the field today, it is important to go beyond overall group outcomes and understand for whom the curriculum works, and why, and match interventions to the particular needs of subgroups of children (Schindler, et al., 2017; Shonkoff, 2017) It was beyond the scope of the current manuscript to explore subgroup differences in outcomes or delve further into the relation of various levels of implementation, classroom, or teacher characteristics, and the association with outcomes. Nevertheless, potential for widespread and effective implementation of interventions designed to impact early childhood self-regulatory skills depend on clear understanding of which groups of children will benefit most, and on determining requisite levels of teacher training, support, and fidelity, and future examinations of these dimensions with the SSEL curriculum should be undertaken. For example, our finding, that parent education and child ethnicity were associated with the EF outcomes in addition to intervention condition warrant further study. In addition, while teacher's in the study did not report that implementation was overly burdensome, and the vast majority felt it improved their ability to meet state standards, a better understanding of how teachers addressed specific preliteracy and math skills in the context of delivering SSEL might be useful given intervention children did not perform better on these skills as was hypothesized.

Limitations of the study include the issues of potential selection bias of children into classrooms already assigned to intervention and control condition in the second year of each cohort (Years 2 and 4). We have chosen to include children from all four years of the study and believe our statistical methods adequately control for any potential selection bias. Further, unlike some other efficacy trials, we provided more modest levels of teacher coaching (e.g., monthly versus weekly), which might have affected strength of implementation and therefore outcomes. While our measures of curriculum fidelity showed adequate to strong delivery in almost all classrooms, more frequent in-classroom coaching might have further strengthened teacher delivery, especially in classrooms where there was turnover. However, the current study's level of teacher support might be more feasible in larger scale dissemination, suggesting that outcomes could be expected from SSEL even in more routine use. In addition, due to secular trends in state early childhood standards during the study, the control classrooms began to implement additional SE and EF activities that reduced the potential differences in exposure to related activities between intervention and control children. Finally, due to the logistics of collecting extensive individual child assessments, outcomes represent the effects of receiving a majority, but not all the SSEL weekly lessons. While this was controlled for by assessing control children on the same schedules, it is possible that there would be stronger outcomes if pretest and posttest assessments were limited to time periods before and after the entire curriculum was delivered.

However, the study has several important strengths. Classrooms were randomized within site, controlling for considerable site-based differences in classroom schedules, teacher education, salaries, and teacher supports beyond the intervention training. Notable is that it included both Head Start and community-based preschool classrooms. While all the programs served largely low-income, at-risk children, including children who were homeless and in protective services, the community preschools typically had lower teacher salaries, fewer teachers with college degrees, and fewer overall supports (such as family advocates and nurses) that were available in all the Head Start sites. Despite these resource differences we found no differences in implementation rates or fidelity by program type suggesting SSEL is a viable option for introduction in a wide range of preschool programs. In addition, while some studies primarily use teacher and/or parent ratings of children's behavior and social competence (e.g., Domitrovich et al., 2007; Webster-Stratton et al., 2008), the current study used only directly assessed child measures, conducted by trained RAs blind to study condition. We also undertook two approaches to documenting the type of SE and EF curricular activities being implemented in the control classrooms in order to verify differences between intervention and control classrooms and, indeed, found that the significant outcomes for difference in children's EF skills seems to match independently observed differential levels of EF activities in the two conditions. Finally, while planned as an efficacy and not a dissemination study, only a modest level of teacher training support was provided, and there was considerable teacher turnover as is typical in community-based preschool programs. Under these normative conditions, the SSEL curriculum seems capable of producing important outcomes.

In summary, we conclude that the SSEL curriculum demonstrated significant effects on improving low-income children's EF, which a growing literature on early childhood development suggests is central to long term educational and functional success. Effects on SE remain to be further investigated under conditions where the control classrooms are receiving less robust SE instruction. While we did not find direct effects on pre-academic outcomes, the exploratory data showing moderate improvements in overall classroom emotional support and instructional support have the potential to impact such outcomes in preschool programs that are of lower quality than those enrolled in this study. Further, the SSEL curriculum has the strength of providing materials at a reasonable cost and instructional specificity that can be easily disseminated and replicated across both Head Start and community-based preschool programs.

Acknowledgements

This study was funded by the US Department of Education, Institute of Education Sciences (#R305A130336). Study design, data collection, analyses, interpretation and manuscript preparation were solely the responsibility of the authors. Study data have been deposited with the Inter-University Consortium for Political and Social Research (ICPSR; Project #103861) for public use after June 30, 2019. The authors wish to thank the staff and families of the participating preschool programs: Guild of St. Agnes Early Education and Child Care, Montachusett Opportunity Council Child Care and Head Start, Rainbow Child Development Center, Worcester Comprehensive Education and Care,

Worcester Public Schools/Child Development-Head Start, and the YWCA of Central Massachusetts. We also would like to thank our data management staff, Bruce Barton, PhD, Aimee Kroll Derosiers, MS, and Elizabeth Orvek, MS; the over 30 part-time study staff who conducted child assessments and classroom observations; and our administrative support team and post-doctoral fellow for years of dedication in keeping a complex and multifaceted study running smoothly: Lorna Chiasson, Jennifer Hazelton, Elizabeth DelaRosa Lawson, and Anthoula Poulakos, PhD. Finally, we have received steadfast and thoughtful advice and support throughout the study from our Scientific Advisory Committee, Estaban Cardimil, PhD, Associate Professor, Department of Psychology, Clark University; Stephanie M. Jones, PhD., Professor, Harvard University Graduate School of Education; and Daniel Shaw, PhD., Distinguished Professor, Department of Psychology, University of Pittsburg.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.appdev.2019.02.008.

References

Ackerman, B. P., Brown, E. D., & Izard, C. E. (2004). The relations between contextual risk, earned income, and the school adjustment of children from economically disadvantaged families. *Developmental Psychology, 40*(2), 204–216. https://doi.org/10.1037/0012-1649.40.2.204.

Bierman, K. L., Domitrovich, C. E., Nix, R. L., Gest, S. D., Welsh, J., Greenberg, M. T., & … & Gill, S. (2008). Promoting academic and social-emotional school readiness: The head start REDI program. *Child Development, 79*(6), 1802–1817. https://doi.org/10.1111/j.1467-8624.2008.01227.x.

Bierman, K. L., Nix, R. L., Greenberg, M. T., Blair, C., & Domitrovich, C. E. (2008). Executive functions and school readiness intervention: Impact, moderation, and mediation in the head start REDI program. *Development and Psychopathology, 20*(3), 821–843. https://doi.org/10.1017/S0954579408000394.

Bierman, K. L., & Torres, M. (2016). Promoting the development of executive functions through early education and prevention programs. In J. A. Griffin, P. McCardle, & L. S. Freund (Eds.). *Executive function in preschool age children: Integrating measurement, neurodevelopment, and translational research.* American Psychological Association: Washington DC.

Blair, C. (2002). School readiness: Integrating cognition and emotion in a neurobiological conceptualization of children's functioning at school entry. *The American Psychologist, 57*(2), 111–127.

Blair, C., & Diamond, A. (2008). Biological processes in prevention and intervention: The promotion of self-regulation as a means of preventing school failure. *Development and Psychopathology, 20*(3), 899–911. https://doi.org/10.1017/S0954579408000436.

Blair, C., & Raver, C. C. (2014). Closing the achievement gap through modification of neurocognitive and neuroendocrine function: Results from a cluster randomized controlled trial of an innovative approach to the education of children in kindergarten. *PLoS One, 9*, e112393. https://doi.org/10.1371/journal.pone.011239.

Blair, C., & Raver, C. C. (2015). School readiness and self-regulation: A developmental psychobiological approach. *Annual Review of Psychology, 66*, 711–713. https://doi.org/10.1146/annurev-psych-010814-015221.

Boyce, W. T. (2016). Differential susceptibility of the developing brain to contextual adversity and stress. *Neuropshopharmacology, 41*, 141–162. https://doi.org/10.1038/npp.2015.294.

Bruce, J., Gunnar, M. R., Pears, K. C., & Fisher, P. A. (2013). Early adverse care, stress neurobiology, and prevention science.: Lessons learned. *Prevention Science, 14*, 247–256. https://doi.org/10.1007/0003-006X.32.7.513.

Campbell, S. B., Shaw, D. S., & Gilliom, M. (2000). Early externalizing behavior problems: Toddlers and preschoolers at risk for later maladjustment. *Development and Psychopathology, 12*(3), 467–488.

Carlo, G., Knight, G., Eisenberg, N., & Rotenberg, K. J. (1991). Cognitive processes and prosocial behaviors among children: The role of affective attributions and reconciliations. *Developmental Psychology, 27*(3), 456–461. https://doi.org/10.1037/0012-1649.27.3.456.

Committee for Children (2011a). *Second step early learning.* Seattle, WA: Committee for Children.

Committee for Children (2011b). *Review of research-second step social-emotional skills for early learning.* Seattle, WA: Committee for Children.

Crick, N. R., & Dodge, K. (1994). A review and reformulation of social information-processing mechanisms in children's social adjustment. *Psychological Bulletin, 115*(1), 74–101. https://doi.org/10.1037/0033-2909.115.1.7.

Davis, H. L., & Pratt, C. (1996). The development of children's theory of mind: The working memory explanation. *Australian Journal of Psychology, 47*, 25–31. https://doi.org/10.1080/00049539508258765.

Denham, S. A., Bassett, H., Mincic, M., Kalb, S., Way, E., Wyatt, T., & Segal, S. (2012). Social-emotional learning profiles of preschoolers' early school success: A person-centered approach. *Learning and Individual Differences, 22*(2), 178–189. https://doi.org/10.1016/j.lindif.2011.05.001.

Denham, S. A., Bassett, H. H., Zinsser, K., & Wyatt, T. M. (2014). How preschoolers' social-emotional learning predicts their early school success: Developing theory-promoting, competency-based assessments. *Infant and Child Development, 23*(4), 426–454. https://doi.org/10.1002/icd.1840.

Denham, S. A., Bouril, B., & Belouad, F. (1994). Preschoolers' affect and cognition about challenging peer situations. *Child Study Journal, 24*(1), 1–21.

Denham, S. A., Wyatt, T. M., Bassett, H. H., Echeverria, D., & Knox, S. S. (2009). Assessing social-emotional development in children from a longitudinal perspective. *Journal of Epidemiology and Community Health, 63*, i37–i52. https://doi.org/10.1136/jech.2007.070797.

Diamond, A., Barnett, W. S., Thomas, J., & Munro, S. (2007). Preschool program improves cognitive control. *Science, 318*(5855), 1387–1388. https://doi.org/10.1126/science.1151148.

Domitrovich, C. E., Cortes, R. C., & Greenberg, M. T. (2007). Improving young children's social and emotional competence: A randomized trial of the preschool "PATHS" curriculum. *The Journal of Primary Prevention, 28*(2), 67–91. https://doi.org/10.1007/s10935-007-0081-0.

Duncan, C. J., & Magnuson, K. (2013). Investing in preschool programs. *Journal of Economic Perspectives, 27*, 109–132. https://doi.org/10.1257/jep.27.2.109.

Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., … Japel, C. (2007). School readiness and later achievement. *Developmental Psychology, 43*(6), 1428–1446. https://doi.org/10.1037/0012-1649.43.6.1428.

Dunn, M., & Dunn, L. M. (2007). *Peabody picture vocabulary test – 4.* Circle Pines, MN: AGS.

Eisenberg, N., Spinrad, T. L., & Eggum, N. D. (2010). Emotion-related self-regulation and its relation to children's maladjustment. *Annual Review of Clinical Psychology, 6*, 495–525. https://doi.org/10.1146/annurev.clinpsy.121208.131208.

Fox, N. A., Calkins, S. D., & Bell, M. A. (1994). Neural plasticity and development in the first two years of life: Evidence from cognitive and socioemotional domains of research. *Development and Psychopathology, 6*, 677–696.

Gullo, D. F., & McLoughlin, C. S. (1982). Comparison of scores for normal preschool children on Peabody picture vocabulary test- revised and McCarthy scales of Children's abilities. *Psychological Reports, 51*(2), 623–626. https://doi.org/10.2466/pr0.1982.51.2.623.

Hedges, L. (2007). Effect sizes in cluster-randomized designs. *Journal of Educational and Behavioral Statistics, 32*, 341–370. https://doi.org/10.3102/1076998606298043.

Hox, J. (2010). *Multilevel analysis: Techniques and applications.* New York: Routledge.

Hughes, C., & Ensor, R. (2007). Positive and protective effects of early theory of mind on problem behaviors in at-risk preschoolers. *Journal of Child Psychology and Psychiatry, 48*, 1025–1032. https://doi.org/10.1111/j.1469-7610.2007.01806.x.

Hughes, C., & Ensor, R. (2011). Individual differences in growth in executive function across the transition to school predict externalizing and internalizing behaviors and self-perceived academic success at 6 years of age. *Journal of Experimental Child Psychiatry, 108*, 663–676. https://doi.org/10.1016/j.jecp.2010.06.005.

Izard, C. E., Haskins, F. W., Schultz, D., Trentacosta, C. J., & King, K. A. (2003). *Emotion matching test.* Delaware, USA: Human Emotions Lab, Department of Psychology, University of Delaware.

Kaiser, A. P., Hancock, T. B., Cai, X., Foster, E. M., & Hester, P. P. (2000). Parent-reported behavioral problems and language delays in boys and girls enrolled in head start. *Behavioral Disorders, 26*, 26–41.

Kane, T. J., & Staiger, D. O. (2012). *Gathering feedback for teaching: Combining high-quality observations with student surveys and achievement gains.* Washington, DC: Bill & Melinda Gates Foundation.

Liu, J. (2004). Childhood externalizing behavior: Theory and implications. *Journal of Child and Adolescent Psychiatric Nursing, 17*(3), 93–103.

Massachusetts Department of Early Education and Care (June 2015). Massachusetts Standards for Preschool and Kindergarten Social and Emotional Learning, and Approaches to Play and Learning. Retrieved 11/15/2017 from http://www.doe.mass.edu/kindergarten/SEL-APL-Standards.pdf

Mattera, S. K., Lloyd, C. M., Fishman, M., & Bangser, M. A. (2013). A first look at the head start cares demonstration: Large-scale implementation of program to improve children's social-emotional competence. OPRE report 2013-47 http://www.acf.hhs.gov/sites/default/files/opre/head_start_cares_implementation_acf_execusumm.pdf, Accessed date: 15 November 2016.

McClelland, M. M., Acock, A. C., Piccinin, A., Rhea, S. A., & Stallings, M. C. (2013). Relations between preschool attention span-persistence and age 25 educational outcomes. *Early Childhood Research Quarterly, 28*(2), 314–324. https://doi.org/10.1016/j.ecresq.2012.07.008.

McClelland, M. M., Cameron, C. E., Connor, C. M., Farris, C. L., Jewkes, A. M., & Morrison, F. J. (2007). Links between behavioral regulation and preschoolers' literacy, vocabulary, and math skills. *Developmental Psychology, 43*(4), 947–959. https://doi.org/10.1037/0012-1649.43.4.947.

McGrew, K. S., Schrank, F. A., & Woodcock, R. W. (2007). *Woodcock-Johnson III Normative Update.* Rolling Meadows, IL: Riverside Publishing.

McMahon, S. D., Washburn, J., Felix, E. D., Yakin, J., & Childrey, G. (2000). Violence prevention program effects on urban preschool and kindergarten children. *Applied and Preventive Psychology, 9*, 271–281. https://doi.org/10.1016/S0962-1849(00)80004-9.

Morrison, F. J., Ponitz, C. C., & McClelland, M. M. (2010). Self-regulation and academic achievement in the transition to school. In S. D. Calkins, & M. A. Bell (Eds.). *Child development at the intersection of emotion and cognition* (pp. 203–224). Washington, DC, US: American Psychological Association. https://doi.org/10.1037/12059-011.

National Scientific Council on the Developing Child (2004). Young children develop in an environment of relationships. Working paper No. 1. Retrieved 3/1/18 from https://developingchild.harvard.edu/resources/wp1/.

Nesbitt, K. T., Farran, D. C., & Fuhs, M. W. (2015). Executive function skills and academic achievement gains in prekindergarten: Contributions of learning-related behaviors. *Developmental Psychology, 51*(7), 865–878. https://doi.org/10.1037/dev0000021.

Nix, R. J., Bierman, K. L., Heinrichs, B. S., Gest, S. D., Welsh, J. A., & Domitrovich, C. E. (2016). The randomized controlled trial of head start REDI: Sustained effects on developmental trajectories of social-emotional functioning. *Journal of Counseling and Clinical Psychology, 84*, 310–322. https://doi.org/10.1037/a0039937.

Pianta, R. C., LaParo, K. M., & Hamre, B. K. (2008). *Classroom assessment scoring system manual: Pre-K.* Baltimore, MD: Brookes.

Raudenbush, S. W., Bryk, A. S., & Congdon, R. (2013). *HLM 7.01 for windows [computer software].* Skokie, IL: Scientific Software International, Inc.

Raver, C. C. (2004). Placing emotional self-regulation in sociocultural and socioeconomic contexts. *Child Development, 75*, 346–353.

Raver, C. C., Jones, J. M., Li-Grining, C. P., Metzger, M., Champion, K. M., & Sardin, L. (2008). Improving preschool classroom processes: Preliminary findings from a randomized trial implemented in head start settings. *Early Childhood Research Quarterly, 23*, 10–26. https://doi.org/10.1016/j.ecresq.2007.09.001.

Raver, C. C., Jones, S. M., Li-Grining, C., Zhai, F., Bub, K., & Pressler, E. (2011). CSRP's impact on low-income preschoolers' preacademic skills: Self-regulation as a mediating mechanism. *Child Development, 82*(1), 362–378. https://doi.org/10.1111/j.1467-8624.2010.01561.x.

Raver, C. C., Jones, S. M., Li-Grining, C., Zhai, F., Metzer, M., & Solomon, B. (2009). Targeting children's behavior problems in preschool classrooms: A cluster-randomized controlled trial. *Journal of Consulting and Clinical Psychology, 77*, 302–316. https://doi.org/10.1037/a0015302.

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*(2), 182–191. https://doi.org/10.1016/j.ecresq.2010.07.003.

Sasser, T. R., Bierman, K. L., & Heinrichs, B. (2015). Executive function and school adjustment: The mediational role of pre-kindergarten learning-related behaviors. *Early Childhood Research Quarterly, 30*, 70–79. https://doi.org/10.1016/j.ecresq.2014.09.001.

Sasser, T. R., Bierman, K. L., Heinrichs, B., & Nix, R. L. (2017). Preschool intervention can promote sustained growth in the executive-function skills of children exhibiting early deficits. *Psychological Science, 28*, 1719–1730. https://doi.org/10.1177/0956797617711640.

Schindler, H. S., Fisher, P. A., & Shonkoff, J. P. (2017). From innovation to impact at scale: Lessons learned from a cluster of research-community partnerships. *Child Development, 88*, 1435–1446. https://doi.org/10.1111/cdev.12904.

Schultz, D., Izard, C. E., Ackerman, B. P., & Youngstrom, E. A. (2001). Emotion knowledge in economically disadvantaged children: Self-regulatory antecedents and relations to social difficulties and withdrawal. *Development and Psychopathology, 13*(01), 53–67. https://doi.org/10.1017/S0954579401001043.

Seidenfeld, A. M., Johnson, S. R., Cavadel, E. W., & Izard, C. E. (2014). Theory of mind predicts emotion knowledge development in head start children. *Early Education and Development, 25*(7), 933–948. https://doi.org/10.1080/10409289.2014.883587.

Sektnan, M., McClelland, M. M., Acock, A., & Morrison, F. J. (2010). Relations between early family risk, children's behavioral regulation, and academic achievement. *Early Childhood Research Quarterly, 25*(4), 464–479. https://doi.org/10.1016/j.ecresq.2010.02.005.

Shonkoff, J. P. (2017). Rethinking the definition of evidence-based interventions to promote early childhood development. *Pediatrics, 140*(6), e29173136.

Solomon, T., Plamondon, A., O'Hara, A., Finch, H., Goco, G., Chaban, P., … …Tannock, R. (2018). A cluster randomized-controlled trail of the impact of the *Tools of the Mind* curriculum on self-regulation in Canadian preschoolers. *Frontiers in Psychology, 8*, 2366. https://doi.org/10.3389//fpsyg.2017.02366.

Takesian, A. E., & Hensch, T. K. (2013). Balancing plasticity/stability across brain development. *Progress in Brain Research, 207*, 3–34. https://doi.org/10.1016/B978-0-444-63327-9.00001-1.

Taylor, R. L. (1979). Comparison of the McCarthy scales of Children's abilities and the Peabody picture vocabulary test. *Psychological Reports, 45*(1), 196–198. https://doi.org/10.2466/pr0.1979.45.1.196.

Teaching Strategies, LLC (2002 − 2012). The Creative Curriculum®for Preschool. Retrieved 2/17/17 from http://www.teachingstrategies.com/.

Tominey, S. L., & McClelland, M. M. (2011). Red light, purple light: Findings from a randomized trial using circle time games to improve behavioral self-regulation in preschool. *Early Education and Development, 22*, 489–519. https://doi.org/10.1080/10409289.2011.574258.

McClelland, M.M, Tominey, S.L., Schmitt, S.A., & Duncan, R. (2017). SEL interventions in early childhood. The Future of Children*,* 27*,* 33–47. www.futureofchildren.org

Torres, M. M., Domitrovich, C. E., & Bierman, K. L. (2015). Preschool interpersonal relationships predict kindergarten achievement: Mediated by gains in emotion knowledge. *Journal of Applied Developmental Psychology, 39*, 44–52. https://doi.org/10.1016/j.appdev.2015.04.008.

U. S. Department of Health and Human Services, Administration for Children and Families, Office of Head Start (2011). The Head Start Child Development And Early Learning Framework: Promoting Positive Outcomes In Early Childhood Programs Serving Children 3–5 Years Old*.* Retrieved 2/10/17 from http://eclkc.ohs.acf.hhs.gov/hslc/tta-system/teaching/eecd/Assessment/Child%20Outcomes/HS_Revised_Child_Outcomes_Framework(rev-Sept2011).pdf.

U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, What Works Clearinghouse (2017a). Procedures Handbook, version 4.0. Retrieved 10/12/17 from https://ies.ed.gov/ncee/wwc/Docs/referenceresources/wwc_procedures_handbook_v4.pdf.

U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, What Works Clearinghouse (2017b). Standards Handbook, version 4.0. Retrieved 10/12/17 from https://ies.ed.gov/ncee/wwc/Docs/referenceresources/wwc_standards_handbook_v4.pdf.

Upshur, C. C., Heyman, M., & Wenz-Gross (2017). Efficacy trial of the second step early learning (SSEL) curriculum: Preliminary outcomes. *Journal of Applied Developmental Psychology, 50*, 15–25. https://doi.org/10.1016/j.appdev.2017.03.004.

Upshur, C. C., Wenz-Gross, M., & Reed, G. (2013). A pilot study of a primary prevention curriculum to address preschool behavior problems. *Early Childhood Research Quarterly, 34*, 309–327. https://doi.org/10.1007/s10935-013-0316-1.

Ursache, A, Blair, C., Raver, C.C. (2012). The promotion of self-regulation as a means of enhancing school readiness and early achievement in children at risk for school failure. Child Development Perspectives*, 6:* 122-128.

Webster-Stratton, C., Reid, M. J., & Stoolmiller, M. (2008). Preventing conduct problems and improving school readiness: Evaluation of the incredible years teacher and child training programs in high risk schools. *Journal of Child Psychology and Psychiatry, 49*, 471–488.

Welsh, J. A., Nix, R. L., Blair, C., Bierman, K. L., & Nelson, K. E. (2010). The development of cognitive skills and gains in academic school readiness for children from low-income families. *Journal of Educational Psychology, 102*(1), 43–53. https://doi.org/10.1037/a0016738.

Wenz-Gross, M., Yoo, Y., Upshur, C. C., & Gambino, A. J. (2018). Pathways to kindergarten readiness: Second step early learning curriculum and the mediating role of executive functioning, preschool academic skills, and task behavior. *Frontiers in Psychology, 9*.

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001/2007). *Woodcock-Johnson III.* Itasca, IL: Riverside Publishing.

Yoshsikawa, H., Aber, J. L., & Beardslee, W. R. (2012). The effects of poverty on the mental, emotional, and behavioral health of children. *American Psychologist, 67*, 272–284. https://doi.org/10.1037/a0028015.