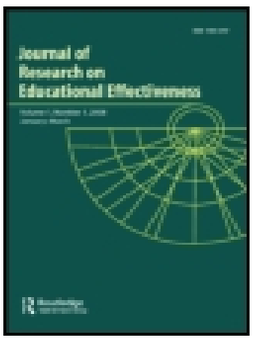

Routledge
Taylor & Francis Group

Journal of Research on Educational Effectiveness

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/uree20

Long-Term Effects of Social-Emotional Learning on Academic Skills: Evidence from a Randomized Trial of *INSIGHTS*

Meghan P. McCormick , Robin Neuhaus , Erin E. O'Connor , Hope I. White , E. Parham Horn , Samantha Harding , Elise Cappella & Sandee McClowry

To cite this article: Meghan P. McCormick , Robin Neuhaus , Erin E. O'Connor , Hope I. White , E. Parham Horn , Samantha Harding , Elise Cappella & Sandee McClowry (2020): Long-Term Effects of Social-Emotional Learning on Academic Skills: Evidence from a Randomized Trial of *INSIGHTS* , Journal of Research on Educational Effectiveness, DOI: 10.1080/19345747.2020.1831117

To link to this article: https://doi.org/10.1080/19345747.2020.1831117

JOURNAL OF RESEARCH ON EDUCATIONAL EFFECTIVENESS
https://doi.org/10.1080/19345747.2020.1831117

Routledge
Taylor & Francis Group

INTERVENTION, EVALUATION, AND POLICY STUDIES

Long-Term Effects of Social-Emotional Learning on Academic Skills: Evidence from a Randomized Trial of *INSIGHTS*

Meghan P. McCormick[a], Robin Neuhaus[b], Erin E. O'Connor[b], Hope I. White[c], E. Parham Horn[b], Samantha Harding[d], Elise Cappella[e] and Sandee McClowry[e]

[a]MDRC, Families & Children Policy Area, New York, USA; [b]Department of Teaching & Learning, New York University, New York, USA; [c]Department of Clinical Psychology, State University of New York - Buffalo, New York, USA; [d]Lesley University, Cambridge, Massachusetts, USA; [e]Department of Applied Psychology, New York University, New York, USA

ABSTRACT

Social-Emotional Learning (SEL) programs are school-based preventive interventions that aim to improve children's social-emotional skills and behaviors. Although meta-analytic research has shown that SEL programs implemented in early childhood can improve academic and behavioral outcomes in the short-term, there is limited work examining program effects on children's math and language skills in the longer-term. Moreover, few studies have considered variation in impacts by children's pre-intervention academic skills. Using an experimental design, the current study leveraged administrative data available through school records ($N = 353$) to examine the impacts of one SEL program—*INSIGHTS into Children's Temperament*—implemented in early elementary school on math and language standardized test scores from third through sixth grade. Findings revealed positive average treatment effects on English/Language Arts (ELA) test scores in third and fourth grade, but not in fifth and sixth grade. Students who had higher academic skills at study enrollment showed lasting impacts on ELA scores in fourth, fifth, and sixth grade. There were no treatment impacts on math skills, and no variation in effects on math achievement by baseline skills. Implications are discussed.

ARTICLE HISTORY
Received 2 January 2020
Revised 7 July 2020
Accepted 14 September 2020

KEYWORDS
Social-emotional learning; intervention; language; math; baseline risk

Children who grow up in poverty are at heightened risk for exhibiting poorer social-emotional skills and psychological and behavioral health, relative to their more affluent peers (McCoy et al., 2018; Sibley et al., 2019). Lower levels of social-emotional competencies, such as the ability to manage and regulate one's behaviors and emotions and to resolve conflicts, are associated with poor academic skills when children begin formal schooling (Raver, 2002). Such findings have helped to spur the development and expansion of Social-Emotional Learning (SEL) programs—school-based preventive

CONTACT Meghan P. McCormick meghan.mccormick@mdrc.org 200 Vesey St. 23rd Floor New York, NY 10281, USA.

Supplemental data for this article can be accessed at https://doi.org/10.1080/19345747.2020.1831117.

interventions that aim to improve children's social-emotional skills and behavioral development—in schools serving primarily low-income students. Although the rigor of the evidence is mixed, there is empirical research demonstrating that SEL programs (Bierman et al., 2010; Brackett et al., 2012; Diamond & Lee, 2011) can improve children's social-emotional and academic outcomes in the short-term (Bierman et al., 2010; Brackett et al., 2012; Corcoran et al., 2018; Diamond & Lee, 2011; Durlak et al., 2011). Perhaps due in part to this evidence base, there is considerable popular and political support for expanding access to SEL programs. For example, the U.S. House of Representatives recently passed legislation to dramatically expand federal funding for "whole-child initiatives" in public schools, including SEL programs (Stringer, 2019).

Few studies, however, have been able to track SEL program participants across time in order to understand the *long-term impacts* of interventions on students' outcomes. For example, a recent meta-analysis by Taylor and colleagues (2017) found that among 82 SEL program impact studies with at least 6 months of follow-up data, only six studies examined effects on child outcomes four or more years post-intervention. There is even more limited research examining whether programs benefit children's *academic outcomes* in the long-term. In order to determine whether early investments in SEL programs translate into on-going benefits for students' academic skills, there is a critical need to conduct long-term evaluation research, and to identify the groups for whom these interventions work best across time.

To this end, the current study adds to the literature by examining the effects of the *INSIGHTS into Children's Temperament* SEL program—implemented in kindergarten and first grade classrooms serving primarily low-income students—on children's English Language Arts (ELA) and math standardized test scores in third through sixth grade. We further test how long-term effects vary by children's academic skills at kindergarten entry. Taken together, findings from this work can help inform efforts to implement structures to support continued benefits for students across time, and to target and/or adjust programming for different groups of students depending on the skills they have when they start school.

Effects of Social-Emotional Learning (SEL) Programs on Students in Low-Income Schools

As noted in a review by McKown (2017), social-emotional competencies include skills such as the ability to infer others' thoughts and feelings (thinking skills), the ability to initiate a positive interaction (behavioral skills), and the ability to stay calm when upset (self-control skills). Children who enter school with better social-emotional skills are likely to receive more positive teacher feedback that motivates subsequent learning (Rabiner et al., 2016). In contrast, poorly developed social-emotional skills at school entry—including lower levels of attention and inhibitory control—may negatively affect students' development of academic skills because attention difficulties undermine the benefit that children obtain from formal instruction. If children fail to develop basic academic skills at school entry as a result, they may fall further behind over time and become increasingly less engaged in academic work (Rabiner & Coie, 2000). As discussed further by Rabiner, Carrig, et al. (2016), children with poorly developed social

skills are more likely to experience peer rejection and have difficulty establishing supportive relationships with teachers. This can result in an aversion to school and reduced classroom participation (Hamre & Pianta, 2005; O'Connor et al., 2014), continuing to negatively affect long-term academic success (Véronneau & Dishion, 2011). Yet, many children—up to one third as estimated by Epstein et al. (2008)—do begin kindergarten exhibiting low levels of social-emotional competencies.

Social-emotional learning (SEL) programs aim to intervene to support the development of these early skills. SEL (as distinct from social-emotional skills or competencies) is defined by the Collaborative for Academic, Social, and Emotional Learning (CASEL; Weissberg et al., 2015) as the process through which children and adults understand and manage emotions, set and achieve positive goals, feel and show empathy for others, establish and maintain positive relationships, and make responsible decisions. As discussed in meta-analytic work (Durlak et al., 2011; Taylor et al., 2017), SEL programs can be implemented by teachers, non-school personnel, or a be a multi-component combination of the two. In teacher-implemented interventions, teachers are trained by curriculum developers, coaches, or other program staff to deliver classroom-based interventions. These programs typically take the form of a specific curriculum and set of instructional strategies (e.g., behavior rehearsal, cooperative learning) that seek to develop specific social and emotional skills. Programs implemented by non-school personnel require outside staff (such as university researchers or external consultants) to come to the school to deliver the intervention. Finally, multi-component interventions generally have two components and often supplement teacher-administered classroom programs with a parent component or schoolwide initiative.[1] Although studies vary in their methodological rigor, there is a large body of research demonstrating that SEL programs implemented in early childhood and elementary school settings improve children's social-emotional skills and behaviors in the short-term (Corcoran et al., 2018; Durlak et al., 2011; Jones et al., 2011; Webster-Stratton et al., 2008). For example, individual studies and narrative reviews report positive outcomes of SEL programming on student concentration and problem behaviors (Durlak et al., 2011; Kellam et al., 2008; Webster-Stratton et al., 2008; Zins et al., 2004).

Importantly, some studies also report positive effects of SEL programs on children's academic achievement (Durlak et al., 2011). For example, a large, cluster randomized control trial of the Chicago School Readiness Project (CSRP)—an emotionally and behaviorally focused classroom-based intervention designed to support low-income preschoolers' school readiness—revealed positive effects on children's preacademic skills prior to the start of kindergarten, as measured by vocabulary, letter naming, and math skills (Raver et al., 2011). Similarly, an efficacy trial of the Incredible Years Program, a comprehensive intervention designed to promote social competence and reduce

[1]In their meta-analysis of SEL program impacts ($N = 213$), Durlak and colleagues (2011) found that 114 (53%) were primarily implemented by teachers, 44 (21%) were implemented by non-school personnel, and 55 (26%) were multicomponent. Effects on academic performance were largest and statistically significant in the studies where teachers implemented the model directly (E.S. = .34) followed by multicomponent interventions (E.S. = .26). The effects of interventions on academic performance implemented by non-school personnel were not statistically significant (E.S. = .12). In the Taylor et al. (2017) meta-analysis examining long-term impacts of SEL programs, of the 82 studies reviewed, 32 (39%) examined programs implemented by teachers, 27 (33%) examined programs implemented by non-school personnel, and 23 (28%) examined multi-component interventions. The authors did not find any evidence that long-term program effects varied by intervention format.

disruptive behavior, has demonstrated positive impacts on children's early school readiness skills (Webster-Stratton et al., 2008). As summarized in a meta-analysis by Yang and colleagues (2019), the impacts of SEL have been particularly pronounced in studies implemented in schools serving primarily low-income students.

Such findings have spurred the proliferation of SEL programs in schools across the country. All fifty states have now integrated SEL into their educational standards in some way (CASEL, 2019). Yet, despite this evidence, there are conflicting findings demonstrating largely null effects of these models on students' *academic skills*. For example, a large-scale evaluation of seven social and character development, including SEL programs, conducted by the Institute of Education Sciences found no direct effects on third- to fifth-grade students' social-emotional skills or academic achievement (Social & Character Development Research Consortium, 2010). Other researchers have also questioned the underlying premise that promoting children's social and emotional skills can improve their academic and behavioral outcomes (Zeidner et al., 2002). For example, work by Duncan and colleagues (2007) examining observations from multiple large-scale national datasets found no statistically significant associations between early social-emotional skills and math and reading skills in third grade.

Long-Term Effects of SEL Programs on Academic Outcomes

Given this mixed evidence base, further research is needed to understand whether SEL programs that demonstrate short-term impacts on children's academic skills continue to have such benefits in the long-term. Leveraging experimental data can help inform whether exposure to early social-emotional supports is causally linked to children's academic outcomes in the long-term. For example, after participating in SEL programs in early elementary school, students then move through later grades in elementary and middle school. This transition to early adolescence presents a new set of developmental tasks, including changes in self-concept and identity along with increasing demands in family, classroom, and peer contexts. Children with strong social-emotional skills may be better suited to navigating this challenging transition, with its changes in instruction, decreases in teacher support, and often increased exposure to negative peer behaviors (Akos, 2005; Eccles, 2004; Rudolph et al., 2001; Simmons & Blyth, 1987; Steinberg, 2005). For example, life course/cascading theories of development (Masten & Cicchetti, 2010) theorize that psychological well-being is reciprocally and positively associated with how well people meet the social demands encountered at each life stage (Ialongo et al., 2006). Based in this theoretical framework, one would expect to see lasting impacts of SEL programs on academic skills, assuming that there were short-term effects on social-emotional, behavioral, and academic outcomes. However, few studies have tested this theory empirically. Indeed, Taylor and colleagues (2017) conducted a meta-analysis examining 82 studies of SEL programs implemented from kindergarten through high school with follow-up periods ranging from 6 months to 18 years. They found that 61% of the studies had follow-up periods under one year.

This is a key limitation in extant work given theory suggesting that advanced academic skills are the outcomes that SEL programs are most likely to change in the long-

term. For example, Bailey and colleagues (2017) have argued that in order to observe lasting effects of an early childhood intervention on *students' skills*, the outcomes that are measured to assess impact must be "trifecta skills"—*fundamental* to children's development, *malleable* to an intervention like an SEL program, and *not likely to develop in the absence of the program*. SEL programs do target a number of skills that are both malleable and fundamental, including basic academic skills, self-concept, academic motivation, emotional self-regulation, executive functioning, and social and relationship skills (Greenberg & Abenavoli, 2017). Yet, Bailey and colleagues (2017) argue that both theoretical and empirical research suggests that the majority of these competencies are also likely to develop even in the absence of an intervention. They cite advanced mathematics skills (e.g., algebra, analysis), language/literacy and communication skills, and academic motivation as potential targets of SEL programs that one could describe as "trifecta skills." When examining the long-term effects of SEL programs on students' academic outcomes, it may be important to measure competencies that explicitly fit within this "trifecta skills" designation.

Indeed, the meta-analysis conducted by Taylor et al. (2017) only identified eight studies of SEL programs with follow-up periods longer than 6 months that also examined impacts on academic performance, defined as grades or standardized test scores. Across these studies, the average effect size on academic outcomes measured in the long-term (average of 195 weeks, or 3.75 years) was .33 standard deviations (95% CI [.17, .49]). The magnitude of this effect was about twice as large as the effect sizes for the other outcomes the authors examined (e.g., SEL skills, attitudes, positive social behaviors, conduct problems emotional distress). Even so, this sample size is limited to eight studies, includes studies of interventions implemented in middle or high school, and aggregates studies across a large range of follow-up periods. It is thus unclear whether and how the magnitude of program effects differ over time as children move through later schooling. This is an important consideration because as children transition from early childhood to middle childhood and then to preadolescence, they face substantial changes in the social and academic demands of their environments (Akos, 2005; Eccles, 2004; Rudolph et al., 2001; Simmons & Blyth, 1987; Steinberg, 2005).

Even so, language and math skills assessed by standardized tests can serve as good examples of "trifecta" skills, as described by Bailey and colleagues (2017). Both assessments are fundamental, with prior work showing that third grade math and language test scores are predictive of critical long-term outcomes, including high school completion and college enrollment (e.g., Hernandez, 2011). They are both also malleable, with policy and intervention studies demonstrating causal evidence that standardized test scores can be improved with targeted supports. Finally, and perhaps most difficult to achieve, assessments in both domains may be less likely to develop in the absence of intervention, meaning that control group students will not necessarily eventually develop advanced language and math skills tested in third to sixth grade, regardless of whether they participate in SEL intervention or not. In contrast, examples of outcomes that would be likely to develop regardless of intervention would be more constrained competencies like letter and word knowledge and early numeracy skills (Bailey et al., 2017).

Variation in Program Impacts by Academic Skills at School Entry

It is also critical to understand *for whom* SEL programs work best in the long-term. Prior work examining whether SEL programs are more effective for some participants than others has tended to consider demographic factors such as socioeconomic status, age, and urbanicity as potential sources of variation (Durlak et al., 2011; Taylor et al., 2017). Yet, on average, studies examining long-term impacts of SEL programs have shown that the benefits of these interventions are similar across participants regardless of racial and socioeconomic demographic characteristics (Taylor et al., 2017). When considering heterogeneity in impacts of SEL programs, it may be more fruitful to consider how effects differ for children who arrive at school with higher versus lower levels of academic skills. Children who enter kindergarten with low academic skills are the students at highest risk for low academic achievement throughout schooling (von Hippel et al., 2018). The strongest predictor of future academic success is the level of academic skills children have when they start kindergarten (Duncan et al., 2007). Interventions that aim to support academic development are thus often explicitly targeted at these lower-skilled students.

We can draw on work from prior social-emotional learning research—cutting across targeted developmental stages—to consider theory suggesting variation in program impacts. For example, the compensatory (McClelland et al., 2017) hypothesis posits that children who enter school with the lowest levels of academic skills are most likely to benefit from intervention because they have the most room for improvement, relative to children with stronger skills and less room to grow. There is some causal evidence from older samples of children to support this theory (e.g., Rimm-Kaufman et al., 2014). In an evaluation of Second Step, a school-randomized universal SEL program implemented over two years in early elementary school, children with initial deficits in skills for learning (e.g. listening, attention), problem solving, and emotion management evidenced the greatest gains from the program with respect to those same skills (Low et al., 2019). In contrast, and aligned with the accumulated advantages hypothesis, it may also be that children who begin school with the highest levels of skills will benefit more from SEL programs because they are better prepared to take advantage of the learning opportunities offered by these interventions. Further examination of variation in SEL program impacts by skills at school entry can more directly inform theory about who benefits most from early SEL programs, and educational practice about how to explicitly target different social-emotional supports as children begin elementary school.

INSIGHTS into Children's Temperament

The current study aims to estimate long-term impacts of one particular SEL program—*INSIGHTS into Children's Temperament (INSIGHTS)*—a comprehensive intervention with teacher, parent, and classroom programs. *INSIGHTS* is an SEL program that uses a temperament framework as a lens through which to provide strategies to support children's social-emotional skills and behaviors. Temperament is an individual's consistent reaction style of responding to people, events, and other environmental stimuli, particularly those involving stress or change. Temperament is multidimensional, biologically

based, but sensitive to environmental inputs, and relatively stable through childhood (Rothbart & Bates, 2006). Key to temperament theory is the concept of goodness of fit, or notion that it is important for a child's temperament to be in consonance with the demands, expectations, and opportunities of the child's environment (Chess & Thomas, 1984). Although temperament itself should not be targeted by intervention, the environment can be modified to improve goodness of fit. Understanding key dimensions of temperament serves as a tool to support teachers, parents, and students by recognizing variation in how different types of students are likely to respond to various situations. Temperament in this way supports SEL implementation in a way that is not "one size fits all."

Using this framework, trained *INSIGHTS* facilitators deliver the intervention to parents and teachers with the goal of helping them to understand key dimensions of a child's temperament and respond with warmth and discipline strategies that support adaptive social-emotional and behavioral outcomes (McClowry, 2014). Trained facilitators also work with elementary school students, delivering classroom curricula designed to enhance empathy for individuals with different temperament styles and to use problem-solving techniques when confronted with daily dilemmas. As further described in Appendix A, the intervention theorizes that after participating in *INSIGHTS*, the quality of the classroom context and the home environment will be enhanced, in turn enhancing students' behaviors, self-regulation, and academic skills. Complementary work examining the short-term impacts of *INSIGHTS* has demonstrated some empirical evidence for this theory of change.

There is also causal evidence that *INSIGHTS* improves low-income racial/ethnic minority students' academic skills in the short-term. Results from a cluster-randomized control trial of *INSIGHTS* revealed positive impacts on students' math and reading achievement approximately 4 months after the end of the intervention. The effect sizes for math (ES = .31) and reading ability (ES = .55) were comparable to the effect sizes found in meta-analytic findings reporting on short-term impacts (e.g. Taylor et al., 2017). The current study builds on this earlier work, leveraging the existing experimental design from this randomized trial to estimate the impacts of *INSIGHTS* on children's English Language Arts (ELA) and math standardized test scores in third, fourth, fifth, and sixth grade. In addition, the study examines variation in impacts of *INSIGHTS* by children's academic skills at kindergarten entry. Given the complexity of the *INSIGHTS* program model and its focus on supporting a broad set of social-emotional skills, one could theorize that the intervention stands to have particular benefits for children who enter school with higher levels of skills and are well-prepared to take advantage of the lessons from the intervention. Findings will inform whether an early SEL program can have lasting effects on students' academic skills, and build evidence about the students who benefit most.

Method

Between 2008 and 2012, 22 elementary schools from three New York City school districts located in one borough (out of 32 school districts total spread across five boroughs) were randomly assigned to participate in the *INSIGHTS* intervention or to an

Table 1. Examination of missing test score data by treatment status.

	INSIGHTS		*Attention-control*			
Outcome	Group attrition (%)	*N*	Group attrition (%)	*N*	Total sample	Diff. attrition (%)
3rd grade ELA test score	17	187	21	166	353	−4
4th grade ELA test score	19	182	21	166	348	−2
5th grade ELA test score	20	180	23	161	341	−3
6th grade ELA test score	22	175	25	157	332	−3
3rd grade math test score	17	187	21	166	353	−4
4th grade math test score	19	182	21	166	348	−2
5th grade math test score	20	180	23	161	341	−3
6th grade math test score	22	175	25	157	332	−3
Total sample	17	187	21	166	353	−4

attention-control condition. The three districts included in the current study were selected because the student populations were primarily low-income and racial/ethnic minority at the time of the study, and they were located in a section of the city that was relatively accessible to the research team. Children began participating in the study when they were in kindergarten (Study Year 1) and continued participating through the end of their first grade year (Study Year 2). Each cohort entered the study over three consecutive years (2008, 2009, and 2010) and participated in two years of data collection (kindergarten classrooms participated in Year 1 and first grade classrooms participated during Year 2). The current paper achieves the study aims by leveraging baseline and administrative test score data for the students who consented to participate in the study in the fall of kindergarten and first grade. This is the same group of students for whom short-term impacts were reported in prior work.

Participants and Setting

Participants in this study include students ($N = 353$) who consented to enroll in the randomized trial of *INSIGHTS* and have at least one valid administrative test score outcome for third, fourth, fifth, or sixth grade. Students in the current study sample ranged from four to seven years of age at baseline ($M = 5.38$, $SD = 0.61$). Half (50%) of the children were male and 87% qualified for free or reduced lunch programs. Approximately 72% of children were Black, 19% were Hispanic non-Black, and the remaining children were biracial. Approximately 28% of children's parents did not finish high school; 26% had a high school or GED diploma; 24% had attended some college; and 22% had completed college. Children enrolled in the study were similar in demographic characteristics to the other students at the schools who were invited but did not consent to participate in data collection activities.

The student participants in the current study are a subset of the 435 ($N = 225$ *INSIGHTS*; $N = 210$ control) students who enrolled in the trial at baseline and participated in the short-term efficacy trial. As illustrated in Table 1, the follow-up sample includes 353 total students ($N = 187$ *INSIGHTS*, $N = 166$ control). The team accessed administrative data on these students for the years they were enrolled in the study and subsequent years through 2017. Data thus covered 2008—when the first cohort enrolled in kindergarten—through the spring of 2017 when, assuming typical promotion from

grade to grade, the bulk of the first cohort was enrolled in eighth grade, the majority of the second cohort was enrolled in seventh grade, and most of the third cohort was enrolled in sixth grade.[2] We matched the original study sample to students' school records in third, fourth, fifth, and sixth grade when we would have ELA and math test scores for all three cohorts. A number of students in the original sample ($N = 82$; 19% of the original sample) left the public school district by the third grade and thus attrited from the study sample. Although students were nested in 22 schools at the end of the intervention, follow-up data revealed that they were distributed across 48 schools by third grade (range of 1−22 students per school in third grade), 54 schools by fourth grade (range of 1−19 students per school in fourth grade) 61 schools by fifth grade (range of 1−17 students per school in fifth grade), and 141 total schools in sixth grade (range of 1−7 students per school in sixth grade) when the large majority of the sample made a shift to a middle school.[3] Overall attrition rates were similar for students originally enrolled in both treatment and control group schools. More specifically, of the students enrolled in one of the 11 schools assigned to treatment, 17% ($N = 38$) attrited from the study. Of the students originally enrolled in one of the 11 schools assigned to the control group, 21% ($N = 44$) attrited from the study (see Table 1 for more detailed attrition rates by outcome). There were no other differences between those who attrited from the sample and those who did not.

Research Procedures

Selection and Randomization of Schools

Elementary schools serving low-income students in one geographic area of New York City were targeted for the study. Recruitment began by contacting principals of schools with more than 50% of students eligible for free or reduced-price lunch. Twenty-three principals agreed to participate. One school withdrew during baseline data collection after an administrative transition. A random numbers table was used to assign schools to receive the *INSIGHTS* intervention or the attention-control condition. All kindergarten and first grade teachers in participating schools were eligible for the study. Ninety-eight percent of eligible teachers consented and none withdrew.

Study Enrollment and Timeline

Students participated in intervention activities across two years—kindergarten (study year 1) and first grade (study year 2). Students' kindergarten teachers were invited to participate in the study when they were enrolled in kindergarten (study year 1); first grade teachers were invited to participate when students were enrolled in first grade

[2]Cohort 1 students enrolled in the study in kindergarten in the fall of 2008 and were enrolled in 3rd, 4th, 5th, and 6th grades between 2011–2012 (3rd grade) and 2014–2015 (6th grade). Cohort 2 students enrolled in the study in kindergarten in the fall of 2009 and were enrolled in 3rd, 4th, 5th, and 6th grades between 2012–2013 (3rd grade) and 2015–2016 (3rd grade). Cohort 3 students enrolled in the study in kindergarten in the fall of 2010 and were enrolled in 3rd, 4th, 5th, and 6th grades between 2013–2014 (3rd grade) and 2016–2017 (6th grade).

[3]This distribution may be unique to NYC where students have middle school choice and there is a public transportation infrastructure to support students to attend middle schools across the full city.

(study year 2). Children and classrooms were recruited for the study during the beginning of kindergarten and baseline data were collected from students and classrooms in the Fall of the kindergarten year. The *INSIGHTS* intervention was implemented for 10 weeks in the winter of the kindergarten year. Data were then collected from classrooms and students in the spring of the kindergarten year. Following the summer, data were collected from the same set of students enrolled in the study in the Fall of the first grade year, as well as the new first grade classrooms participating in the second year of the study. The intervention was implemented across 10 weeks in the late fall and early winter of the first grade year. Follow-up data were then collected in the spring of first grade from students and classrooms.

Students in *INSIGHTS* schools received the full intervention in both kindergarten and first grade. Children had the opportunity to participate in 2 years of the study, but teachers only participated in the year of the study when the children they were teaching were targeted. None of the teachers in the study participated in more than one year of intervention. Parents who enrolled their children in data collection activities participated in the intervention in kindergarten only.

Data Collection

Pretreatment data were collected within each study year via direct child assessments by trained data collectors. Teachers completed background questionnaires and parents reported on demographics and child temperament. Administrative data on school demographic characteristics were publicly available. Teachers, intervention facilitators, and fidelity coders completed intervention fidelity checklists to help understand the extent to which students, teachers, and parents experienced the *INSIGHTS* program. The team then obtained further data from the New York City Department of Education (NYC DOE) through the Research Alliance for New York City Schools (RANYCS). The NYC DOE collects and records administrative data on all students at the beginning (October) and end (June) of each school year. The research team received administrative data through the spring of 2017 for all students who were enrolled in kindergarten in one of the schools participating in the study. The team then used ID numbers, names, and birthdates to match the administrative data to the study sample in order to have access to information on students' school records through the spring of 2017. As noted above, this article explicitly focuses on outcomes through sixth grade, when the large majority of students made a critical transition to middle school.

INSIGHTS and Attention-Control Procedures

During the 2 years of study participation, schools assigned to *INSIGHTS* received parallel sessions for teachers, parents, and children within their classroom. Parent and teacher sessions were held at each school in ten, two-hour meetings, with makeup sessions offered as needed. Teacher sessions were held during the school day or after school, and parent sessions were also offered both during and after school. Parents received $20 for each session attended and teachers received professional development credit and $40 for classroom resources. All parent and teacher sessions occurred in

groups with facilitators at the school. Make-up sessions were available to parents and teachers who missed sessions but were used minimally.

In the *INSIGHTS* parent and teacher program sessions, caregivers learn to recognize a child's temperament based on four empirically derived temperament typologies (McClowry, 2002): Hilary the Hard Worker who is industrious; Gregory the Grumpy who is high maintenance; Fredrico the Friendly who is social and eager to try; and Coretta the Cautious who is shy. Even so, because approximately 50% of children are not represented by one of the temperament profiles (McClowry, 2002), parents and teachers also learn the dimensions of temperament and how to understand children uniquely with that information. Parents and teachers are then encouraged to reframe their perceptions more positively and to select strategies that match a child's particular temperament. For example, parents and teachers are taught to recognize a shy temperament (e.g., the child has a tendency to withdraw) and encouraged to appreciate that shy children are often astute observers who are sensitive and cautious when encountering new situations or meeting people. Teachers and parents also learn to use a scaffold-and-stretch approach when children encounter challenging situations for their temperament. For example, when assigning a classroom activity that is challenging to a shy child, such as participating in a school assembly, teachers are encouraged to assess the student's distress level. If the child is likely to be unduly distressed, the teacher can make the activity more manageable by first arranging a practice session with a friend and then with a small group of classmates before a whole-class rehearsal is set to take place (McClowry, 2014).

Acknowledging the child's efforts in each progressively challenging step is another part of this responsive strategy. In *INSIGHTS* classroom sessions, activities focus on empathy and problem-solving skills. The students are introduced to puppets exemplifying the same four temperament typologies. The children are encouraged to understand the puppets' respective strengths and challenges. For example, the Coretta the Cautious puppet thinks carefully before she acts but warms up when provided more time. The children also use the puppets to resolve videotaped dilemmas and those they encounter in their own lives. In a dilemma involving the Coretta the Cautious puppet, the children and the puppets encourage her to be assertive when she hesitates to ask her teacher for help.

To maintain model fidelity, facilitators followed scripts, used material checklists, and documented sessions. Deviations or clinical concerns were discussed weekly in supervision with the program developer. Supervision focused on challenges related to conducting sessions, implementation logistics, and participant concerns. All teacher and parent sessions were videotaped and reviewed for coverage of content and effectiveness of facilitation (Hulleman & Cordray, 2009). Videotapes demonstrated that 94% of the curriculum was adequately covered in the teacher sessions and 92% of the curriculum was covered for the parent sessions.

Intervention Dosage

The average number of teacher sessions attended was 9.44 ($SD = .91$). The average number of classroom sessions attended by consented children was 8.30 ($SD = 2.25$). The average number of parent sessions attended by parents of participating children was

5.93 (SD = 4.15). Twenty-five percent of parents were present for all sessions and 30.3% were present for eight or nine sessions.

Attention-Control

Schools not assigned to *INSIGHTS* participated in a 10-week, supplemental reading program after school hours. Teachers and parents attended two 2-h workshops in which reading coaches provided reading materials and presented strategies to enhance early literacy. Parents received $20 and teachers received professional development credit and $40 for classroom resources for each workshop. We retain the terminology "attention-control" throughout the paper in order to align this study with prior work evaluating the *INSIGHTS* intervention.

Measures

ELA and Math Test Scores

As noted above, we accessed students' scores on the state English Language Arts (ELA) and math standardized test scores through administrative records. We used z-scoring to standardize ELA and math scores within grade, subject, and testing year. This scoring approach was necessary because the state test changed across years and cohorts and we needed to make the test score interpretation similar across cohorts and years. As such, a score of 0 represents the mean value on the test, relative to all students in the district. As discussed below, we performed sensitivity analyses to determine how robust our results were to the exclusion of students who had been retained in grade at some point. This was necessary because there were concerns in the data about whether retained students' test scores were accurately linked to the test year. We used the full test score measure, rather than the categorical measure assessing each student's level of proficiency on the assessment, to both align with prior work using these test score measures as outcomes in intervention and policy studies (e.g., Legewie et al., 2019; Martorell & Mariano, 2018) and because the continuous score yielded greater statistical power to detect intervention impacts.

Baseline Academic Skills

We used baseline scores on the Woodcock Johnson III Letter-Word ID (LWID) and Applied Problems (AP) subtests (Woodcock et al., 2001) to create a composite of students' academic skills at study enrollment. The Letter-Word ID subtest assesses letter naming and word decoding skills by asking children to identify a series of letters and words. Possible scores range from 0 to 76. The Applied Problems subtest assesses simple counting skills and the ability to analyze and solve mathematical word problems. Possible scores range from 0 to 64. The WJ-III typically correlates with measures of cognitive ability (rs = .66–.73 with the Wechsler Preschool and Primary Scale of Intelligence—Revised; Weschler, 1989). The reliabilities for the subtests are .88 for LWID and .84 for AP.

We created the composite score by calculating the average of the Letter Word ID and Applied Problems scores for each student. There were four reasons why we decided to

create a composite of baseline academic skills rather than examine baseline literacy and math skills as separate moderators. First, given the number of outcomes we aimed to examine in 3rd, 4th, 5th, and 6th grade, we were wary about then also doubling the number of moderation analyses. Using a composite measure allowed us to examine variation in outcomes by baseline academic skills in the most parsimonious possible way. Second, we did not have a particular theory as to why we would expect to see variation in impacts for baseline literacy versus math skills specifically and thus chose to consider a more holistic assessment of academic skills at school entry across both literacy and math domains. Third, our baseline measures of math and literacy skills were correlated at $r = .68$ suggesting that there was sufficient overlap between them to warrant creation of a composite. And finally, prior work (O'Connor & McCartney, 2007) has used a similar approach for creating this composite, setting a helpful precedent for its utility in demarcating academic skills in early schooling.

Baseline Disruptive Behaviors

The teacher-reported Sutter-Eyberg Student Behavior Inventory (SESBI) was used to measure children's behavior problems (Eyberg & Pincus, 1999), and is included as a pretreatment covariate in analytic models. The SESBI contains 36 items that measure disruptive or conduct problem behaviors. Using a 7-point Likert scale, teachers reported on the frequency with which each student engaged in a series of problem behaviors on a 7-point scale, ranging from 1 (*never*) to 7 (*always*). The overall average of the frequency of problem behaviors was taken to calculate the average score for each student. The SESBI has demonstrated high internal consistency, as well as high test-retest reliability and convergent and discriminant validity (Eyberg & Pincus, 1999; $\alpha = .96$ in current study).

Demographic Characteristics

At public school enrollment, parents and guardians reported on their children's demographic characteristics—race, ethnicity, gender, eligibility for free or reduced price lunch, birthdate, and whether their child spoke a language other than English. Given the distribution of these variables across the sample, we coded race as Black = 1, Other race = 0; ethnicity as Hispanic = 1, not Hispanic = 0; language status as Dual Language Learner = 1 and 0 otherwise; and gender as female = 1 and 0 otherwise. Child age on September 1st of the kindergarten year was calculated using the birthdate and included as a covariate in analytic models. Administrative data were used to determine whether children had attended public PreK (coded as attended PreK = 1 and 0 otherwise). Parents also reported on their own education level when they enrolled their child in the study. We used this information to create four parental education groups—less than HS diploma, HS diploma/GED, 2 year college degree, and BA or greater. Analyses include the first three groups as covariates with BA or greater as the reference group. Taken together, these demographic characteristics are used as covariates in predictive models.

School-Level Characteristics

We included a set of covariates at the baseline school-level as well in our predictive models that match the set of school-level covariates we used in our short-term follow-up study.

We used publicly available administrative data to capture the school size (number of students enrolled in the year prior to the intervention implementation), school attendance rate, the percent of students in the school who were proficient on the state ELA test in the prior year, and the percent of students in the school proficient on the state math test in the prior year. We further adjusted for study cohort at the school-level including dummy variables for Cohort 2 and Cohort 3, with Cohort 1 as the reference group. This approach mirrored the modeling we used to account for cohort in the earlier short-term impact study. We used additional school-level demographic characteristics to examine baseline equivalence at the school-level. These were school-level versions of the student-level demographic characteristics (percent female, percent Black, percent Hispanic, percent eligible for free/reduced price lunch, percent immigrant, percent DLL, and percent who attended a district PreK program).

Analytic Approach

Missing data analysis

As noted above, the total analytic sample size for this study is $N = 353$ students, with 82 students attriting from the sample since baseline. Of the total number of students, 332 (76% of the baseline sample) remained in the sample through the end of the sixth grade, 341 (78% of the baseline sample) through fifth grade, 348 (80% of the baseline sample) through fourth grade, and 353 through third grade (81% of the baseline sample). Follow-up analyses revealed that students in the treatment and attention-control group attrited at similar rates (across the 353 sample size, 21% of the baseline attention-control group attrited and 17% of the treatment group attrited). In Table 2 we further illustrate how the baseline characteristics are similar across the non-attrited students in the treatment and attention-control groups. According to the standards from the What Works Clearinghouse (WWC) (2017), this study constitutes a low-attrition RCT, meaning that, given the study sample size, the overall attrition and differential attrition indicate a tolerable level of bias for the estimated intervention effect.

Overall, there was a relatively low amount of missing data across study variables. All covariates had less than 9% missing and 17% of the sample was missing some covariate data. Using the assumption that data were Missing at Random (MAR), the team used a multiple data imputation method in STATA MICE and imputed 100 datasets in order to generate complete data on all covariates, which had minimal levels of missingness (Graham et al., 2007). We did not impute outcome data. We present the results from models using multiple imputation in the main analysis but also did consider how results varied using a list wise deletion approach. As described further below, results were robust across models (see further details in Appendix B).

Impact Analysis

Because school was the original unit of random assignment and most students stayed in the same elementary school that they enrolled in during kindergarten, we expected that student outcomes would not be independent at the third, fourth, and fifth grade follow-ups. However, we did observe wider distribution of the sample across 141 middle

Table 2. Baseline descriptive statistics for *INSIGHTS* and attention-control groups.

	Attention-control		*INSIGHTS*			
Characteristic	Mean	SD	Mean	SD	Std. diff	Sig. diff.
School-level	0.88	–	0.86	–	0.02	
Annual attendance rate	0.51	–	0.49	–	0.02	
Female (%)	0.76	–	0.79	–	−0.03	
Black (%)	0.39	–	0.43	–	−0.04	
Hispanic (%)	0.11	–	0.09	–	0.02	
White (%)	0.13	–	0.12	–	0.01	
Other race (%)	0.72	–	0.75	–	−0.03	
Eligible for free or reduced price lunch (%)	0.08	–	0.06	–	0.02	
Dual Language Learner (%)	0.06	–	0.05	–	0.01	
From recent immigrant family (%)	0.59	–	0.62	–	−0.03	
Attended public PreK (%)	0.53	–	0.49	–	0.04	
% students scoring average on ELA state test	0.63	–	0.61	–	0.02	
% students scoring average on math state test	505.07	190.09	493.21	138.62	0.06	
Number of students in school						
Sample size	11		11			
Student characteristics for original study sample						
Baseline age	5.37	0.72	5.41	0.66	−0.06	
Female (%)	0.48	–	0.48	–	0.00	
Black (%)	0.73	–	0.77	–	−0.04	
Hispanic (%)	0.35	–	0.38	–	−0.03	
Eligible for free or reduced price lunch (%)	0.72	–	0.75	–	−0.03	
Dual Language Learner (%)	0.09	–	0.11	–	−0.02	
From recent immigrant family (%)	0.04	–	0.06	–	−0.02	
Attended public PreK (%)	0.62	–	0.66	–	−0.04	
Baseline academic skills composite	33.14	10.98	32.81	11.14	0.03	
Disruptive behaviors at baseline	2.25	1.19	2.28	1.24	−0.03	
Sample size	225		210			
Student characteristics for non-attrited sample						
Baseline age	5.56	0.36	5.61	0.31	−0.14	
Female (%)	0.48	–	0.46	–	0.02	
Black (%)	0.76	–	0.81	–	−0.05	
Hispanic (%)	0.14	–	0.16	–	−0.02	
Eligible for free or reduced price lunch (%)	0.96	–	0.95	–	0.01	
Dual Language Learner (%)	0.07	–	0.03	–	0.04	
Attended public PreK (%)	0.72	–	0.73	–	−0.01	
Baseline academic skills composite	32.86	11.12	32.17	11.09	0.06	
Disruptive behaviors at baseline	2.25	1.03	2.37	1.18	−0.12	*
Outcomes for non-attrited sample[a]						
3rd grade ELA test score (*z* scored)	−0.03	0.85	−0.08	0.96	0.05	
4th grade ELA test score (*z* scored)	−0.11	1.09	−0.17	0.95	0.06	
5th grade ELA test score (*z* scored)	−0.28	0.98	−0.24	0.91	−0.04	
6th grade ELA test score (*z* scored)	−0.29	1.01	−0.26	0.97	−0.03	
3rd grade math test score (*z* scored)	−0.15	0.99	−0.11	0.89	−0.04	
4th grade math test score (*z* scored)	−0.34	0.90	−0.29	0.84	−0.06	
5th grade math test score (*z* scored)	−0.34	0.91	−0.31	0.99	−0.03	
6th grade math test score (*z* scored)	−0.34	0.91	−0.31	0.99	−0.03	
Sample size	187		166			

Note. [a]As noted in the text, we use z scores from the school district to assess children's test scores.
A score of "0" represents the average student score across the entire city in the year the student took the test.
A positive score indicates that the student scores above the city average, while a negative score indicates that a student scored below the city average.
***$p < .01$; **$p < .05$; *$p < .10$.

schools following the transition to sixth grade. In order to determine the most appropriate fit to the data, we considered a number of different models and compared their log likelihoods and fit statistics (AIC and BIC). This approach for model selection has been recommended by Scott et al. (2013). For both ELA and math test scores in third grade, we first ran unconditional OLS regression models with clustered standard errors for

baseline school. Then, we compared that model fit to a multi-level model (Fitzmaurice et al., 2004) that included random effects for the school that participated in the original trial. After comparing the log likelihoods of the models and the AIC/BIC statistics across these specifications, we found that the multi-level model was a better fit to the data for both outcomes. We then considered a number of covariance structures, including the unstructured, independent, exchangeable, and identity structures. Results were almost identical across approaches so we retained the unstructured covariance structure for our main set of analyses. We aimed to keep the analytic approach consistent with prior work examining short-term impacts and thus chose not to center the Level 1 covariates, which Enders and Tofighi (2007) argues is the level at which centering can influence estimation in a two-level model. Because our main variable of interest was assignment to *INSIGHTS*, which was measured at Level 2, the analysis did not focus explicitly on the Level 1 covariates taken on their own. The base model examining average treatment effects is illustrated in Equation (1).

$$TESTSCORE_{ij} = \beta_1 + \beta_2 INSIGHTS_j + \delta_{ij} + \alpha_j + \zeta_j + \varepsilon_{ij} \quad (1)$$

In this equation, $TESTSCORE_{ij}$ refers to the test score outcome of interest for student i in school j, β_1 is the intercept, β_2 is the treatment impact for students who participated in the study in school j, δ_{ij} is a vector of student-level covariates, α_j is a vector of school-level covariates, ζ_j is a random intercept for school, and ε_{ij} captures unexplained residual error. After estimating the main effects of *INSIGHTS* on the study outcomes, we then considered how impacts varied by baseline academic skills by creating cross-level interactions between assignment to *INSIGHTS* and the composite variable representing baseline academic skills included in the vector of covariates. Equation (2) illustrates the test of cross-level interactions.

$$TESTSCORE_{ij} = \beta_1 + \beta_2 INSIGHTS_j + \beta_3 BaselineRisk_{ij} + \beta_4 BaselineRisk_{ij}\ XINSIGHTS_j + \delta_{ij} + \alpha_j + \zeta_j + \varepsilon_{ij} \quad (2)$$

All findings are reported as unstandardized coefficients predicting the z scored outcome. For statistically significant effects we then calculated effect sizes by dividing the coefficient of interest by the standard deviation of the outcome.

We considered four sets of robustness checks. First (discussed in Appendix B), we examined whether and how effects differed when we used multiple imputation to handle missing covariate data as compared to a listwise deletion approach. Second, given the relatively small number of schools in the study and potential for imbalance in findings, we tested models using OLS regression with clustered standard errors for study school (discussed in Appendix C). Third, we removed any students who had been retained in grade from the impact analysis to determine whether our results were at all affected by the test scores applied to retained students. This was a necessary check because in our data cleaning work we realized that it was sometimes impossible to ensure that students were assigned to the correct test grade if they had been retained (see full discussion in Appendix D). Finally, we included IEP status in kindergarten (which was measured post random assignment) as a covariate to determine whether impacts were consistent after considering receipt of special education services (see Appendix E).

As discussed further in the Appendices, we found that our impact findings were robust across these checks.

Results

Checking for Baseline Equivalence

We compared characteristics of study students and schools randomly assigned to the *INSIGHTS* and attention-control conditions to assess baseline equivalence. Although our prior work conducting the short-term impact analysis of *INSIGHTS* had demonstrated evidence for baseline equivalence, it was important to continue to assess equivalence for the longer-term follow-up sample given the sample attrition described above. A summary of our baseline equivalence checks is included in Table 2. We found that although the attention-control group appeared to be less disadvantaged than the treatment group on some measures, these differences were not statistically significant at $p < .05$. Moreover, consistent with the findings from the original short-term impact RCT, we did not find any statistically significant differences in the school-level characteristics between the *INSIGHTS* and attention-control groups.

Descriptive Statistics for Outcome Variables

Descriptive statistics (Mean, *SD*) on the standardized test scores for the overall sample are reported in Table 2. Independent samples *t*-tests revealed similar levels of ELA and math test scores across the *INSIGHTS* and attention-control groups in third, fourth, fifth, and sixth grade.

Impacts of INSIGHTS on ELA and Math Standardized Test Scores

Analyses examining impacts on outcomes are presented in Table 3 (ELA) and Table 4 (math). Our adjusted models revealed that students assigned to *INSIGHTS* schools had higher ELA standardized test scores than students assigned to attention-control schools in third ($\gamma = .27$, $SE = .11$, $p = .03$) and fourth grade ($\gamma = .23$, $SE = .12$, $p = .04$). However, there were no statistically significant differences between the groups on ELA test scores in fifth ($\gamma = -.06$, $SE = .18$, $p = .43$) and sixth grade ($\gamma = -.05$, $SE = .14$, $p = .38$). The treatment impact in third grade translates into .28 SD units, while the impact in fourth grade can also be expressed as .24 SD units—small to moderate effect sizes that align with the magnitude of effects found in other long-term studies of SEL programs (Taylor et al., 2017). There were no statistically significant impacts of *INSIGHTS* on math test scores in any of the tested grades (Table 4).

Variation in Impacts of INSIGHTS by Baseline Academic Skills

Further probing of these results revealed that the impact of *INSIGHTS* on ELA test scores in third grade did not vary by baseline academic skills ($\gamma = .004$, $SE = .11$, $p = .86$). However, we did find that the interactions between baseline academic skills and ELA scores in fourth ($\gamma = .026$, $SE = .009$, $p = .03$), fifth ($\gamma = .022$, $SE = .01$, $p = .04$),

Table 3. Impacts of *INSIGHTS* on reading achievement in third, fourth, fifth grade, and sixth grade.

	3rd grade		4th grade		5th grade		6th grade	
Fixed effects	γ	SE	γ	SE	γ	SE	γ	SE
Intercept	−0.42	1.05	−1.09	1.05	−1.21	1.06	−1.28	1.12
Student-level demographic characteristics								
Child eligible for free lunch	−0.72***	0.17	−0.63***	0.18	−0.63***	0.19	−0.59***	0.21
Child Black	0.02	0.25	0.06	0.25	0.33	0.27	0.38	0.22
Child Hispanic	0.05	0.27	0.11	0.28	0.35	0.31	0.41	0.38
Dual Language Learner	0.28	0.26	0.25	0.29	−0.03	0.33	−0.06	0.29
Child female	0.28***	0.17	0.25**	0.11	0.24**	0.11	0.21**	0.09
Child age at baseline	−0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
Parent education less than high school grad.	−0.21	0.15	−0.35***	0.12	−0.13	0.14	−0.11	0.09
Parent education high school grad.	0.02	0.14	−0.11	0.11	0.11	0.13	0.09	0.11
Parent education 2 year degree	−0.11	0.11	−0.12	0.07	−0.12	0.09	−0.17	0.12
Attended public PreK	0.11	0.12	0.21*	0.12	0.03	0.12	0.08	0.05
School-level characteristics								
Cohort 2	0.09	0.06	0.11	0.08	0.13	0.08	0.08	0.08
Cohort 3	0.14**	0.05	0.12**	0.04	0.11**	0.05	0.09	0.05
School size	0.03**	0.01	0.03**	0.01	0.02*	0.01	0.01	0.01
School % competent ELA test	0.22**	0.07	0.25**	0.08	0.19**	0.08	0.11*	0.06
School % competent math test	0.15*	0.08	0.14*	0.07	0.14	0.08	0.12	0.09
School average attendance	0.02**	0.01	0.02**	0.01	0.01	0.01	0.01	0.01
Baseline covariates from original RCT								
Disruptive behaviors at baseline	−0.01	0.07	−0.11	0.08	−0.05	0.08	−0.03	0.05
Assessed academic skills at baseline	0.04***	0.01	0.04***	0.01	0.03***	0.01	0.04***	0.01
Assignment to *INSIGHTS*	0.27**	0.11	0.23**	0.12	−0.06	0.18	−0.05	0.14
Random effects								
Between-school variance	0.02*	0.01	0.02*	0.01	0.03	0.03	0.04	0.03
Residual variance	0.51***	0.05	0.56***	0.05	0.52***	0.06	0.51***	0.05
Log likelihood	−211.73		−242.32		−226.23		−222.54	
N	353		348		341		332	

***$p < .01$; **$p < .05$; *$p < .10$.

Table 4. Impacts of *INSIGHTS* on math achievement in third, fourth, fifth, and sixth grade.

	3rd grade		4th grade		5th grade		6th grade	
Fixed effects	γ	*SE*	γ	*SE*	γ	*SE*	γ	*SE*
Intercept	−0.08	1.09	−1.71*	0.97	−0.71	1.03	−0.86	1.43
Student-level demographic characteristics								
Child eligible for free lunch	−0.21	0.18	−0.23	0.17	−0.19	0.18	0.17	0.12
Child Black	−0.07	0.26	0.18	0.23	0.12	0.26	0.19	0.22
Child Hispanic	−0.19	0.28	0.18	0.26	0.05	0.29	0.04	0.31
Dual Language Learner	0.69**	0.27	0.38	0.28	0.71***	0.32	0.68**	0.29
Child female	−0.06	0.11	−0.02	0.09	0.06	0.11	0.06	0.11
Child age at baseline	−0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
Parent education less than high school grad.	0.17	0.18	−0.11	0.18	−0.08	0.14	−0.09	0.09
Parent education high school grad.	0.05	0.11	0.08	0.27	−0.09	0.13	0.08	0.11
Parent education 2 year degree	0.04	0.31	0.12	0.38	−0.13	0.18	0.05	0.12
Attended public PreK	−0.04	0.13	−0.01	0.01	0.26**	0.12	0.18*	0.09
School-level characteristics								
Cohort 2	0.11	0.08	0.12	0.14	0.12	0.15	0.12	0.16
Cohort 3	0.11**	0.04	0.12**	0.05	0.13**	0.06	0.11**	0.05
School size	0.02**	0.01	0.02**	0.01	0.02*	0.01	0.01	0.01
School % competent ELA test	0.17**	0.08	0.16**	0.04	0.15**	0.05	0.11*	0.06
School % competent math test	0.11	0.08	0.12*	0.07	0.12	0.08	0.13	0.09
School average attendance	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
Baseline covariates from original RCT								
Disruptive behaviors at baseline	−0.09	0.07	−0.09	0.07	−0.13*	0.07	−0.11*	0.06
Assessed academic skills at baseline	0.05***	0.01	0.06***	0.01	0.06***	0.01	0.04***	0.01
Assignment to *INSIGHTS*	0.14	0.12	−0.03	0.11	−0.11	0.11	−0.09	0.12
Random effects								
Between-school variance	0.08**	0.03	0.07**	0.02	0.07**	0.02	0.06*	0.03
Residual variance	0.54***	0.06	0.46***	0.04	0.51***	0.05	0.49***	0.04
Log likelihood	−211.73		−242.32		−226.23		−229.43	
N	353		348		341		332	

***$p < .01$; **$p < .05$; *$p < .10$.

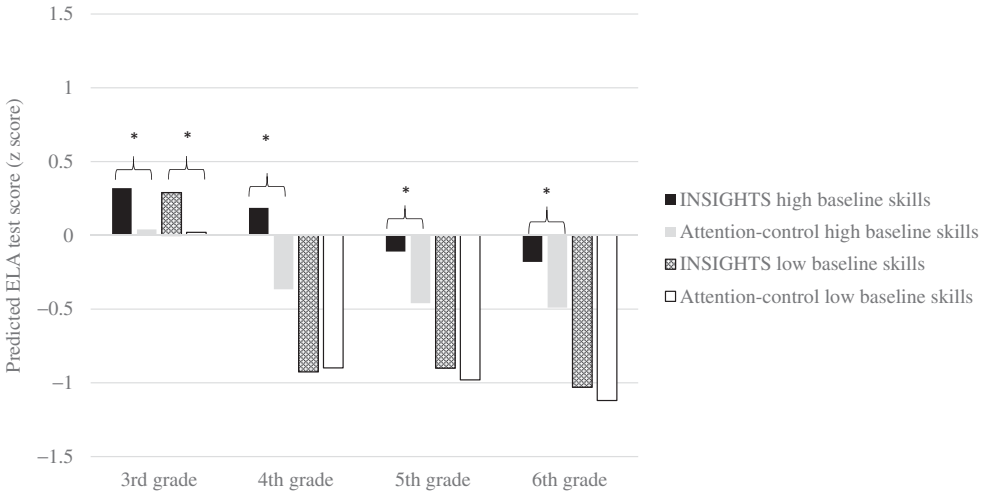

Figure 1. Variation in impacts of *INSIGHTS* in third through sixth grade, by baseline academic skills. *Note.* $^{***}p < .01$, $^{**}p < .05$, $^{*}p < .10$. Figure demonstrates that the following group-based differences in predicted ELA test scores are statistically significant at the .05 level: (1) the difference between the *INSIGHTS* high baseline skills group and the attention-control high baseline skills group in 3rd grade; (2) the difference between the *INSIGHTS* low baseline skills group and the attention-control low baseline skills group in 3rd grade; (3) the difference between the *INSIGHTS* high baseline skills group and the attention-control high baseline skills group in 4th grade; (4) the difference between the *INSIGHTS* high baseline skills group and the attention-control high baseline skills group in 5th grade; and (5) the difference between the *INSIGHTS* high baseline skills group and the attention-control high baseline skills group in 6th grade. The prototypical high baseline skills group scored 1 SD more than the mean on the baseline composite while the prototypical low baseline skills group scored 1 SD less than the mean on the baseline composite.

and sixth grade ($\gamma = .025$, $SE = .012$, $p = .04$) were statistically significant. As illustrated in Figure 1, findings revealed that students in schools assigned to *INSIGHTS* who had higher levels of baseline academic skills (defined as scoring more than one SD above the mean on the baseline composite) performed better than similarly skilled attention-control groups students on ELA tests in fourth, fifth, and sixth grade. There were no treatment differences in ELA test scores at any follow-up time point for students in the low baseline academic skill group (defined as scoring one SD below the mean on the baseline composite). The interaction between baseline academic skills and assignment to *INSIGHTS* was not statistically significant in any of the models predicting math test scores.

Robustness Checks

We found that results were robust across approaches for handling missing data (Appendix B), different modeling specifications (Appendix C), removing students who were retained in grade from analyses (Appendix D), and including IEP status as a covariate in impact models (Appendix E). Results from the robustness checks are discussed in more detail in Appendices.

Discussion

This study aimed to add to the literature on social-emotional learning by conducting experimental research examining the long-term impacts of one SEL program–*INSIGHTS into Children's Temperament*—implemented at the transition to elementary school on students' standardized ELA and math test scores in third through sixth grades. Results from analyses examining the full sample of students revealed that *INSIGHTS* had lasting impacts on children's ELA test scores in third and fourth grades. There were no impacts on math test scores or ELA test scores in fifth and sixth grade. To our knowledge, this is one of the few examples of a study that has found lasting impacts of an SEL program implemented in early schooling on academic skills assessed with standardized tests three years following the end of the intervention. The effect sizes for the statistically significant impacts on ELA test scores are aligned with the average magnitude of effects that Taylor and colleagues (2017) found in their meta-analysis summarizing results across eight studies estimating long-term effects of SEL programs on academic performance (measured with grades or test scores). Impacts can be considered substantial when benchmarked against normative trajectories in academic skills, with effect sizes in both third and fourth grade translating into about 4 months of language/literacy development (Lipsey et al., 2012). Additionally, it is important to remember that in this study, the effects of *INSIGHTS* were compared to an attention-control condition where students, parents, and teachers were exposed to an after-school reading program. Results suggest that the benefits students accrue from exposure to SEL supports may have a more lasting effect that non-intensive activities focused on academic content.

Yet, although we observed some lasting impacts of *INSIGHTS* on students' test scores, a key finding from this work was that effects were not sustained during children's final year of elementary school *and*, importantly, during the shift to middle school. During this time period students were transitioning to early adolescence, a unique and challenging phase of development where peer relationships become increasingly salient and children experience downward shifts in self-esteem, efficacy, and engagement (Mikami et al., 2017; Wang et al., 2018). The skills that children learned and developed after participating in *INSIGHTS* in kindergarten and first grade may have enhanced their behavioral regulation and academic engagement in middle childhood, thus supporting their academic skills and performance on ELA standardized tests during these grades. However, it may be that the intervention was insufficient or inappropriate for helping students develop the more demanding set of social-emotional skills needed to successfully navigate the transition to adolescence (Chung & McBride, 2015; Yang et al., 2018), thus resulting in largely null impacts by fifth grade. This is a somewhat different pattern of findings than other evaluations have detected (Taylor et al., 2017). For example, a long-term evaluation of the Chicago School Readiness Project—an intensive intervention to support students' behavioral regulation in PreK—demonstrated lasting effects on students' self-reported grades 10–11 years following program completion (Watts et al., 2018). Importantly, this study relied on self-reported grades as an outcome and considered a longer-term follow-up period than the current study which could help explain differences in the findings. It may be that although we conceptualized our outcomes in this study as meeting Bailey and colleagues' (2017) definition of "trifecta skills," the

competencies assessed were likely to develop for students regardless of exposure to intervention or not. We unfortunately lack item-level information on the outcomes so cannot fully examine the face validity of the outcomes to better explore this possibility.

Moreover, we observed a different pattern of results depending on the outcome domain of interest—language or math. After having detected short-term impacts of *INSIGHTS* on both math and language through the end of first grade, the current study showed that only the effects on language were sustained two and three years later. We may have observed impacts on math skills in kindergarten and first grade because the *INSIGHTS* intervention immediately supported students to develop the self-regulatory skills they needed to engage in cognitively demanding daily math activities in the early elementary classroom (Blair et al., 2015). However, *INSIGHTS* did not provide any direct math instruction nor did teachers integrate *INSIGHTS* discussions of daily dilemmas into math learning. As such, when the *INSIGHTS* supports did not continue into the later elementary school grades, children lacked consistent exposure to the sustaining environment necessary to reinforcing and building on the key self-regulatory skills that supported their math skills (Bailey et al., 2017). In contrast, we may have observed early impacts on language and literacy because *INSIGHTS* helped students develop attentional and behavioral skills that allowed them to better engage in daily classroom activities specifically used in early elementary school (e.g., call and response, shared reading), continuing through third and fourth grade. Moreover, many teachers embedded *INSIGHTS* content into everyday language and literacy instruction and students had opportunities to practice these skills in those learning contexts. As such, they may have continued to benefit from intervention through fourth grade because they continued to have ample opportunities to practice these skills in similar settings. As instruction became focused on more complex skills (e.g., reading comprehension) in fifth and into sixth grade (Li et al., 2016), further intervention would be needed to continue to sustain these impacts.

Related, recent work by Eisenberg et al. (2018) has argued that the relationship between temperament and academic achievement over time is influenced by multiple mediating variables such as liking school or the quality of relationships with teachers and peers, especially for children who are high in negative reactivity. Although *INSIGHTS* supported academic achievement in the short term for such students, there was no continuing intervention across time to enhance these mediators. Thus, in line with the "sustaining environments" theory discussed by Bailey and colleagues (2017), an SEL program like this using a temperament framework may not be able to show longer-term impacts without continued supports across time. This finding suggests that future work may consider the need for a school-wide intervention that cuts across the elementary (and perhaps) middle school grades.

At the same time, an important contribution of the current study is that we did find empirical support for the accumulated advantages hypothesis, in that the impacts of *INSIGHTS* on ELA test scores were sustained through sixth grade for the students who entered kindergarten with higher levels of academic skills. This finding aligns with recent meta-analytic work by Simonsmeier and colleagues (2018) showing that across different types of early interventions, *less* cognitively demanding interventions showed bigger effects for students with lower baseline skills, and *more* cognitively demanding interventions demonstrated larger benefits for students with higher baseline skills.

Universal SEL programs, although designed to enhance the social-emotional skills of all students in a given context, may introduce some cognitively demanding content to young children (Greenberg & Abenavoli, 2017). For example, in the *INSIGHTS* program, children ages five and six years old learn to identify daily dilemmas and then process the various emotions they feel in these contexts, aligned with their particular temperament profile. They then receive supports for how to problem-solve when faced with these dilemmas, recognizing the role that their emotions and temperament play in that process. SEL programs like *INSIGHTS* are not teaching basic skills like letter and number knowledge, and are instead promoting a range of social-emotional competencies that may require higher-order thinking to build and develop (Paris, 2005; Snow & Matthews, 2016). It may be that the children who are most likely to benefit from these interventions in the longer-term—particularly for standardized test outcomes—are those who begin schooling with the strongest cognitive foundation for learning and immediately benefit from the content.

Strengths, Limitations, and Directions for Future Research

This study has a number of strengths, most notably the experimental design, the ability to access administrative data on students' standardized test scores rather than relying on self- or teacher-reports of academic ability, an explicit focus on low-income students at heightened risk for poorer academic outcomes in the short- and long-term, minimal differential attrition between treatment and control group members, and a focus on longer-term outcomes measured up to five years after the end of the intervention. Even given these strengths, however, there are key limitations. First, we are limited by the baseline data in the number of covariates that we are able to use to assess baseline equivalence. Second, although the study explicitly focuses on a low-income sample of students, the demographics of the study sample do not mirror New York City, thus limiting the external validity of the study findings. In addition, we had limited observable data on the students who did not consent to participate in the study. As such, it is possible that the study results were not generalizable to the broader population of students at the participating schools. Third, although we did have administrative data on students' attendance rates during the academic year, we lacked daily information on attendance. As such, we were unable to determine whether students were in school on the days when the intervention took place, which would yield a more precise measure of program dosage and allow us to test whether the impacts of the intervention varied by a combination of dosage and academic risk. Next, a key outcome from our original impact study was student behavior. We did not have access to reliable administrative data on students' behaviors in this study and thus were unable to determine whether the effects observed on standardized tests were also true for behaviors or whether the impacts on standardized test scores were in some way related to reductions in behavior problems. Relatedly, given our desire to maintain the internal validity of the study afforded by the experimental design, we did not explore mechanisms that could potentially explain why we observed lasting effects of the *INSIGHTS* program for children who entered kindergarten with higher levels of academic skills. A complementary study that our team conducted found that the intervention reduced children's receipt of

special education services through fifth grade, hypothesizing that this might be attributed to reductions in behavior problems. Yet, more work to identify mechanisms underlying these effects is warranted.

Implications for Research, Practice, and Policy

Although the results from this study reflect long-term impacts of just one SEL program—*INSIGHTS*—on a narrow range of standardized test score outcomes, the findings speak to the importance of conducting long-term follow-up of SEL programs implemented in early childhood and elementary school settings. As noted by Taylor and colleagues (2017), there are limited examples of SEL programs that have been evaluated in both the short- and longer-term, making it difficult to understand whether early investments in SEL programming stand to benefit children across time. Researchers working to evaluate SEL programs should put structures in place to allow for longer-term follow-up, including language about follow-up in consent forms and asking parents to provide multiple pieces of contact information to facilitate future data collection activities.

Practitioners and policymakers also stand to learn from the current study. Although social-emotional learning is currently in-demand in schools, there are numerous available programs that schools can decide to implement and the level of evidence about these programs varies, both in the short- and long-term. Work identifying programs that demonstrate lasting impacts on students' academic skills can inform schools and districts about how they could potentially prioritize selecting interventions. Findings from this study also suggest the potential importance of implementing SEL programming throughout the course of students' educational trajectories, rather than solely during one or two grades or developmental stages. Finally, policymakers at district, local, and state levels looking to expand the availability of evidence-based SEL programming can help identify programs that have demonstrated rigorous evidence of impact by accessing studies like ours. Efforts by policymakers to allocate scarce resources toward the sets of SEL programs that do stand to benefit students in the short- and long-term can help support returns to substantial investments in SEL.

Acknowledgments

The authors thank James Kemple and Patricia Chou from the Research Alliance for NYC Schools for all their work in helping the research team access the data used for this study. The authors also thank the NYC Department of Education for sponsoring and supporting this study for the past 13 years.

Funding

The research reported here was conducted as a part of a study funded by Grant R305A160177 from the Institute of Education Sciences to New York University with a subcontract to MDRC. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

References

Akos, P. (2005). The unique nature of middle school counseling. *Professional School Counseling*, *9*(2), 95–103. https://doi.org/10.5330/prsc.9.2.c3172218176248x4

Bailey, D., Duncan, G., Odgers, C., & Yu, W. (2017). Persistence and fadeout in the impacts of child and adolescent interventions. *Journal of Research on Educational Effectiveness*, *10*(1), 7–39. https://doi.org/10.1080/19345747.2016.1232459

Bierman, K. L., Coie, J. D., Dodge, K. A., Greenberg, M. T., Lochman, J. E., McMahon, R. J., & Pinderhughes, E., Conduct Problems Prevention Research Group (2010). The effects of a multi-year universal social–emotional learning program: The role of student and school characteristics. *Journal of Consulting and Clinical Psychology*, *78*(2), 156–168. https://doi.org/10.1037/a0018607

Blair, C., Ursache, A., Greenberg, M., Vernon-Feagans, L., & Family Life Project Investigators. (2015). Multiple aspects of self-regulation uniquely predict mathematics but not letter–word knowledge in the early elementary grades. *Developmental Psychology*, *51*(4), 459–472. https://doi.org/10.1037/a0038813

Brackett, M. A., Rivers, S. E., Reyes, M. R., & Salovey, P. (2012). Enhancing academic performance and social and emotional competence with the RULER feeling words curriculum. *Learning and Individual Differences*, *22*(2), 218–224. https://doi.org/10.1016/j.lindif.2010.10.002

Chess, S., & Thomas, A. (1984). *Origins and evolution of behavior disorders*. Bruner/Mazel.

Chung, S., & McBride, A. M. (2015). Social and emotional learning in middle school curricula: A service learning model based on positive youth development. *Children and Youth Services Review*, *53*, 192–200. https://doi.org/10.1016/j.childyouth.2015.04.008

Collaborative for Academic, Social, and Emotional Learning. (2019). *CASEL state scan*. https://casel.org/state-scan-scorecard-project-2/

Corcoran, R. P., Cheung, A. C. K., Kim, E., & Xie, C. (2018). Effective universal school-based social and emotional learning programs for improving academic achievement: A systematic review and meta-analysis of 50 years of research. *Educational Research Review*, *25*, 56–72. https://doi.org/10.1016/j.edurev.2017.12.001

Diamond, A., & Lee, K. (2011). Interventions shown to aid executive function development in children 4 to 12 years old. *Science*, *333*(6045), 959–964. https://doi.org/10.1126/science.1204529

Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., Pagani, L. S., Feinstein, L., Engel, M., Brooks-Gunn, J., Sexton, H., Duckworth, K., & Japel, C. (2007). School readiness and later achievement. *Developmental Psychology*, *43*(6), 1428–1446. https://doi.org/10.1037/0012-1649.43.6.1428

Durlak, J. A., Weissberg, R. P., Dymnicki, A. B., Taylor, R. D., & Schellinger, K. B. (2011). The impact of enhancing students' social and emotional learning: A meta-analysis of school-based universal interventions. *Child Development*, *82*(1), 405–432. https://doi.org/10.1111/j.1467-8624.2010.01564.x

Eccles, J. S. (2004). *Race as a developmental context for adolescent development. Presidential address*. Society for Research on Adolescence.

Eisenberg, N., Spinrad, T. L., & Valiente, C. (2018). Emotion-related self-regulation and children's social, psychological, and academic functioning. In K. Shigemasu, S. Kuwano, T. Sato, & T. Matsuzawa (Eds.), *Diversity in harmony – insights from psychology: Proceedings of the 31st International Congress of Psychology* (pp. 268–295). John Wiley & Sons.

Enders, C. K., & Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: A new look at an old issue. *Psychological Methods*, *12*(2), 121.

Epstein, M., Atkins, M., Cullinan, D., Kutash, K., & Weaver, K. (2008). Reducing behavior problems in the elementary school classroom. *IES Practice Guide*, *20*(8), 12–22.

Eyberg, S. M., & Pincus, D. (1999). *Eyberg child behavior inventory and Sutter-Eyberg student behavior inventory-revised: Professional manual*. Psychological Assessment Resources.

Fitzmaurice, G. M., Laird, N. M., & Ware, J. H. (2004). *Applied longitudinal analysis*. Wiley.

Graham, J. W., Olchowski, A. E., & Gilreath, T. D. (2007). How many imputations are really needed? Some practical clarifications of multiple imputation theory. *Prevention Science: The*

Official Journal of the Society for Prevention Research, *8*(3), 206–213. https://doi.org/10.1007/s11121-007-0070-9

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

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

Hernandez, D. J. (2011). *Double jeopardy: How third-grade reading skills and poverty influence high school graduation*. Annie E. Casey Foundation.

Hulleman, C. S., & Cordray, D. S. (2009). Moving from the lab to the field: The role of fidelity and achieved relative intervention strength. *Journal of Research on Educational Effectiveness*, *2*(1), 88–110. https://doi.org/10.1080/19345740802539325

Ialongo, N., Rogosch, F. A., Cicchetti, D., Toth, S. L., Buckley, J., Petras, H., & Neiderhiser, J. (2006). A developmental psychopathology approach to the prevention of mental health disorders. In D. Cicchetti & D. Cohen (Eds.), *Developmental psychopathology (2nd ed.): Theory and Method (Vol. 1)* (pp. 968–1018). Wiley.

Jones, S. M., Brown, J. L., & Aber, J. L. (2011). Two-year impacts of a universal school-based social-emotional and literacy intervention: An experiment in translational developmental research. *Child Development*, *82*(2), 533–554. https://doi.org/10.1111/j.1467-8624.2010.01560.x

Kellam, S. G., Brown, C. H., Poduska, J. M., Ialongo, N. S., Wang, W., Toyinbo, P., Petras, H., Ford, C., Windham, A., & Wilcox, H. C. (2008). Effects of a universal classroom behavior management program in first and second grades on young adult behavioral, psychiatric, and social outcomes. *Drug and Alcohol Dependence*, *95*, S5–S28. https://doi.org/10.1016/j.drugalcdep.2008.01.004

Legewie, J., Farley, C., & Stewart, K. (2019). *Aggressive policing and academic outcomes: Examining the impact of police "surges" in NYC students' home neighborhoods* (Policy brief). https://research.steinhardt.nyu.edu/scmsAdmin/media/users/ks191/policing/Aggressive_Policing_and_Academic_Outcomes_Brief.pdf

Li, M., Murphy, P. K., Wang, J., Mason, L. H., Firetto, C. M., Wei, L., & Chung, K. S. (2016). Promoting reading comprehension and critical–analytic thinking: A comparison of three approaches with fourth and fifth graders. *Contemporary Educational Psychology*, *46*, 101–115. https://doi.org/10.1016/j.cedpsych.2016.05.002

Lipsey, M. W., Puzio, K., Yun, C., Hebert, M. A., Steinka-Fry, K., Cole, M. W., Roberts, M., Anthony, K. S., & Busick, M. D. (2012). *Translating the statistical representation of the effects of education interventions into more readily interpretable forms* (NCSER 2013–3000). U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance.

Low, S., Smolkowski, K., Cook, C., & Desfosses, D. (2019). Two-year impact of a universal social-emotional learning curriculum: Group differences from developmentally sensitive trends over time. *Developmental Psychology*, *55*(2), 415–433. https://doi.org/10.1037/dev0000621

Martorell, P., & Mariano, L. T. (2018). The causal effects of grade retention on behavioral outcomes. *Journal of Research on Educational Effectiveness*, *11*(2), 192–216. https://doi.org/10.1080/19345747.2017.1390024

Masten, A. S., & Cicchetti, D. (2010). Developmental cascades. *Development and Psychopathology*, *22*(3), 491–495. https://doi.org/10.1017/S0954579410000222

McClelland, M. M., Tominey, S. L., Schmitt, S. A., & Duncan, R. (2017). Social emotional learning interventions in early childhood. *The Future of Children*, *27*(1), 33–48. https://doi.org/10.1353/foc.2017.0002

McClowry, S. G. (2002). The temperament profiles of school-age children. *Journal of Pediatric Nursing*, *17*(1), 3–10. https://doi.org/10.1053/jpdn.2002.30929

McClowry, S. G. (2014). *Temperament-based elementary classroom management*. Rowman & Littlefield.

McCoy, D. C., Jones, S., Roy, A., & Raver, C. C. (2018). Classifying trajectories of social-emotional difficulties through elementary school: Impacts of the Chicago School Readiness Project. *Developmental Psychology*, *54*(4), 772–787. https://doi.org/10.1037/dev0000457

McKown, C. (2017). Social-emotional assessment, performance, and standards. *The Future of Children*, *27*(1), 157–178. https://doi.org/10.1353/foc.2017.0008

Mikami, A. Y., Ruzek, E. A., Hafen, C. A., Gregory, A., & Allen, J. P. (2017). Perceptions of relatedness with classroom peers promote Adolescents' Behavioral Engagement and Achievement in Secondary School. *Journal of Youth and Adolescence*, *46*(11), 2341–2354. https://doi.org/10.1007/s10964-017-0724-2

O'Connor, E. E., Cappella, E., McCormick, M. P., & McClowry, S. G. (2014). Enhancing the academic development of shy children: A test of the efficacy of INSIGHTS. *School Psychology Review*, *43*(3), 239,259.

O'Connor, E. E., & McCartney, K. (2007). Examining teacher-child relationships and achievement as part of an ecological model of development. *American Educational Research Journal*, *44*(2), 340–369. https://doi.org/10.3102/0002831207302172

Paris, S. G. (2005). Reinterpreting the development of reading skills. *Reading Research Quarterly*, *40*(2), 184–202. https://doi.org/10.1598/RRQ.40.2.3

Rabiner, D. L., & Coie, J. D. (2000). Early attention problems and children's reading achievement: A longitudinal investigation. The Conduct Problems Prevention Research Group. *Journal of the American Academy of Child and Adolescent Psychiatry*, *39*(7), 859–867. https://doi.org/10.1097/00004583-200007000-00014

Rabiner, D. L., Carrig, M., & Dodge, K. A. (2016). Attention problems and academic achievement: Do persistent and earlier-emerging problems have more adverse long-term effects? *Journal of Attention Disorders*, *20*(11), 946–957. https://doi.org/10.1177/1087054713507974

Rabiner, D. L., Godwin, J., & Dodge, K. A. (2016). Predicting academic achievement and attainment: The contribution of early academic skills, attention difficulties, and social competence. *School Psychology Review*, *45*(2), 250–267. https://doi.org/10.17105/SPR45-2.250-267

Raver, C. C. (2002). Emotions matter: Making the case for the role of young children's emotional development for early school readiness. *Social Policy Report*, *16*(3), 1–18. https://doi.org/10.1002/j.2379-3988.2002.tb00041.x

Raver, C. C., Jones, S., 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

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

Rothbart, M. K., & Bates, J. E. (2006). Temperament. In W. Damon, R. Lerner, & N. Eisenberg (Eds.), *Handbook of child psychology: Vol. 3. Social, emotional, and personality development* (6th ed., pp. 99–166). Wiley.

Rudolph, K. D., Lambert, S. F., Clark, A. G., & Kurlakowsky, K. D. (2001). Negotiating the transition to middle school: The role of self-regulatory processes. *Child Development*, *72*(3), 929–946. https://doi.org/10.1111/1467-8624.00325

Scott, M. A., Simonoff, J. S., & Marx, B. D. (Eds.). (2013). *The SAGE handbook of multilevel modeling*. SAGE.

Sibley, E., Thomson, D., Longo, F., & Dearing, E. (2019). Student learning and development in economically disadvantaged family and neighborhood contexts. In S. B. Sheldon & T. A. Turner-Vorbeck (Eds.), *The Wiley handbook of family, school, and community relationships in education* (pp. 399–417). John Wiley & Sons, Inc.

Simmons, R. G., & Blyth, D. A. (1987). *Moving into adolescence: The impact of pubertal change and school context*. Aldine de Gruyter.

Simonsmeier, B., Flaig, M., Deiglmayr, A., Schalk, L., & Schneider, M. (2018). *Domain-specific prior knowledge and learning: A meta-analysis*. ZPID (Leibniz Institute for Psychology Information).

Snow, C. E., & Matthews, T. J. (2016). Reading and language in the early grades. *The Future of Children*, *26*(2), 57–74. https://doi.org/10.1353/foc.2016.0012

Social and Character Development Research Consortium (2010). *Efficacy of schoolwide programs to promote social and character development and reduce problem behavior in elementary school children* (NCER 2011–2001). National Center for Education Research, Institute of Education Sciences, U.S. Department of Education.

Steinberg, L. (2005). Cognitive and affective development in adolescence. *Trends in Cognitive Sciences*, *9*(2), 69–74. https://doi.org/10.1016/j.tics.2004.12.005

Stringer, K. (2019, July 17). *SEL in the house: Democrats approve millions in landmark federal funding for social-emotional learning in bill that now faces test in Senate*. http://laschoolreport.com

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

Véronneau, M. H., & Dishion, T. J. (2011). Middle school friendships and academic achievement in early adolescence: A longitudinal analysis. *The Journal of Early Adolescence*, *31*(1), 99–124. https://doi.org/10.1177/0272431610384485

von Hippel, P. T., Workman, J., & Downey, D. B. (2018). Inequality in reading and math skills forms mainly before kindergarten: A replication, and partial correction, of 'Are schools the great equalizer?' *Sociology of Education*, *91*(4), 323–357. https://doi.org/10.1177/0038040718801760

Wang, M. T., Kiuru, N., Degol, J. L., & Salmela-Aro, K. (2018). Friends, academic achievement, and school engagement during adolescence: A social network approach to peer influence and selection effects. *Learning and Instruction*, *58*, 148–160. https://doi.org/10.1016/j.learninstruc.2018.06.003

Watts, T. W., Gandhi, J., Ibrahim, D. A., Masucci, M. D., & Raver, C. C. (2018). The Chicago School Readiness Project: Examining the long-term impacts of an early childhood intervention. *PLoS One*, *13*(7), e0200144. https://doi.org/10.1371/journal.pone.0200144

Webster-Stratton, C., Reid, M., & Stoolmiller, M. (2008). Preventing conduct problems and improving school readiness: Evaluation of the Incredible Years Teacher and Child Training Programs in high-risk schools. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *49*(5), 471–488. https://doi.org/10.1111/j.1469-7610.2007.01861.x

Weissberg, R. P., Durlak, J. A., Domitrovich, C. E., & Gullotta, T. P. (2015). Social and emotional learning: Past, present, and future. In J. A. Durlak, C. E. Domitrovich, R. P. Weissberg, & T. P. Gullotta (Eds.), *Handbook for social and emotional learning: Research and practice* (pp. 3–19). Guilford.

Weschler, D. (1989). *Weschler Preschool and Primary Scale of Intelligence–Revised*. San Antonio, TX: The Psychological Corporation.

What Works Clearinghouse (WWC). (2017). *WWC procedures handbook (Version 4.0)*. https://ies.ed.gov/ncee/wwc/Docs/ReferenceResources/wwc_procedures_handbook_v4_draft.pdf

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock Johnson psychoeducational battery-third edition*. Riverside Publishing.

Yang, C., Bear, G. G., & May, H. J. (2018). Multilevel associational between school-wide social-emotional learning approach and student engagement across elementary, middle, and high schools. *School Psychology Review*, *47*(1), 45–61. https://doi.org/10.17105/SPR-2017-0003.V47-1

Yang, W., Datu, J. A. D., Lin, X., Lau, M. M., & Li, H. (2019). Can early childhood curriculum enhance social-emotional competence in low-income children? A meta-analysis of the educational effects. *Early Education and Development*, *30*(1), 36–59. https://doi.org/10.1080/10409289.2018.1539557

Zeidner, M., Roberts, R. D., & Matthews, G. (2002). Can emotional intelligence be schooled? A critical review. *Educational Psychologist*, *37*(4), 215–231. https://doi.org/10.1207/S15326985EP3704_2

Zins, J. E., Weissberg, R. P., Wang, M. C., & Walberg, H. J. (Eds.). (2004). *Building school success through social and emotional learning*. Teachers College Press.