Social-Emotional Learning and Academic Achievement: Using Causal Methods to Explore Classroom-Level Mechanisms

Meghan P. McCormick
MDRC
Elise Cappella
Erin E. O'Connor
Sandee G. McClowry
New York University

Social-emotional learning (SEL) programs have demonstrated positive effects on children's social-emotional, behavioral, and academic outcomes, as well as classroom climate. Some programs also theorize that program impacts on children's outcomes will be partially explained by improvements in classroom social processes, namely classroom emotional support and organization. Yet there is little empirical evidence for this hypothesis. Using data from the evaluation of the SEL program INSIGHTS*, this article tests whether assignment to* INSIGHTS *improved low-income kindergarten and first grade students' math and reading achievement by first enhancing classroom emotional support and organization. Multilevel regression analyses, instrumental variables estimation, and inverse probability of treatment weighting (IPTW) were used to conduct quantitative analyses. Across methods, the impact of* INSIGHTS *on math and reading achievement in first grade was partially explained by gains in both classroom emotional support and organization. The IPTW method revealed that the program impact on reading achievement in first grade was partially explained through an improvement in classroom organization. Implications for research and practice are discussed.*

Keywords: *social-emotional learning, academic achievement, elementary school, child development, education*

Early social-emotional competencies, such as behavioral regulation, attentional skills, and the ability to problem solve, are critical to children's academic outcomes (Blair, 2002; Diamond & Lee, 2011). Such findings have prompted the development and implementation of social-emotional learning (SEL) programs, which are a type of school-based preventive intervention explicitly designed to foster children's academic skills by supporting their social-emotional and behavioral development (Durlak, Weissberg, Dymnicki, Taylor, & Schellinger, 2011). Given evidence that children raised in poverty are more likely to start school with lower levels of social-emotional skills (Ursache, Blair, & Raver, 2012), numerous SEL program models have been implemented in underresourced elementary schools. Although some SEL programs have improved academic achievement (e.g., Jones, Brown, & Aber, 2011; Raver et al., 2011), others have shown no impact on children's academic outcomes (Catalano, Berglund, Ryan, Lonczak, & Hawkins, 2002; Greenberg et al., 2003; Spivak & Farran, 2014). Moreover, because most evaluations of SEL programs have used intent-to-treat research designs, little is known about how SEL programs work. Understanding more about mechanisms of SEL

programs would entail identifying and testing how interventions improve proximal outcomes (or mediators), which in turn link to more distal outcomes like academic achievement. This information can help determine the critical factors explaining variation in SEL programs' ability to support students' academic skills, and thus inform future development and scaling of such interventions.

A number of universal SEL programs—which target curricula toward the teacher and/or all students in a school or classroom—hypothesize initial improvements in classroom quality and within-classroom social interactions. These classroom-level impacts are theorized to mediate program impacts on children's social-emotional and academic outcomes (e.g., Cappella et al., 2012; Hawkins, Kosterman, Catalano, Hill, & Abbott, 2005). In separate studies, some SEL programs have demonstrated positive effects at both the classroom and child levels (Brown, Jones, LaRusso, & Aber, 2010; Raver et al., 2011; Rivers, Brackett, Reyes, Elbertson, & Salovey, 2013). Yet few studies, if any, have rigorously tested whether improvements in critical dimensions of classroom quality explain more distal impacts of SEL programs on children's academic achievement. Thus, there is a lack of

empirical evidence for the hypothesized theories of change that many classroom-based SEL program espouse. The current study aims to address this critical gap in the literature by examining the classroom-level mechanisms explaining impacts of one SEL program—*INSIGHTS Into Children's Temperament*—on low-income urban children's reading and math achievement during kindergarten and first grade. Understanding the explanatory mechanisms linking SEL intervention to student achievement will provide crucial information for further program research, development, implementation, and scale-up.

Social-Emotional Learning and Achievement in Early Schooling

SEL programs broadly aim to enhance an interrelated set of cognitive, emotional, and behavioral skills regarded as foundational for academic performance. Skills targeted by SEL programs include the recognition and management of emotions, appreciating others' perspectives, initiating and maintaining positive relationships, and using critical thinking skills to make responsible decisions and handle interpersonal situations (Zins & Elias, 2006). Such competencies promote children's engagement in instructional activities and the classroom setting that, in turn, enhance academic achievement (Eisenberg, Valiente, & Eggum, 2010). Children who successfully develop core social-emotional and behavioral competencies (e.g., emotional and behavioral regulation, attention skills) in preschool are more likely to successfully navigate the transition to elementary school (Fantuzzo et al., 2007; Rimm-Kaufman, Curby, Grimm, Nathanson, & Brock, 2009). Notably, evidence from longitudinal studies suggests that links between social-emotional skills and academic achievement in early schooling are causal (Raver, 2002).

Universal SEL programs include curricula designed to promote social-emotional skills among all children in a classroom. Some programs also integrate schoolwide and family components that target services at multiple developmental contexts (Greenberg et al., 2003). Most effective SEL programs are implemented by providing professional development (PD) and training to teachers. It is anticipated that teachers will then use the strategies taught in the PD sessions in their classrooms. Such implementation is often manualized in the curriculum and can be embedded into instruction or stand alone as its own program. Work by Abry, Hulleman, and Rimm-Kaufman (2015) done on one SEL program—the Responsive Classroom approach—has found that fidelity to the intervention model is important for promoting children's academic outcomes. However, there is very little research identifying the specific training, coaching, or PD critical for enhancing fidelity or impacts.

In addition, there are no data to determine the exact number of SEL programs being implemented across the country.

However, the Collaborative for Academic and Social Emotional Learning (CASEL) has identified 19 effective program models, meaning there is some rigorous evidence for their ability to improve elementary school children's social-emotional, behavioral, and/or academic outcomes (CASEL, 2014). Such programs are typically developed and implemented in preschool and early elementary school settings by community-based organizations or by academic partners (Durlak et al., 2011). Rigorously evaluated programs collect implementation information on dosage and program fidelity from teachers, students, and program staff. In addition, states vary in their implementation of SEL programs. Although all 50 states require schools to implement programs and policies related to students' social-emotional development, no states mandate use of specific SEL program models (CASEL, 2014).

Yet several rigorously evaluated SEL programs, tested in low-income pre-K and elementary schools, have been successful in improving low-income children's social-emotional skills (e.g., 4Rs: Jones et al., 2011; CSRP: Raver et al., 2011; Incredible Years: Webster-Stratton, Reid, & Stoolmiller, 2008) and academic outcomes (e.g., Brackett, Rivers, Reyes, & Salovey, 2012; Jones et al., 2011; Raver et al., 2011; Webster-Stratton et al., 2008). Perhaps most compelling, in a meta-analysis of 213 programs, Durlak and colleagues (2011) found that, across all participants, SEL participants evidenced an 11-percentile-point gain in academic achievement postintervention as compared to children in the control group.

Other evaluations, such as the Institute of Education Sciences' recent study of seven Social and Character Development Programs, have found no direct effects on elementary school students' academic outcomes (Social and Character Development Research Consortium, 2010). Given the current policy focus in K–12 education on testing and achievement, policy makers and practitioners are particularly interested in whether school-based programs do improve academic outcomes. Inconsistent findings related to academic outcomes, however, are difficult to understand and potentially misleading because few studies of SEL programs have even considered academic outcomes. In Durlak and colleagues' (2011) review, only 16% of the studies collected information on academic achievement (i.e., standardized tests, grades). There are numerous explanations for inconsistent findings regarding impacts on academic outcomes. For example, SEL programs' theories of change hypothesize distal effects on academic outcomes. Yet most evaluations only examine immediate effects, measured in the short-term postintervention. As such, it may be that longer term follow-up is needed to examine impacts on academic outcomes. Alternatively the effects of interventions aimed at enhancing social-emotional development simply may not "spill over" to improve achievement when measured with traditional assessments. Even given these possibilities, a critical, understudied

limitation in understanding null effects on academic outcomes is the lack of research examining the proximal pathways through which SEL programs improve more distal outcomes. Failing to test program mechanisms limits the field's ability to understand variation in the efficacy of SEL interventions and truly unpack the key ingredients that are critical for promoting longer term impacts on achievement.

Quality of Classroom Interactions, SEL Programs, and Achievement

Ecological and systems theories offer an additional perspective on school-based interventions by specifying the multiple levels of the ecology in which children are embedded (Bronfenbrenner & Morris, 1998). A systems framework for social settings relates these ideas to classrooms, positing that the norms and relationships extant in the classroom setting are central mechanisms through which children develop in school (Tseng & Seidman, 2007). If an SEL program can improve the overall quality of relationships between teachers, students, and peers in the classroom setting, students may be more likely to engage in classroom activities, listen to their teacher, and ask teachers and peers for help with academic tasks. These behaviors will likely support positive academic outcomes. A systems framework is helpful in the current article because several SEL programs (e.g., Brackett et al., 2012; Brown et al., 2010; Cappella et al., 2012; Raver et al., 2011) were indeed designed to induce changes in the pattern of social interactions between students and adults, and between students and peers in classroom settings.

In this vein, researchers have developed measures of classroom social processes, in part to determine whether programs do improve classroom interactions (see Pianta, La Paro, & Hamre, 2008). Two particular dimensions of classroom social processes that have been extensively examined are classroom emotional support and organization (e.g., Brown et al., 2010; Cappella et al., in press; Rivers et al., 2013). Classroom emotional support and organization are the two classroom-level variables of interest in the current study. According to the framework of Pianta, La Paro, et al. (2008), emotionally supportive classrooms exhibit teaching practices that are highly responsive to students' needs and interests. Classrooms with high levels of organization show effective strategies for behavior management, maximization of learning time, and engagement in successful implementation of academic activities (Pianta, La Paro, et al., 2008). In a series of studies with mostly White and/or middle-income samples, both classroom emotional support and organization have been linked to students' math and reading achievement in early elementary school (Pianta, Belsky, Vandergrift, Houts, & Morrison, 2008; Ponitz, McClelland, Mathews, & Morrison, 2009; Rimm-Kaufman et al., 2009; Roorda, Koomen, Spilt, & Oort, 2011). Recent work focused on preschool settings indicates similar associations for lower income, racial/ethnic minority pre-K students (e.g., Hamre, Hatfield, Pianta, & Jamil, 2014).

Rigorous evidence from randomized trials demonstrates that SEL programs can improve both emotional support and organization in low-income and/or largely racial/ethnic minority elementary schools (Brown et al., 2010; Hagelskamp, Brackett, Rivers, & Salovey, 2013; Webster-Stratton et al., 2008). Additional work suggests that these SEL programs also improved low-income urban students' academic skills and achievement (4Rs: Jones et al., 2011; Responsive Classroom: Abry et al., 2013; PATHS: Bierman et al., 2014; Incredible Years: Webster-Stratton et al., 2008; RULER: Brackett et al., 2012). Yet few studies have empirically examined whether improvements in domains of classroom quality are pathways through which SEL programs enhance student academic outcomes in low-income urban schools. Doing so may provide insight into why some SEL programs appear to improve students' achievement, and others do not. Ecological and systems theories suggest that the salient contexts in which children develop and the social processes inherent in those settings are critical to children's adaptive outcomes. The call now is to examine data on extant SEL program evaluations, identify which programs improved children's academic skills, and find evidence for *how* these programs worked.

INSIGHTS Into Children's Temperament

Broadly, *INSIGHTS* is a preventive school-based intervention designed to enhance the development of low-income primary grade students at-risk for academic and behavioral difficulties. As depicted in its general theory of change (see Appendix A), *INSIGHTS* is a comprehensive temperament-based intervention with programs for teachers, parents, and students. The intervention's framework integrates theory, research, and clinical strategies that support the academic learning context. Using a temperament interventionist perspective, *INSIGHTS* aims to enhance goodness of fit, or the match between the environment and the child's temperament. *INSIGHTS* features four empirically based temperament profiles as exemplars (McClowry, 2002; McClowry et al., 2013): Coretta the Cautious who is shy, Gregory the Grumpy who is high maintenance (i.e., low in task persistence, and high in negative reactivity and motor activity), Fredrico the Friendly who is social and eager to try, and Hilary the Hard Worker who is industrious (i.e., high in task persistence, and low in negative reactivity and motor activity).

The curriculum for the parent and teacher programs has three parts. In Part 1, "The *3 Rs* of Child Management: Recognize, Reframe, and Respond," participants are taught to *recognize* the unique qualities that children exhibit as an expression of their temperament. Intentionality, the belief that a child consciously misbehaves, is reduced when participants recognize that many reactions to specific situations

are related to one's temperament. Participants are encouraged to *reframe* their perceptions with the understanding that every temperament has strengths and challenges. They also learn that while temperament is not amenable to change, parent and teacher *responses* are and can, in turn, influence the behavior of children. Recognition and acceptance of a child's temperament, however, does not imply permissiveness. In Part 2, "Gaining Compliance," temperament-based management strategies are implemented to improve children's behavior. Parents and teachers are assisted in replacing negative patterns of interaction with child management strategies that are matched to specific types of temperaments. Finally, Part 3 focuses on strategies that support children in becoming more socially competent, particularly when encountering situations that are temperamentally challenging. Appendix B provides a brief overview of the *INSIGHTS* curriculum.

The content of the parent and teacher programs is delivered over a 10-week period during 2-hour, weekly facilitated sessions using a structured curriculum that includes didactic content, professionally produced videotaped vignettes (25 for parents, 27 for teachers), session handouts, and group discussions. The scripted manual for the parent and teacher programs is 178 pages and includes objectives, activities for each session, and participant handouts.

During the same 10-week period, the participating children and their classmates participate in a 45-minute classroom component. Children are introduced to four puppets that represent the four temperament profiles. The facilitator and participating teacher engage the children in the content using the puppets. Workbooks and vocabulary flash cards are also incorporated into the sessions. The children are taught that, based on temperament, some situations are easy for individuals while others may be challenging. Professionally produced vignettes guide the children in problem solving daily dilemmas with the help of the puppets and their classmates. Children also identify dilemmas in their own lives and apply the same problem-solving process to resolving these dilemmas.

The *INSIGHTS* theory of change (see Appendix A) theorizes that following implementation, the program will first enhance classroom emotional support and organization. Next, children will show improvements in social-emotional and academic skills. A recent evaluation did reveal positive impacts of *INSIGHTS* across kindergarten and first grade on math and reading skills (O'Connor, Cappella, McCormick, & McClowry, 2014). Other findings include short-term effects on classroom emotional support in kindergarten and first grade, and on classroom organization in first grade (Cappella et al., in press). The rich data from this larger study provide the opportunity to refine results by examining potential pathways through which *INSIGHTS* improved student achievement outcomes. To limit the scope of this article, we focus on linking classroom-level mechanisms to study achievement outcomes and do not consider more distal social-emotional outcomes. In this vein, we are interested in answering two fairly broad research questions:

1. Did the *INSIGHTS* program improve classroom emotional support, thus leading to improvements in student math and reading achievement in kindergarten and first grade?
2. Did the *INSIGHTS* program improve classroom organization, thus leading to improvements in student math and reading achievement in kindergarten and first grade?

In answering these two questions, we use a series of rigorous quantitative methods (multilevel regression, instrumental variables estimation, inverse probability of treatment weighting) that aim to maintain a high level of internal validity when assessing mechanisms in the context of a randomized trial. Thus, three statistical methods will be used to examine the two research questions. The broader goal of the study is to (a) provide information on research methodology that enhances causal inference for noncausal questions in randomized trials, (b) increase evidence for multilevel SEL program theories of change, and (c) inform future SEL program design and implementation.

Method

Participants and Setting

This study took place in 22 public elementary schools in a large city. In all, 120 teachers and 435 parent/child dyads were participants. Teachers were mainly female (94.2%) and identified as Hispanic or Latino (11.9%), Black or African American (56.4%), White (24.3%), or mixed race/other (7%). All teachers reported having a bachelor's degree; 96% had a master's degree. All classrooms were classified as regular education, with an average of 16.57 students ($SD = 3.54$).

The large majority of students were 5 years old when they began the study ($M = 5.38$, $SD = 0.61$). Half (52%) of the students were male, 87% qualified for free or reduced-price lunch, 75% were Black, non-Hispanic, 16% were Hispanic, non-Black, and the remaining students were biracial. Approximately 28% of students' parents had education levels less than a high school degree; 26% had at least a high school degree or GED diploma; 24% had at least some college experience; and the remaining 22% had graduated from a 4-year college. Descriptive statistics are listed in Table 1.

Students enrolled in the study were similar to the other students at the schools who were invited to participate in the study but did not enroll. Participating schools had high percentages of students who were racial/ethnic minorities (Black, $M = 0.77$, $SD = 0.13$; Hispanic, $M = 0.40$, $SD = 0.27$) and eligible for free/reduced lunch ($M = 0.80$, $SD = 0.16$).

TABLE 1

Descriptive Statistics for Variables of Interest, Pretest and Posttest

| | Kindergarten pretest (T1) | | | | Kindergarten posttest (T2) | | | | First grade pretest (T3) | | | | First grade posttest (T4) | | | |
| | Treatment | | Control | | Treatment | | Control | | Treatment | | Control | | Treatment | | Control | |
Variable	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
Student-level variables																
Sustained attention	46.14	12.60	45.58	12.85	51.34	11.99	54.59	9.02	57.95	9.48	56.05	8.74	61.44	9.03	60.54	9.45
Behavior problems	2.28	1.24	2.15	1.19	2.48	1.39	2.15	1.02	2.18	1.14	2.26	1.03	2.28	1.36	2.46	1.42
Reading achievement	16.74	7.20	17.76	7.15	20.51	7.56	23.11	7.95	23.65	9.20	26.73	8.37	33.54	9.07	33.57	8.11
Math achievement	14.46	4.69	14.45	5.24	16.99	4.46	17.48	4.83	18.46	4.86	19.80	3.87	23.34	4.62	23.17	3.86
Negative reactivity	2.92	0.76	2.86	0.79	—	—	—	—	—	—	—	—	—	—	—	—
Task persistence	3.79	0.72	3.72	0.73	—	—	—	—	—	—	—	—	—	—	—	—
Activity	2.89	0.90	2.82	0.90	—	—	—	—	—	—	—	—	—	—	—	—
Withdrawal	2.44	0.76	2.48	0.80	—	—	—	—	—	—	—	—	—	—	—	—
Child female	0.47	—	0.49	—	—	—	—	—	—	—	—	—	—	—	—	—
Child Black	0.77	—	0.75	—	—	—	—	—	—	—	—	—	—	—	—	—
Child Hispanic	0.17	—	0.14	—	—	—	—	—	—	—	—	—	—	—	—	—
Eligible free/reduced lunch	0.84	—	0.83	—	—	—	—	—	—	—	—	—	—	—	—	—
Classroom-level variables															—	—
Emotional support	4.55	0.95	5.03	0.84	4.59	0.86	4.91	0.92	4.66	0.87	4.65	0.84	4.90	0.81	4.46	0.69
Classroom organization	4.17	1.19	4.40	0.91	3.85	0.95	4.26	1.09	4.40	0.88	4.37	0.91	4.45	0.86	4.01	0.68
Teacher years teaching	13.91	7.52	12.62	7.56	—	—	—	—	14.40	8.22	13.90	11.61	—	—	—	—
# of adults in class	1.38	0.63	1.59	0.67	—	—	—	—	1.26	0.34	1.48	0.64	—	—	—	—
Class size	15.29	3.15	15.87	3.01	—	—	—	—	17.01	4.19	18.20	2.99	—	—	—	—
School-level variables																
% Black	83.63	7.78	74.63	16.89	—	—	—	—	—	—	—	—	—	—	—	—
% Hispanic	50.80	27.89	37.63	25.68	—	—	—	—	—	—	—	—	—	—	—	—
% eligible free lunch	81.43	12.63	78.50	18.78	—	—	—	—	—	—	—	—	—	—	—	—
Average attendance																
School size	499	130	501	194	—	—	—	—	—	—	—	—	—	—	—	—

Schools had an average attendance rate of 86.26% ($SD = 0.19$) and size of 465 total students ($SD = 158.46$). Schools in this study had higher levels of eligibility for free/reduced lunch and lower state test scores than schools in the city taken as a whole.

Procedure

Principals at 23 elementary schools made a 2-year commitment to participate in the study. Prior to randomization, however, one school withdrew during a principal transition. Recruitment of the kindergarten teachers began each September. First grade teachers were recruited from the same schools. In all, 96% of the kindergarten and first grade teachers consented to participate, with no attrition. Teachers reported on student behaviors, academic competencies, and relationships for each participating student and received $50 in gift cards for classroom supplies to thank them for their time. Parents from the participating kindergarten teachers' classrooms were recruited at school in September and October. Parents reported on demographic characteristics, child temperament, and family involvement. Parents received a $20 gift card to thank them for their time. Student assent was then acquired.

Significant efforts were made to recruit a representative group of students for the study within each participating classroom. Recruitment efforts took place over 6 to 8 weeks in each year of the study. The number of students in each class who enrolled ranged from 4 to 10. Although some parents did consent to participate early in the recruitment period, all possible efforts were made to recruit additional parents until data collection was scheduled to begin. As noted above, students enrolled in the study were similar to the students at the schools who chose not to participate in the study.

Random Assignment

Schools were the unit of random assignment. After baseline data were collected in kindergarten, a random numbers table was used to randomly assign schools to *INSIGHTS* or a supplemental reading program (i.e., comparison group). Of the 22 schools, 11 were randomized to *INSIGHTS* ($n = 225$ students; $n = 57$ teachers) and 11 hosted the supplemental reading program, which served as the attention-control

condition (n = 210 students; n = 63 teachers). As discussed in a future section, random assignment was largely successful in creating comparable treatment and comparison groups (O'Connor et al., 2014). Longitudinal multi-informant data (e.g., administrative data, direct assessment, teacher report, parent report) were collected from schools, teachers, classrooms, parents, and students in the fall and spring of kindergarten (T1 and T2; kindergarten pre- and posttest) and the fall and spring of first grade (T3 and T4; first grade pre- and posttest).

INSIGHTS Intervention Procedures

INSIGHTS included (a) teacher sessions, (b) parent sessions, and (c) universal classroom sessions. Teacher and classroom sessions were implemented within the regular school day. Parent sessions were held at the school, typically after school. Teachers and parents attended 10 weekly 2-hour facilitated sessions based on a structured curriculum that included didactic content and professionally produced vignettes as well as handouts and group activities. Parents received $20 and teachers received PD credit and $40 gift cards for each session they attended. During the same 10 weeks, the classroom program was delivered in 45-minute lessons to all students in participating classrooms. The curriculum materials included puppets, workbooks, flash cards, and videotaped vignettes and aimed to help students resolve challenging dilemmas at home and school.

Facilitator Training. Facilitators were selected based on their professional experience. The eight facilitators were graduate students in psychology, education, and educational theater from varied racial/ethnic backgrounds. The facilitators first attended a graduate-level course to learn the theory and research underlying the intervention. Then they learned how to use the intervention materials from an experienced facilitator. Each facilitator conducted the full intervention (teacher, parent, and student) in the schools to which she or he was assigned.

Intervention Fidelity. Facilitators followed scripts, used material checklists, documented sessions, and received ongoing training and supervision. Deviations were discussed weekly in meetings with the program developer. Supervision focused on challenges related to conducting sessions, implementation logistics, and participant concerns. Parent and teacher sessions were videotaped and reviewed for coverage of content and effective facilitation. Fidelity coding, conducted by an experienced clinician, revealed that 94% of the curriculum was adequately covered in the teacher sessions; 92% was covered in parent sessions.

Attention-Control Condition. Schools not assigned to *INSIGHTS* participated in a 10-week supplemental reading program after school for children whose parents consented. Teachers and parents attended two 2-hour workshops in which reading coaches provided materials and presented strategies to enhance literacy. Parents received $20 and teachers received PD credit and $40 for resources for each workshop. Of the children enrolled, 24% participated in the full 10 sessions; an additional 19% took part in 8 or 9 sessions. In all, 30% of parents and 83% of teachers attended both sessions. Reading program facilitators had weekly meetings with the project director to ensure that all components were implemented. Curriculum fidelity was high: 95% to 100% of topics were covered across the 10-week program.

Measures

Below, we provide details about the outcomes and hypothesized mechanisms, as well as classroom- and child-level covariates. Assignment to *INSIGHTS* was dummy coded so that 1 = INSIGHTS *participant*, 0 = *attention-control participant* (explained in more detail below). All the measures that we report on in this section were also collected as part of the larger intent-to-treat evaluation of *INSIGHTS* (see O'Connor et al., 2014).

Outcome Variables. Reading achievement and math achievement were assessed using raw scores from the Letter-Word Identification and Applied Problems subtests of the Woodcock–Johnson III Tests of Achievement, Form B (Woodcock, McGrew, & Mather, 2001). The Letter-Word ID subtest assesses letter naming and word decoding skills by asking children to identify a series of letters and words presented in isolation. The Applied Problems subtest assesses children's counting skills and the ability to analyze and solve mathematical word problems presented orally. The WJ-R is a nationally normed and widely used achievement test with demonstrated internal consistency and validity (e.g., Wechsler, 1989; Woodcock & Johnson, 1989). Baseline levels of reading and math achievement were also used as confounding covariates in this article.

Mediator Variables. Classroom emotional support and classroom organization were observed with the Classroom Assessment Scoring System (CLASS; Pianta, La Paro, et al., 2008). The CLASS measures three domains: emotional support, classroom organization, and instructional support (Hamre et al., 2013). Emotional support includes four dimensions of teacher practices: positive climate, negative climate, teacher sensitivity, and regard for student perspectives. Classroom organization includes three dimensions: productivity, behavior management, and instructional learning formats. The CLASS has a third domain—instructional support—which describes teaching behaviors that enhance the ability of students to engage in higher level thinking, integrate knowledge across disciplines, and apply knowledge in

real-world contexts. Instructional support is not considered in this study, however, because *INSIGHTS* was not designed to improve this component of classroom interactions. Hamre and colleagues (2013) reviewed the three-factor structure of the CLASS across 4,035 classrooms in the preschool and elementary school grades. Results of these analyses provided evidence that the CLASS's three-factor latent structure provided a better fit to observational data than alternative one- and two-domain models of classroom interactions.

All CLASS dimensions were live-coded by observers on a 7-point scale: 1 or 2 (*low*) to 6 or 7 (*high*; negative climate is reverse coded). Observations were conducted by a single data collector blind to intervention condition and trained to reliability according to the following procedures: (a) 2-day training with a certified CLASS trainer and (b) scoring within one point of gold-standard codes on 80% of CLASS dimensions across five 15-minute video segments. All data collectors were master's- or doctoral-level students in psychology. During the first live classroom observation, data collectors observed with a master coder to ensure agreement in live coding. Agreement between data collectors and master coders was above 80% for all observations. Between data collection periods, each data collector passed a video-based continuing reliability test (Pianta, La Paro, et al., 2008). Data collectors reviewed five videotaped classroom segments and scored within one point of the gold standard codes 80% of the time, prior to beginning the next round of data collection. Intraclass correlations (ICCs) were then used to assess interobserver agreement between data collectors and the gold standard codes before each wave of data collection. ICCs ranged from .82 to .94 for emotional support and from .84 to .93 for classroom organization across data collection time points.

To build familiarity and reduce reactivity, data collectors introduced themselves to teachers before observing and sat in a location suggested by the teacher for an unobstructed view of the classroom with minimal impact on instruction. Data collectors observed for 15 minutes and recorded scores for 10 minutes four times during the first 100 minutes of the school day. Dimensions were coded during each 15-minute shorter observation and values were averaged for a dimension score. The four dimension scores for emotional support were averaged to form the domain score (current article α = .86–.88) and the three dimension scores for classroom organization were averaged to create the domain score (current article α = .85–.88). Note that baseline levels of classroom emotional support and organization were also used as confounding covariates in this article.

Confounding Covariates. *Confounding covariates* represent the variables that might influence the mechanisms (classroom emotional support/organization) and/or outcomes (reading/math achievement). Extant work has identified associations between each of the confounding covariates described below and the mechanisms or outcomes.

Demographic characteristics. Parents reported on demographic characteristics. A number of studies have found that White children, more affluent children, children of older parents who are married and have more education, and children of parents whose parents are employed are more likely to have higher academic achievement (Morrissey, Hutchison, & Winsler, 2014). The evidence is less clear that gender is related to achievement in kindergarten and first grade (Robinson & Lubienski, 2011). Child-level covariates included race/ethnicity (*Hispanic or Black* = 1, *other* = 0), gender (*male* = 1, *female* = 0), age (in years), and free lunch eligibility (*eligible* = 1, *not eligible* = 0). Parent-level covariates included age (in years), ethnicity (*Hispanic or Black* = 1, *other* = 0), parent education (in years), parent is single (*single* = 1, *not single* = 0), and parent work status (*works full-time* = 1, *works part-time or not at all* = 0).

Child sustained attention was measured with the Attention Sustained subtest from the Leiter International Performance Scale–Revised (Roid & Miller, 1997). This measure was included as a covariate given links between attention and academic achievement (Razza, Martin, & Brooks-Gunn, 2010). Children were shown a page with pictures of a variety of objects scattered throughout and a target object at the top. They were asked to cross out as many of the objects matching the target as possible without accidentally crossing out any other objects in a limited time frame. Performance across four trials was averaged to yield two attention scores—focused attention and lack of impulsivity—which were added together. The task has demonstrated high internal consistency reliability (α = .83) and reliability (Roid & Miller, 1997).

Child behavior problems were measured with the 36-item Sutter–Eyberg Student Behavior Inventory, the teacher version of the Eyberg Child Behavior Inventory (Eyberg & Pincus, 1999). This measure was included as a covariate given links between behavior problems and lower academic achievement (van Lier et al., 2012). On a frequency scale ranging from 1 to 7 (1 = *never* to 7 = *always*), teachers reported on the frequency with which each consented child engaged in a range of problematic behaviors. A mean score was calculated from the scale items, and possible scores ranged from 1 to 7. The scale has shown evidence of validity and reliability (Conners, Sitarenios, Parker, & Epstein, 1998; Querido & Eyberg, 2003). Cronbach's alpha in this article was .97 in both kindergarten and first grade.

Teacher–child relationship quality. The 15-item teacher-reported Student–Teacher Relationship Scale (STRS; Pianta, 2001) was used to assess teacher perceptions of the quality of the teacher–child relationship at T1. This measure was included as a covariate given links between teacher–child relationships and academic achievement (McCormick, O'Connor, Cappella, & McClowry, 2013) as well as classroom emotional support and organization (Cadima, Doumen,

Verschueren, & Buyse, 2015). Using a 5-point Likert-type scale that ranged from 1 (*definitely does not apply*) to 5 (*definitely applies*), teachers rated how applicable statements were to their current relationship with a particular child. This scale contains two subdimensions: Closeness and Conflict. The Closeness subscale consists of 8 items and measures the amount of warmth and communication present in the relationship. The Conflict subscale consists of 7 items and measures the extent to which the relationship is marked by disharmonious interactions. The mean of each scale was taken to calculate a dimension score. Mashburn, Hamre, Downer, and Pianta (2006) have identified reliability and validity for the closeness and conflict subdimensions of the scale. Cronbach's αs in this study were .92 for Closeness at T1 and .88 for Conflict at T1.

Teacher perceptions of academic competence. Two subscales of the Academic Competence Evaluation Scale (DiPerna & Elliott, 2000) measured teacher perceptions of children's achievement-related behaviors in reading and mathematics. These scales have demonstrated evidence of reliability and validity (DiPerna & Elliott, 2000). We included this measure as a covariate to control for teacher perceptions of achievement, which have been linked to standardized assessments of achievement (Südkamp, Kaiser, & Möller, 2012). Teachers rated students' academic skills in comparison with the grade-level expectations at their school (1 = *far below*, 3 = *grade level*, 5 = *far above*). Consisting of 11 items, the Reading/Writing subscale includes items about the skills necessary for generating and understanding written language. The 8-item Mathematics subscale reflects skills related to use and application of numbers, including computation, and problem solving. The mean of each scale was taken to calculate an average score. Internal consistencies were high in this study (α = .96 for T1 reading; α = .97 for T1 math).

Parent involvement in elementary school was assessed with the parent-reported Family Involvement Questionnaire for Elementary School (FIQ-E; Manz, Fantuzzo, & Power, 2004). This measure was included as a covariate given correlations between parent involvement and academic achievement in elementary school (Jeynes, 2012). Consisting of 44 items, the FIQ-E was developed for lower income urban families. The FIQ-E asks parents to report on the frequency with which they engage in behaviors related to their child's schooling on a scale from 1 (*never*) to 4 (*always*). A mean score was calculated from the scale items, and possible scores thus range from 1 to 4. The average alpha was .96 at T1.

Child temperament was measured with the School-Age Temperament Inventory (SATI; McClowry, 2002). This measure was included as a covariate given correlations between dimensions of temperament and academic achievement in elementary school (Valiente et al., 2013). The SATI

is a 38-item 5-point Likert-type scale (ranging from 1 = *never* to 5 = *always*) that was standardized with a racially/ethnically and socioeconomically diverse sample of 883 parents reporting on their children (McClowry, 2002). This study also found evidence of reliability and validity. The instrument has four dimensions derived from principal factor analysis: Negative Reactivity (12 items; intensity/frequency of negative affect), Task Persistence (11 items; self-direction in fulfilling task responsibilities), Withdrawal (9 items; child's initial response to new situations), and Activity (6 items; large motor activity). For each subscale, the mean of the items was taken to calculate an average score. Cronbach's alphas for the SATI (completed at enrollment) were .77 for Activity, .81 for Withdrawal, .85 for Task Persistence, and .87 for Negative Reactivity.

Classroom characteristics. Participating teachers reported on their years of teaching experience. During classroom observations, research assistants collected information on class size and number of adult staff present during academic activities.

Analytic Approach

Mediation Analyses in Randomized Control Trials. In this study, we are explicitly interested in the classroom-level mechanisms (emotional support and organization) theorized to explain the impacts of *INSIGHTS* on student achievement. Baron and Kenny (1986) developed the classic method for addressing this problem. In that framework, one would be interested in decomposing the total effect of the treatment on the outcome (Path C) into the part attributed to the effect of the treatment on the mediator (Path A) and the mediator on the outcome (Path B). Yet this correlational approach is problematic in the context of a randomized control trial because the mediators were assessed posttreatment. As such, although the treatment difference in the mediators themselves would be unbiased, attempts to attribute gains in student outcomes to treatment impacts on the mediators could be biased (Page, 2012; Reardon & Raudenbush, 2013).

This study thus uses a number of methods to attempt to improve the rigor of the mediation analysis, over and above the traditional correlational approach. The first method, called instrumental variables estimation (IVE), has increasingly been used to assess mediation in the context of program evaluation (e.g., Gennetian, Magnuson, & Morris, 2008; Hill, Brooks-Gunn, & Waldfogel, 2003). IVE uses a two-stage approach to first estimate the effect of an instrument on the proposed mediator. In the second stage, the regression-predicted estimates from the first stage are used to estimate the effect of the mediator on the outcome. In the context of this article, assignment to *INSIGHTS* is the instrument. Classroom emotional support and organization are the theorized mediators, and student math and reading achievement are the

outcomes. There are a number of assumptions inherent in inferring causality from this set of analyses, the tenability of which are discussed in the analysis section. One critical assumption in IVE, however, is the exclusion restriction that posits that the hypothesized mediator is the *only mechanism* linking the instrument to the outcomes of interest.

It is highly unlikely that this analysis meets the exclusion restriction, as there are a number of theoretical mechanisms linking assignment to *INSIGHTS* to students' math and reading achievement (see Jones & Bouffard, 2012). Still, we argue that by using a variety of methods to examine the same research questions, it may still be possible to increase confidence in the findings from the mediation models (assuming results across methods are consistent). As such, after examining IVE mediation estimates, we will run additional models using inverse probability of treatment weighting (IPTW) to examine classroom emotional support and organization as mediators linking *INSIGHTS* and students' academic outcomes. Like instrumental variables, IPTW addresses possible selection bias at the level of the mediator but does not assume that the proposed mediator is *the one* explanatory pathway. Instead, there is a different critical assumption that there is ignorability of selection into mediation groups (typically discussed as ignorability of selection into treatment), conditional on covariates (Page, 2012). Generally, IPTW is a framework whereby the potential values of a postrandomized variable pertaining to a question of interest (i.e., the mediator) are used to reweight the sample based on observed pretreatment characteristics. Conducting this procedure will enable generation of an unbiased effect of emotional support and classroom organization on reading and math achievement, assuming that we have included all possible confounding covariates in the model. We argue that if we do find consistent results across these three different modeling approaches—multilevel regression, IVE and IPTW—we will be able to improve the empirical understanding of classroom level mechanisms linking *INSIGHTS* to student achievement.

Missing Data Analyses. There were no missing classroom-level data in this study. However, for the child-level variables, there was 0% to 20% missing data across study variables. As such, we first compared students who were missing and not missing individual data points on a series of baseline characteristics, specifically, school, teacher, cohort, child ethnicity (e.g., Hispanic or Black), child's gender, child age, child free lunch eligibility, behavior problems, sustained attention, math achievement, reading achievement, parent gender, parent age, parent ethnicity, parent education, parent marital status, and parent work status. Although we did not find substantial differences in rates of missingness between students by treatment status or student outcomes of interest, missingness patterns between baseline variables were not random. Students with lower levels of

parental education, parents who were not married, and those with more behavior problems were most likely to be missing outcome data, equally across the treatment and comparison groups. As such, the assumptions for complete case analysis were not met (Hill, Waldfogel, Brooks-Gunn, & Han, 2005; Rubin & Little, 2002).

A multiple imputation (MI) method was thus employed, and 20 separate data sets were imputed by chained equations, using Stata MICE in Stata version 12 (Enders, 2013; Schafer & Graham, 2002). MI assumes data are missing at random. MI replaces missing values with predictions based on all the information observed in the study and accounts for uncertainty about missing data by imputing several values for each missing value, generating multiple data sets. Stata ran each set of analyses 20 times and aggregated findings across them.

Between-Group Variance. Given that these analyses were unlikely to meet the regression assumption of independence, we first ran unconditional multilevel models to decompose the between- and within-school variance for the classroom level mechanisms, and the between- and within-classroom variance for the student-level outcomes. ICCs (see Raudenbush & Bryk, 2002) indicated nonnegligible between-school and between-classroom variation. For classroom-level variables, ICCs were .16 and .18 for kindergarten classroom emotional support and organization, and .18 and .21 for first grade classroom emotional support and organization. For achievement outcomes ICCs at the classroom level were .12 and .15 for math and reading achievement in kindergarten, and .13 and .17 for math and reading achievement in first grade.

Regression Models for Research Questions 1 and 2. Next, we used regression analyses and the Baron and Kenny (1986) mediation approach to determine whether there was correlational evidence for the hypothesized mechanisms. All regression models used robust standard errors to account for nonnegligible clustering. There was previous evidence for effects of *INSIGHTS* on classroom emotional support and organization (Cappella et al., in press), as well as student math and reading skills (O'Connor et al., 2014). Thus, we regressed math and reading achievement (outcomes) on classroom emotional support and classroom organization (mediators), adjusting for confounding covariates (Path B in Baron and Kenny model). Then, we tested for mediation by regressing math and reading achievement on the hypothesized mediators and assignment to *INSIGHTS*, adjusting for confounding covariates (Path C'). If the coefficient for classroom emotional support and/or organization shrank after accounting for assignment to *INSIGHTS*, there would be correlational mediation evidence (O'Connor & McCartney, 2007).

Instrumental Variables Estimation for Research Questions 1 and 2. Next, IVE was used to examine those same

pathways. IVE uses a two-stage least squares regression with the *ivregress* command in Stata 12 to estimate (a) the within-grade posttest effect of *INSIGHTS* on classroom emotional support and organization (Stage 1) and (b) the within-grade effect of classroom emotional support and organization on kindergarten and first grade students' academic outcomes posttest for those classrooms in which emotional support and organization improved as a result of *INSIGHTS* (Stage 2; Angrist, Imbens, & Rubin, 1996; Gelman & Hill, 2007). In the first stage, separate ordinary least squares regressions with robust standard errors were used to predict classroom emotional support and organization from exogenous instruments (random assignment to *INSIGHTS* or control condition), adjusting for pretreatment covariates. Classroom emotional support and organization in the IVE approach are operationalized continuously, in their original raw score form. The Stage 1 equation is

$$M_{ij} = \alpha_0 + \alpha_1 Z_{ij} + \gamma_1 X_{ij} + v_{ij} \tag{1}$$

In the second stage, the intervention-predicted values of classroom emotional support and organization were used to generate unbiased estimates of the effect of classroom emotional support and organization on children's academic outcomes (Yij), adjusting for pretreatment covariates. Models predicting emotional support controlled for classroom organization and vice versa. The equation for the second stage is

$$\hat{Y}_{ij} = \beta_0 + \beta_1 M_{ij} + \varphi_1 X_{ij} + \varepsilon_{ij} \tag{2}$$

In this model Y_{ij} represents achievement, M_{ij} is the intervention-predicted value of the explanatory variables (classroom emotional support or organization), φ_1 represents covariates, and ε_{ij} takes account of residual error.

The *estimand* of interest is the effect of the explanatory variables (classroom emotional support or organization) on the outcome for those who would experience an improvement in the hypothesized mechanisms if assigned to *INSIGHTS*, and would not experience an improvement in the hypothesized mechanisms if not assigned to *INSIGHTS*. All models used the *vce(cluster)* command in Stata to account for the nonindependence of student observations within classrooms and general robust standard errors.

Instrumental variables analysis, however, requires a number of assumptions, four of which are tenable in this analysis. First, the assumption that there is a nonzero correlation between the instrument and the mechanisms has been established in a previous study (see Cappella et al., in press).[1] Second, the assumption that there is ignorability of the instrument is tenable because the instrument was manually randomized. Third, we can assume monotonicity, or that there are no defiers in this analysis, because there were no children who left a treatment or control group to enroll in a school participating in the alternate study condition.

Finally, the stable unit treatment value assumption assumes that the outcome for a given individual is not dependent on the experienced treatment of another individual in the sample. Although this assumption is difficult to account for given the nested data structure, using robust standard errors increases the likelihood of capturing the nonindependence of observations.

The critical assumption not met in this analysis, however, is the exclusion restriction. This assumption requires that the hypothesized mediator (classroom emotional support or organization) is the only mechanism linking assignment to *INSIGHTS* with student achievement (Gelman & Hill, 2007). Given that we are testing two mechanisms, this assumption is inherently untenable. This is problematic because if an analysis does not meet the exclusion restriction, there is additional bias in the model that depends on both the effect of the instrument (randomization) on the group called noncompliers (those that improved in the mediators when assigned to the comparison condition and those that did not improve in the mediators when assigned to *INSIGHTS*) and the odds that participants are noncompliers (Gelman & Hill, 2007). The instrument in this article is strong—school-based random assignment to *INSIGHTS* versus the attention-control. Even so, we also consider the same models using a different approach—IPTW—to build further evidence for these mediated pathways.

Inverse Probability of Treatment Weighting for Research Questions 1 and 2. IPTW is a variant of propensity score matching that assumes there is ignorability of selection into mediation groups conditional on covariates (see Hill, Weiss, & Zhai, 2011; Page, 2012). The overall goal of IPTW is to reweight the control group so it looks comparable to the treatment group in terms of confounding characteristics. To accomplish this goal, one must weight each group by the inverse of the estimated probability that they were assigned to the group. To operationalize this approach, we first used a logistic regression with school fixed effects to estimate the likelihood that assignment to *INSIGHTS* was associated with receiving high-quality classroom emotional support and organization, conditioned on pretreatment confounding covariates. Based on work by Burchinal, Vandergrift, Pianta, and Mashburn (2010) identifying threshold effects for classroom quality on child outcomes, we used a score of 4.7 to identify classrooms with "high emotional support and organization." Thus, treatment is binarized for this approach, rather than treated continuously. The Step 1 equation is below and was run only for the students assigned to the *INSIGHTS* program:

$$\text{Logit } (Dij) = \beta_0 + \beta_1 C_{ij} + \alpha_j, \tag{3}$$

C_{ij} is a vector that represents pretreatment characteristics for student i in school j that possibly influence that child's

likelihood for being exposed to high levels of classroom emotional support or organization. β_1 is a vector of the predicted probabilities for all covariates. α_i represents school fixed effects.

We then used the coefficients for the *INSIGHTS* group and applied them to covariate data for the group of students originally randomly assigned to the attention-control group. We conducted this procedure separately for emotional support and classroom organization. Third, we used IPTW to weight the students assigned to the attention control group so that they looked like the group assigned to the *INSIGHTS* condition with respect to all pretreatment covariates. Assuming that all relevant pretreatment characteristics were included, the weighted control group should provide an appropriate comparison group with average values of the baseline variables similar to the treatment participants with high levels of the proposed mediators (i.e., classroom emotional support and organization). The *estimand* in this analysis is the treatment on the treated, i.e., the effect of the proposed mediators (i.e., the "treatment"—high classroom emotional support or organization in either kindergarten or first grade) on math or reading achievement at the end of the school year, compared to what would have happened to the child's achievement if he or she had not had experienced high classroom emotional support and/or organization earlier in the school year. Finally, we used separate multivariate regressions to estimate the impact of high classroom emotional support and organization on math and reading achievement in kindergarten and first grade, applying the weights so that the control group approximated the treatment group.

IPTW also requires a number of assumptions to infer causality. Like instrumental variables, one must assume ignorability and the stable unit treatment value assumption. In addition, there must be sufficient overlap between treatment and control. Because of the randomized design, and the only differentiating factor being assignment to *INSIGHTS*, one can assume overlap. Finally, there must be evidence of balance between the treatment group (in this case, high emotional support / classroom organization) and the control group (low emotional support / classroom organization). Although this assumption is strong, balance statistics presented in the results section suggest it may be tenable. The stable unit treatment value assumption is somewhat untenable, given the nested nature of the data. However, using school fixed effects to account for group differences may address some concern related to the analysis meeting this assumption.

Results

Descriptive Findings for Baseline Variables

Descriptive statistics are listed in Table 1. Independent samples *t* tests showed significant differences between groups in kindergarten teacher emotional support, favoring schools in the comparison group, $t(58) = 2.03, p < .05$. There

were no statistically significant differences between treatment and control for classroom emotional support and organization in first grade or for classroom organization in either grade.

Independent samples *t* tests demonstrated significant pretreatment differences between children enrolled in *INSIGHTS* and the supplemental reading group on reading achievement, $t(433) = 3.12, p < .01$. At baseline, children in *INSIGHTS* evidenced lower overall scores on reading achievement than their peers in the supplemental reading program. Pretreatment differences on other covariates and outcomes were nonsignificant.

Regression Results for Research Questions 1 and 2

Findings from regression analyses from kindergarten for classroom emotional support and organization are presented in Tables 2 and 3, respectively. In kindergarten, the B paths linking the mediators with achievement were nonsignificant for both classroom emotional support and organization. Thus, there was no correlational evidence for the hypothesized mediators in kindergarten.

First grade regression findings are presented in Tables 4 and 5. In first grade, emotional support measured posttest was associated with math achievement, $B = 0.89, SE = 0.38, p = .02$. In addition, in the test of the C' path the coefficient for classroom emotional support was reduced but still significant, $B = 0.72, SE = 0.38, p = .03$. In first grade, classroom organization measured posttest was associated with math achievement, $B = 0.97, SE = 0.40, p = .02$. In addition, in the test of the C' path, after adjusting for assignment to *INSIGHTS*, the coefficient for classroom emotional support was reduced but still significant, $B = 0.83, SE = 0.40, p = .03$. There was no evidence of mediation through classroom variables on reading achievement in first grade.

Instrumental Variables Estimation for Research Questions 1 and 2

As observed in Table 6, there was no evidence in the Stage 1 model that assignment to *INSIGHTS* (the instrument) significantly predicted either classroom emotional support or organization in kindergarten. Thus, as expected, in Stage 2 (Table 7), there were no significant mediated effects of either mediator on math or reading achievement in kindergarten.

In first grade, the Stage 1 model revealed that assignment to *INSIGHTS* had a significant impact on both classroom emotional support, $B = 0.36, SE = 0.09, p < .01$ and classroom organization, $B = 0.39, SE = 0.09, p < .01$ (see Table 8). After accounting for these predicted values in Stage 1, the results of Stage 2 models revealed evidence for the impact of first grade classroom emotional support on math achievement at the end of first grade, $B = 0.88, SE = 0.41, p = .04$, ES = 0.13 (see top panel of Table 9). This is the effect of

TABLE 2

Regressions Predicting Academic Achievement From Assignment to INSIGHTS *and Classroom Emotional Support in Kindergarten*

	Math achievement		Reading achievement			Math achievement		Reading achievement	
Fixed effects—Path B	B	SE	B	SE	Fixed effects—Path C'	B	SE	B	SE
Intercept	19.92**	1.62	25.85**	2.79	Intercept	20.03**	1.67	26.45**	2.81
Classroom emotional support	−0.08	0.31	−0.04	0.48	Assignment to INSIGHTS	−0.48	0.48	2.55	1.98
School % free lunch eligibility	−1.86	1.78	−4.25	2.85	Classroom emotional support	−0.10	0.32	−0.17	0.46
School % Black	1.18	2.14	5.28	3.79	School % free lunch eligibility	−1.71	1.74	−3.49	2.79
School % Hispanic	−1.81*	0.89	−2.22†	1.20	School % Black	1.29	2.18	5.88	3.83
School avg. attendance	1.63	1.24	0.99	1.66	School % Hispanic	−1.71†	0.89	−1.65	1.21
School size	0.01	0.01	−0.01	0.01	School avg. attendance	−1.26	1.30	0.97	1.80
Baseline reading achievement	0.12**	0.04	0.67**	0.07	School size	0.01	0.01	−0.01	0.01
Baseline math achievement	0.43**	0.08	0.14*	0.07	Baseline reading achievement	0.12**	0.04	0.65**	0.07
Baseline behavior problems	−0.17	0.20	−0.56†	0.29	Baseline math achievement	0.44**	0.08	0.18*	0.07
Baseline sustained attention	−0.01	0.02	0.01	0.03	Baseline behavior problems	−0.15	0.19	−0.45	0.29
Child age at study entry	0.57	0.40	−0.70	0.53	Baseline sustained attention	−0.01	0.02	−0.01	0.02
Child Black	−0.88	0.88	0.51	0.99	Child age at study entry	−0.55	0.41	−0.56	0.53
Child Hispanic	−0.77	0.79	−0.40	1.06	Child Black	−0.87	0.89	0.59	0.99
Child biracial	−0.21	0.99	−1.51	1.08	Child Hispanic	−0.69	0.79	0.08	1.08
Child male	−0.83 f	0.47	−0.66	0.65	Child biracial	−0.25	0.99	−1.69	1.07
Parent age	−0.03	0.02	0.05	0.03	Child male	−0.81*	0.47	−0.56	0.64
Parent education (in years)	0.08	0.09	0.15	0.12	Parent age	−0.03	0.02	0.04	0.03
Parent is single	0.23	0.46	−0.71	0.71	Parent education (in years)	−0.08	0.09	0.14	0.13
Parent in Black	−0.28	0.63	−0.32	1.02	Parent is single	0.25	0.46	−0.58	0.70
Parent is Hispanic	−0.87	1.01	−1.01	0.78	Parent in Black	−0.30	0.63	−0.38	0.98
Parent works full-time	0.34	0.62	0.13	0.97	Parent is Hispanic	0.79	0.88	−0.99	0.781
Child negative reactivity	−0.44	0.36	−0.36	0.52	Parent works full-time	0.37	0.62	0.32	0.99
Child task persistence	0.08	0.32	0.14	0.43	Child negative reactivity	−0.47	0.35	−0.52	0.49
Child withdrawal	−0.29	0.28	−0.37	0.41	Child task persistence	0.09	0.32	0.18	0.43
Child motor activity	0.72*	0.28	0.96*	0.48	Child withdrawal	−0.28	0.27	−0.32	0.40
					Child motor activity	0.74*	0.29	1.06*	0.46

$^{†}p < .10.$ *$p < .05.$ **$p < .01.$

classroom emotional support on math achievement for those students who would have experienced higher emotional support if their classroom was assigned to *INSIGHTS* and would have experienced lower emotional support if their classroom was not assigned to *INSIGHTS*.

Similarly, there was also evidence for the impact of first grade classroom organization on math achievement at the end of first grade for the children who experienced high classroom organization, relative to students experiencing lower levels of classroom organization, $B = 1.13$, $SE = 0.41$, $p = .03$, ES $= 0.47$ (see bottom panel of Table 9). This is the effect of classroom organization on math achievement for those students who would have experienced higher organization if their classroom was assigned to *INSIGHTS* and would have experienced lower organization if their classroom was not assigned to *INSIGHTS*.

Inverse Probability of Treatment Weighting (IPTW) for Research Questions 1 and 2

We did not test for mediated impacts in kindergarten using IPTW given nonsignificant effects for both regression and IVE analyses. In addition, before examining mediated impacts in first grade, we examined balance between the predicted probabilities for the high and low emotional support and classroom organization groups. As illustrated in Tables 10 (emotional support) and 11 (classroom organization), we found few significant differences in standardized differences or standard deviations for both emotional support and classroom organization following weighting procedures. Overall differences in the weighted groups were minimal for both emotional support and classroom organization.

TABLE 3

Regressions Predicting Academic Achievement From Assignment to INSIGHTS *and Classroom Organization in Kindergarten*

Fixed effects—Path B	Math achievement		Reading achievement	
	B	SE	B	SE
Intercept	18.91**	1.75	24.70**	2.87
Classroom organization	0.38	0.30	0.44	0.28
School % free lunch eligibility	−0.72	1.99	−2.96	2.98
School % Black	1.60	2.05	5.66†	3.41
School % Hispanic	−1.96*	0.90	−2.29*	1.18
School avg. attendance	0.99	1.25	0.28	0.18
School size	0.01	0.01	−0.01	0.01
Baseline reading achievement	0.12**	0.05	0.67**	0.07
Baseline math achievement	0.44**	0.08	0.15*	0.07
Baseline behavior problems	−0.14	0.19	−0.53†	0.30
Baseline sustained attention	−0.01	0.02	−0.01	0.02
Child age at study entry	−0.57	0.40	0.69	0.53
Child Black	−0.85	0.87	0.54	0.98
Child Hispanic	−0.60	0.77	−0.21	1.06
Child biracial	−0.30	1.00	−1.60	1.08
Child male	−0.75	0.47	−0.58	0.64
Parent age	−0.03	0.02	0.04	0.03
Parent education (in years)	0.10	0.09	0.14	0.12
Parent is single	0.17	0.47	−0.77	0.73
Parent in Black	−0.26	0.63	−0.29	1.01
Parent is Hispanic	−0.60	0.77	−0.21	1.06
Parent works full-time	0.29	0.62	0.09	0.97
Child negative reactivity	−0.46	0.35	−0.38	0.52
Child task persistence	0.23	0.32	0.18	0.43
Child withdrawal	−0.27	0.27	−0.33	0.41
Child motor activity	0.73*	0.29	0.97*	0.48

Fixed effects—Path C'	Math achievement		Reading achievement	
	B	SE	B	SE
Intercept	19.08**	1.85	25.87**	2.93
Assignment INSIGHTS	−0.37	0.50	−2.01	1.94
Classroom organization	0.34	0.31	0.17	0.46
School % free lunch eligibility	−0.71	1.99	−2.90	2.97
School % Black	1.71	2.09	6.39†	3.49
School % Hispanic	−1.86*	0.92	−1.76	1.19
School avg. attendance	−0.76	1.30	1.30	1.88
School size	0.01	0.01	−0.01	0.01
Baseline reading achievement	0.12**	0.04	0.65**	0.07
Baseline math achievement	0.45**	0.08	0.19*	0.07
Baseline behavior problems	−0.12	0.19	−0.44	0.28
Baseline sustained attention	−0.02	0.02	−0.01	0.02
Child age at study entry	−0.55	0.40	−0.57	0.53
Child Black	−0.84	0.88	0.60	0.90
Child Hispanic	−0.54	0.77	0.19	1.08
Child biracial	−0.32	1.01	−1.74	1.08
Child male	−0.74	0.47	−0.51	0.64
Parent age	0.03	0.02	0.04	0.03
Parent education (in years)	0.10	0.09	0.13	0.12
Parent is single	0.19	0.47	−0.62	0.72
Parent in Black	−0.27	0.63	−0.37	0.97
Parent is Hispanic	0.75	0.78	−0.76	0.65
Parent works full-time	0.32	0.62	0.28	0.99
Child negative reactivity	−0.48	0.35	−0.53	0.50
Child task persistence	0.13	0.32	0.21	0.43
Child withdrawal	−0.25	0.27	−0.30	0.40
Child motor activity	0.75*	0.28	1.07*	0.46

†$p < .10$. *$p < .05$. **$p < .01$.

IPTW findings revealed similar patterns as the results from the regression and IV models. Interpreted causally, experiencing high classroom emotional support (>4.7) attributed to *INSIGHTS* resulted in a higher math score of 3.80 points, relative to students who experienced a lower level of classroom emotional support (<4.7), $B = 3.80$, $SE = 1.88$, $p = .03$ (see Table 12). Similarly, high classroom organization (>4.7) in first grade attributed to *INSIGHTS* resulted in a higher math score of 1.84 points, relative to students who experienced lower levels of classroom organization (<4.7), $B = 1.84$, $SE = 0.73$, $p = .04$. High classroom organization also resulted in a higher reading score of 3.03 points, relative to students who experienced lower classroom organization, $B = 3.03$, $SE = 1.45$, $p = .04$ (see results in Table 12). Notably, although this is the only method that demonstrated a mediated effect on reading achievement, the magnitude of the mediated effect for the other methods was similarly large. The standard errors for regression and IVE, however, were larger, and the effects were thus nonsignificant.

Discussion

The aim of this study was to use a series of models to test whether impacts of *INSIGHTS* on math and reading achievement were mediated through improvements in classroom emotional support and organization in kindergarten and first grade. Results consistently demonstrated that program impacts on math achievement in first grade were mediated through improvements in both classroom emotional support and organization. Findings from one method—IPTW—showed that program impacts on reading

TABLE 4

Regressions Predicting Academic Achievement From Assignment to INSIGHTS *and Classroom Emotional Support in First Grade*

Fixed effects—Path B	Math achievement		Reading achievement		Fixed effects—Path C'	Math achievement		Reading achievement	
	B	SE	B	SE		B	SE	B	SE
Intercept	25.04**	2.40	26.44	8.84	Intercept	25.44**	2.42	27.30**	8.29
Classroom emotional support	0.89*	0.38	0.76	1.14	Assignment to INSIGHTS	0.64	0.59	1.39	1.78
School % free lunch eligibility	−1.84	2.34	3.13	7.98	Classroom emotional support	0.72*	0.38	0.67	1.09
School % Black	−0.92	3.10	2.94	5.91	School % free lunch eligibility	−2.43	2.42	−2.43	2.42
School % Hispanic	1.05	1.02	−6.12	5.45	School % Black	−0.92	3.09	−0.92	3.09
School avg. attendance	1.51	1.32	5.36	3.49	School % Hispanic	1.11	1.02	1.11	1.02
School size	−0.01	0.01	−0.01	0.01	School avg. attendance	1.10	1.40	1.10	1.40
Baseline reading achievement	0.20**	0.04	0.72**	0.22	School size	−0.02	0.01	−0.01	0.01
Baseline math achievement	0.32**	0.06	0.22	0.17	Baseline reading achievement	0.21**	0.04	0.21**	0.04
Baseline behavior problems	−0.90**	0.24	−2.28	0.52	Baseline math achievement	0.30**	0.06	0.30**	0.06
Baseline sustained attention	0.01	0.02	0.13*	0.05	Baseline behavior problems	−0.92**	0.24	−0.92**	0.24
Child age at study entry	−0.55	0.42	−0.67	1.28	Baseline sustained attention	0.02	0.02	0.02	0.02
Child Black	−0.90	0.93	2.23	1.8	Child age at study entry	−0.56	0.42	0.91	1.25
Child Hispanic	−1.19	0.91	1.26	1.26	Child Black	−1.01	0.94	1.99	1.74
Child biracial	−0.97	1.18	0.42	3.44	Child Hispanic	−1.47	0.97	0.65	1.74
Child male	−0.34	0.56	1.53	1.53	Child biracial	−1.03	1.21	0.30	1.26
Parent age	−0.04	0.04	−0.05	0.13	Child male	−0.36	0.56	1.48	1.48
Parent education (in years)	−0.06	0.10	0.38	0.25	Parent age	−0.04	0.04	−0.05	0.14
Parent is single	0.29	0.61	0.54	2.01	Parent education (in years)	−0.07	0.10	0.37	0.25
Parent in Black	−0.90	0.93	−0.87	1.66	Parent is single	0.23	0.62	0.39	1.92
Parent is Hispanic	−1.19	0.91	0.12	0.14	Parent in Black	0.92	0.75	−1.06	1.75
Parent works full-time	−0.48	0.66	1.27	2.39	Parent is Hispanic	−0.12	0.15	−0.11	0.13
Child negative reactivity	−0.23	0.38	0.21	1.01	Parent works full-time	−0.48	0.66	1.27	2.42
Child task persistence	−0.44	0.39	−0.88	1.17	Child negative reactivity	−0.17	0.38	0.32	1.01
Child withdrawal	−0.40	0.44	−2.64†	1.55	Child task persistence	−0.45	0.38	0.91	1.18
Child motor activity	0.22	0.37	0.62	0.68	Child withdrawal	−0.38	0.43	−2.62†	1.52
					Child motor activity	0.18	0.37	0.54	0.73

†$p < .10$. *$p < .05$. **$p < .01$.

achievement in first grade were mediated through improvements in classroom organization. Notably, there was no evidence of any classroom-level mediated pathways in kindergarten.

Classroom Social Processes and Achievement in Urban Schools

Current findings reflect recent work showing that emotional support and organization in classroom settings are particularly important for the development of low-income urban children's elementary school math skills (Crosnoe et al., 2010; Hughes, 2011; McCormick et al., 2013; Spilt, Hughes, Wu, & Kwok, 2012). This is an important finding given the strong links between early math skills and later achievement (Duncan et al., 2007). The mediated effect on math but not reading for two of the methods may reflect the fact that children who are in positive, secure, and safe environments are more comfortable taking the types of cognitive risks (e.g., possibility of failure) necessary to learn new math skills in first grade (Curby, Rimm-Kaufman, & Ponitz, 2009). Reading achievement, however, may be less related to the pattern of social interactions at the classroom level and more associated with individual time spent on-task practicing literacy skills. Indeed, previous correlational research suggests a mediated effect of *INSIGHTS* on reading achievement through an improvement in sustained attention—an individual student-level competency (O'Connor et al., 2014). It is also possible that effects on reading may have been diluted by the comparison children taking part in a program targeted at reading skills. Even so, findings from the IPTW method did reveal that program impacts on reading achievement in first grade were partially explained by improvements in classroom organization.

TABLE 5

Regressions Predicting Academic Achievement From Assignment to INSIGHTS *and Classroom Organization in First Grade*

Fixed effects—Path B	Math achievement		Reading achievement		Fixed effects—Path C'	Math achievement		Reading achievement	
	B	SE	B	SE		B	SE	B	SE
Intercept	24.57**	2.33	28.11**	7.13	Intercept	28.81**	6.78	28.82**	6.78
Classroom organization	0.97*	0.40	2.27	2.15	Assignment to INSIGHTS	0.50	0.57	1.22	1.63
School % free lunch eligibility	−2.09	2.29	1.43	6.70	Classroom organization	0.83*	0.40	2.58	2.08
School % Black	−1.23	3.05	6.37	6.25	School % free lunch eligibility	−2.54	2.36	0.31	6.29
School % Hispanic	0.76	1.05	−7.44	6.35	School % Black	−1.25	3.04	6.31	6.22
School avg. attendance	1.68	1.29	3.27	3.19	School % Hispanic	0.79	1.04	−7.37	6.30
School size	−0.01	0.01	0.01	0.01	School avg. attendance	1.37	1.35	2.50	3.43
Baseline reading achievement	0.20**	0.04	0.71**	0.14	School size	−0.01	0.01	−0.01	0.01
Baseline math achievement	0.33**	0.06	0.25†	0.15	Baseline reading achievement	0.21**	0.04	0.72**	0.15
Baseline behavior problems	−0.79**	0.22	−2.24**	0.53	Baseline math achievement	0.31**	0.06	0.22	0.17
Baseline sustained attention	0.01	0.02	0.13*	0.05	Baseline behavior problems	−0.80**	0.23	−2.26**	0.53
Child age at study entry	−0.61	0.43	−1.19	1.10	Baseline sustained attention	0.01	0.02	0.14*	0.06
Child Black	−0.32	0.94	2.64	1.89	Child age at study entry	−0.62	0.43	−1.22	1.09
Child Hispanic	−0.79	0.88	1.42	1.71	Child Black	−0.40	0.95	2.46	1.84
Child biracial	−0.58	1.19	0.99	3.20	Child Hispanic	−0.90	0.93	0.94	1.62
Child male	−0.12	0.54	1.34	1.32	Child biracial	−0.62	1.23	0.89	3.28
Parent age	−0.04	0.03	−0.04	0.12	Child male	−0.13	0.54	1.31	1.30
Parent education (in years)	0.07	0.09	0.27	0.23	Parent age	−0.04	0.04	−0.04	0.04
Parent is single	0.52	0.60	0.66	1.95	Parent education (in years)	0.08	0.09	0.27	0.22
Parent in Black	0.93	0.78	−1.24	1.89	Parent is single	0.46	0.60	0.53	1.89
Parent is Hispanic	1.01	1.23	0.87	0.82	Parent in Black	0.87	0.78	−1.39	0.20
Parent works full-time	−0.71	0.64	0.90	2.39	Parent is Hispanic	−0.98	0.86	−1.01	0.89
Child negative reactivity	−0.17	0.37	0.65	1.16	Parent works full-time	−0.70	0.64	0.93	2.43
Child task persistence	0.29	0.28	−0.81	1.06	Child negative reactivity	−0.13	0.38	0.75	1.24
Child withdrawal	−0.37	0.40	−2.68†	1.53	Child task persistence	0.31	0.38	0.83	1.07
Child motor activity	−0.29	0.38	0.65	1.16	Child withdrawal	−0.36	0.41	−2.66†	1.51
					Child motor activity	0.26	0.36	0.59	0.69

$^\dagger p < .10.$ $^* p < .05.$ $^{**} p < .01.$

With respect to reading instruction more specifically, there is some work to suggest that elementary school teachers are better prepared to teach literacy skills, relative to math (Tatto & Senk, 2011). Most states will grant licenses to preservice elementary school teachers even if they fail the math section of the licensure exam (Epstein & Miller, 2011). At the same time, most SEL programs are either explicitly designed to be delivered in concert with literacy instruction (e.g., 4Rs, RULER) or implemented during literacy instruction given more potential overlap between relevant lessons and themes. For example, with respect to INSIGHTS, there are likely more opportunities to discuss behaviors in relation to storybook characters, than in the context of learning one's numbers or basic numeracy. Given these trends, there may simply be more room for growth in math achievement attributed to gains in classroom climate, relative to reading.

When there are improvements in classroom organization, as there were in this study, there may also be more aggregate teaching time to provide math instruction. Indeed, past research using national longitudinal survey data has shown that teachers in K–3 settings, on average, spend 35 minutes more each day teaching reading than math (Banilower et al., 2013). By improving behavior management and productivity in the classroom, it is possible that there was more time available to allocate to math instruction. In addition, it may be that the intervention shifted the behavioral norms in the classroom due to improvements in teachers' ability to respond clearly and appropriately to students with diverse behavioral needs. In this way, students were more behaviorally engaged as a whole, spent more time learning rather than regulating and redirecting behaviors, and were thus able to improve their individual academic achievement as a result. Findings are in line with

TABLE 6

First-Stage Results for Instrumental Variables Model: Assignment to INSIGHTS *on Classroom Emotional Support and Organization in Kindergarten*

Fixed effects	Classroom emotional support		Classroom organization	
	B	SE	B	SE
Assignment to *INSIGHTS* (instrument)	0.01	0.09	−0.26	0.19
Baseline value of the mediator	0.49**	0.08	0.42**	0.07
Baseline reading achievement	0.01	0.01	0.01	0.01
Baseline math achievement	−0.01	0.01	0.01	0.01
Baseline behavior problems	0.06	0.04	0.04	0.04
Child age at study entry	0.06	0.06	0.01	0.09
Child Black	−0.26*	0.11	−0.30[†]	0.17
Child Hispanic	−0.37**	0.12	−0.33*	0.15
Child biracial	0.34	0.26	0.51*	0.21
Child male	−0.12	0.08	−0.10	0.09
Parent age	0.01	0.01	0.01	0.01
Parent education (in years)	0.01	0.02	0.02	0.02
Parent is single	0.02	0.08	0.13	0.10
Parent in Black	0.05	0.12	0.04	0.16
Parent is Hispanic	0.08	0.11	−0.06	0.08
Parent works full-time	0.09	0.11	0.10	0.15
Child negative reactivity	0.08	0.06	−0.12[†]	0.07
Child task persistence	0.04	0.05	0.05	0.07
Child withdrawal	0.03	0.04	0.01	0.06
Child motor activity	−0.04	0.05	−0.03	0.07
Intercept	4.96**	0.17	4.28**	0.20

[†]*p* < .10. *p* < .05. **p* < .01.

previous work demonstrating links between behavioral regulation, time for learning, and increased achievement (Entwisle & Alexander, 1998; Pianta, Belsky, et al., 2008; Sektnan, McClelland, Acock, & Morrison, 2010).

Differences in Impacts in Kindergarten and First Grade

This study also found evidence of mediation in first grade but not in kindergarten. Given research suggesting the importance of early intervention (Barnett, 2011; Leak et al., 2010), this result was somewhat unexpected. However, it is important to remember that there were no main effects of *INSIGHTS* on classroom organization in kindergarten. Due to the lack of impact on the proposed mediator, one would not anticipate a mediated effect of the program on achievement in kindergarten. In schools serving high proportions of low-income, racial/ethnic minority children, students may arrive at kindergarten without having attended a high-quality prekindergarten or having been sufficiently prepared for the increasing academic demands of kindergarten (Jones, Bub, & Raver, 2013; Reardon, 2011). Kindergarten teachers are thus challenged to help

new students learn about how to conduct themselves, manage behaviors, and engage in productive routines, perhaps making it more difficult for an SEL program to show a significant impact on classroom organization across the kindergarten year. Teachers of kindergarten children may require a more intensive intervention model than teachers of first grade students, who have already gone through the initial adjustment to formal schooling (Perry, Donohue, & Weinstein, 2007). There may, however, be other positive effects of SEL programs on classrooms and students. Indeed, Cappella et al. (in press) identified a positive impact of *INSIGHTS* on the overall behavioral engagement of kindergarten classrooms. Future work should test whether classroom behavioral engagement is a mediator of improvements on students' nonacademic competencies.

It is still notable that *INSIGHTS* improved classroom emotional support in first grade, in turn improving student math achievement. This finding builds on previous work demonstrating impacts of SEL programs on emotionally supportive teaching practices—safety, warmth, sensitivity, and regard for teacher practices (e.g., Brown et al., 2010; Cappella et al., 2012; Raver et al., 2011). In addition,

TABLE 7

Second-Stage Results for Instrumental Variables Model: Classroom Emotional Support and Organization Mediating the Effect of INSIGHTS *Assignment on Math and Reading Achievement in Kindergarten*

	Math achievement		Reading achievement	
Emotional support model Stage 2 fixed effects	B	SE	B	SE
Classroom emotional support (mediator)	−0.26	0.51	−0.30	0.71
Baseline reading achievement	0.11**	0.04	0.67**	0.07
Baseline math achievement	0.44**	0.08	0.13*	0.06
Baseline behavior problems	−0.14	0.18	−0.47	0.37
Child age at study entry	−0.58	0.38	−0.60	0.47
Child Black	−1.22	0.79	0.85	0.97
Child Hispanic	−0.91	0.70	−0.19	1.07
Child biracial	0.01	1.04	−1.15	1.05
Child male	−0.82*	0.39	−0.66	0.56
Parent age	−0.03†	0.01	0.04	0.03
Parent education (in years)	0.07	0.10	0.13	0.11
Parent is single	0.21	0.38	−0.54	0.62
Parent in Black	−0.30	0.60	−0.40	0.97
Parent is Hispanic	−0.21	0.19	−0.28	0.33
Parent works full-time	0.37	0.63	0.39	0.89
Child negative reactivity	−0.41	0.30	−0.10	0.44
Child task persistence	0.07	0.32	0.26	0.36
Child withdrawal	−0.27	0.30	−0.33	0.40
Child motor activity	0.72**	0.24	0.91**	0.39
Intercept	19.91**	2.45	23.33**	3.64
Classroom organization Stage 2 fixed effects				
Classroom organization (mediator)	−0.09	0.51	0.62	0.71
Baseline reading achievement	0.11	0.04	0.66**	0.07
Baseline math achievement	0.44	0.08	0.13*	0.05
Baseline behavior problems	−0.14	0.18	−0.45	0.32
Child age at study entry	−0.61	0.37	−0.71	0.43
Child Black	−1.17	0.78	0.98	0.94
Child Hispanic	−0.83	0.73	0.13	1.03
Child biracial	0.02	1.07	−1.38	1.05
Child male	−0.81*	0.39	−0.54	0.55
Parent age	0.03†	0.02	0.04	0.02
Parent education (in years)	0.07	0.09	0.07	0.11
Parent is single	0.21	0.41	−0.68	0.64
Parent in Black	−0.28	0.61	−0.27	0.92
Parent is Hispanic	−0.14	0.11	−0.17	0.14
Parent works full-time	0.34	0.63	0.18	0.89
Child negative reactivity	−0.41	0.29	−0.18	0.46
Child task persistence	0.09	0.33	0.37	0.37
Child withdrawal	−0.27	0.30	−0.32	0.37
Child motor activity	0.72**	0.24	0.93**	0.37
Intercept	18.98**	1.99	19.35**	3.03

$^†p < .10.$ $^*p < .05.$ $^{**}p < .01.$

empirical (Curby et al., 2009; Mashburn et al., 2008; Pianta, Belsky, et al., 2008) and theoretical (Jennings & Greenberg, 2009; Jones & Bouffard, 2012) studies have linked emotional support and academic competence. Yet this article is one of the first to rigorously map the pathways linking implementation of an SEL program, improvements in

TABLE 8

First-Stage Results for Instrumental Variables Model: Assignment to INSIGHTS *on Classroom Emotional Support and Organization in Kindergarten*

Fixed effects	Classroom emotional support		Classroom organization	
	B	SE	B	SE
Assignment to *INSIGHTS* (instrument)	0.36**	0.09	0.39**	0.09
Baseline value of the mediator	0.65**	0.04	0.57**	0.04
Baseline reading achievement	0.01	0.01	0.01	0.01
Baseline math achievement	0.01	0.01	−0.01	0.01
Baseline behavior problems	0.06*	0.03	0.04	0.03
Child age at study entry	0.05	0.05	0.09	0.06
Child Black	−0.04	0.12	0.09	0.13
Child Hispanic	−0.24*	0.09	−0.19†	0.09
Child biracial	−0.34†	0.18	−0.42†	0.24
Child male	−0.14*	0.07	−0.13†	0.07
Parent age	0.01	0.01	−0.01	0.01
Parent education (in years)	0.01	0.01	0.04	0.01
Parent is single	−0.03	0.06	−0.07	0.07
Parent in Black	−0.18†	0.11	−0.31*	0.12
Parent is Hispanic	−0.12	0.09	−0.14	0.09
Parent works full-time	−0.03	0.08	−0.07	0.11
Child negative reactivity	0.03	0.04	0.01	0.04
Child task persistence	0.09†	0.05	0.07	0.06
Child withdrawal	0.07†	0.04	0.08*	0.04
Child motor activity	0.01	0.01	0.04	0.04
Intercept	4.81**	0.12	4.34**	0.12

$^\dagger p < .10.$ $^* p < .05.$ $^{**} p < .01.$

emotional support, and improvements in math achievement. Continued work is needed to replicate these pathways and to determine the size of the classroom mediation effects relative to individual-level mechanisms targeted by SEL programs.

Limitations

This study has a number of limitations. The causal analyses require strong assumptions. First, the exclusion restriction in the IVE approach is untenable because there is more than one identified mediator linking assignment to *INSIGHTS* and student achievement. In practice, there should be the same number of instruments as there are mediators to address the exclusion restriction. With respect to IPTW, assumptions are more plausible. However, there were still some problems with balance in this article with some variables having standardized differences greater than 0.05. Somewhat related to this, classroom emotional support and organization were operationalized differently across the methods. The regression and IVE approach operationalized the mediators continuously while the

IPTW method binarized the mediators according to thresholds identified as valid in previous work. Future work should aim to better identify comparable treatment effects across modeling approaches.

There are also additional possible mechanisms across levels (classroom, family, child, school) that future work should examine. These include improvements in teaching practice, peer interactions, parent-child relationships, overall classroom behavioral engagement, and specific intervention components (e.g., coaching, training, PD). Additional limitations include the nongeneralizability of the sample to higher income, nonurban schools. Finally, this investigation represents only one attempt to unpack the effects of *INSIGHTS* on student achievement. Indeed, it is unclear which aspects of the comprehensive *INSIGHTS* program were critical factors in inducing the changes in classroom interactions that led to improvements in math achievement in first grade. Future mixed methods studies examining changes in teacher practice both quantitatively and qualitatively can further inform SEL programs on the active ingredients important for inducing changes for students and classrooms.

TABLE 9

Second-Stage Results for Instrumental Variables Model: Classroom Emotional Support and Organization Mediating the Effect of INSIGHTS *Assignment on Math and Reading Achievement in First Grade*

Emotional support model Stage 2 fixed effects	Math achievement		Reading achievement	
	B	*SE*	*B*	*SE*
Classroom emotional support (mediator)	0.88**	0.45	1.68	1.59
Baseline reading achievement	0.16**	0.04	0.45**	0.06
Baseline math achievement	0.28**	0.04	0.33**	0.08
Baseline behavior problems	−0.37$_*$	0.19	−1.29**	0.50
Child age at study entry	−0.38	0.37	−1.46†	0.85
Child Black	0.39	0.76	3.54*	1.70
Child Hispanic	−0.29	0.83	1.39	1.34
Child biracial	−0.72	1.16	1.20	3.22
Child male	0.25	0.47	1.25	1.17
Parent age	−0.01	0.02	−0.02	0.11
Parent education (in years)	−0.01	0.08	0.17	0.14
Parent is single	−0.58	0.52	−0.62	0.99
Parent in Black	0.18	0.75	0.51	1.19
Parent is Hispanic	−0.11	0.09	−0.29	0.28
Parent works full-time	−0.24	0.53	0.26	1.29
Child negative reactivity	−0.30	0.32	−0.61	0.63
Child task persistence	0.03	0.33	1.01	1.11
Child withdrawal	0.19	0.31	−1.18	0.93
Child motor activity	0.47	0.30	0.86	0.61
Intercept	19.01**	2.32	21.59**	9.16
Classroom organization Stage 2 fixed effects				
Classroom organization (mediator)	1.13**	0.41	3.34	2.25
Baseline reading achievement	0.16**	0.04	0.44**	0.06
Baseline math achievement	0.31**	0.04	0.37**	0.08
Baseline behavior problems	−0.33$_*$	0.18	−1.25**	0.26
Child age at study entry	−0.49	0.35	−1.99*	0.87
Child Black	0.41	0.79	3.06†	1.70
Child Hispanic	−0.13	0.79	1.34	1.28
Child biracial	−0.45	1.15	1.47	3.10
Child male	0.43	0.51	1.33	1.16
Parent age	−0.01	0.02	−0.01	0.09
Parent education (in years)	−0.02	0.08	0.11	0.14
Parent is single	−0.39	0.51	−0.41	1.08
Parent in Black	0.33	0.76	−0.06	1.31
Parent is Hispanic	0.13	0.21	−0.11	0.17
Parent works full-time	−0.28	0.50	0.03	1.45
Child negative reactivity	−0.25	0.31	−0.43	0.59
Child task persistence	0.03	0.32	−0.91	0.96
Child withdrawal	0.18	0.29	−1.24	0.95
Child motor activity	0.51	0.28	0.94	0.58
Intercept	18.07**	1.91	15.46**	11.06

$^†p < .10.$ $*p < .05.$ $**p < .01.$

TABLE 10

Balance Statistics for Classroom Emotional Support Matching Procedure Measured at the Beginning of First Grade

Variable	Treatment		Unmatched control		Matched control		Std. difference	Ratio of SDs
	M	SD	M	SD	M	SD		
Baseline classroom emotional support	5.25	0.61	4.16	0.72	5.18	0.62	0.10	0.98
Teacher years teaching	11.68	8.12	16.32	11.89	12.47	8.76	−0.07	0.93
Class size	18.87	3.62	18.82	4.41	18.32	3.35	0.12	1.08
Number of adults in classroom	1.32	0.38	1.28	0.37	1.36	0.35	−0.11	1.09
Baseline reading achievement	18.48	8.09	17.46	7.98	18.56	7.78	−0.01	1.04
Baseline math achievement	14.91	4.93	14.98	5.11	14.41	5.09	0.10	0.97
Baseline behavior problems	2.27	1.21	2.08	0.94	2.28	1.18	−0.01	1.03
Baseline sustained attention	11.89	5.21	10.65	4.68	11.54	5.14	0.07	1.01
Teacher reported reading competence	2.72	0.87	2.61	0.83	2.68	0.88	0.05	0.99
Teacher reported math competence	2.76	0.66	2.78	0.72	2.72	0.70	0.06	0.94
Student teacher conflict	1.90	1.08	1.61	0.81	1.89	1.07	0.01	1.01
Student teacher closeness	4.02	0.80	4.09	0.77	3.98	0.76	0.05	1.05
Parent involvement total score	2.70	0.53	2.65	0.56	2.65	0.56	0.09	0.95
Child negative reactivity	3.00	0.92	2.79	0.83	3.05	0.95	−0.06	0.97
Child task persistence	3.74	0.78	3.86	0.78	3.80	0.81	−0.08	0.96
Child withdrawal	2.47	0.88	2.30	0.95	2.51	0.90	−0.04	0.98
Child motor activity	2.87	0.95	2.70	0.92	2.83	0.98	0.04	0.97
Child age at study start	5.74	0.67	5.50	0.67	5.73	0.64	0.01	1.05
Child Black	0.71	0.46	0.81	0.40	0.70	0.49	0.03	0.94
Child biracial	0.03	0.16	0.02	0.13	0.03	0.17	0.00	0.94
Child Hispanic	0.22	0.42	0.21	0.41	0.21	0.41	0.02	1.02
Child male	0.48	0.50	0.49	0.50	0.48	0.50	0.00	1.00
Parent age	36.10	7.83	33.98	6.85	36.53	7.11	−0.06	1.10
Parent years of education	13.31	2.94	12.83	2.57	13.46	2.85	−0.06	1.03
Parent single	0.46	0.50	0.57	0.50	0.48	0.49	−0.04	1.02
Parent Black	0.67	0.48	0.82	0.39	0.69	0.50	−0.05	0.96
Parent works full-time	0.11	0.32	0.17	0.39	0.10	0.30	0.03	1.07

Implications for Research, Practice, and Policy

This study is one of the first known efforts to identify classroom emotional support and organization as mechanisms explaining impacts of an SEL program on low-income racial/ethnic minority children's math achievement in first grade. The methodologies used can serve as examples for future researchers interested in examining mechanisms in randomized trials of school-based programs. In addition, practitioners can learn from these findings when developing or implementing SEL programs in urban elementary schools. When replicating or scaling up *INSIGHTS* and other SEL programs with similar theories of change, it will be important to consider how classroom emotional support and organization operate as key factors promoting student achievement. Practitioners can assess immediate impacts of SEL programs on classroom emotional support and organization internally at their school and adjust programming if it appears that expected improvements in classroom settings are not being made. In cases where resources are limited, school leaders can identify classrooms that already appear to be emotionally supportive and organized, even without intervention. Resources might be directed to support classroom improvements in settings that are at higher risk for poor student achievement in the short term. Such strategies will require substantial resources to continuously monitor improvements. Given efforts to make teacher assessments more amenable to rich and in-depth feedback (Kane & Staiger, 2012), linking SEL program implementation with assessments of classroom emotional support and organization can be made efficient at a broad scale.

Policy makers can also learn from this study. Primarily, with increased calls from state and federal governments for funding on SEL programs and teacher training (CASEL, 2014), there is a need to determine how to assess

TABLE 11

Balance Statistics for Classroom Organization Matching Procedure Measured at the Beginning of First Grade

Variable	Unmatched treated		Unmatched control		Matched control		Std. difference	Ratio of SDs
	M	SD	M	SD	M	SD		
Baseline classroom organization	5.15	0.78	4.00	0.93	5.08	0.72	0.09	1.08
Teacher years teaching	10.55	7.89	15.63	11.09	11.03	7.17	−0.06	1.10
Class size	18.44	3.09	19.04	4.40	18.73	3.40	−0.09	0.91
Number of adults in classroom	1.22	0.36	1.34	0.38	1.25	0.33	−0.08	1.09
Baseline reading achievement	17.19	8.28	18.18	7.94	17.87	7.66	−0.08	1.08
Baseline math achievement	14.63	4.86	15.09	5.09	14.84	5.33	−0.04	0.91
Baseline behavior problems	2.15	1.12	2.19	1.07	2.25	1.03	−0.09	1.09
Baseline sustained attention	11.60	6.26	11.10	4.28	11.21	6.40	0.06	0.98
Behavioral engagement	0.66	0.23	0.71	0.18	0.65	0.21	0.04	1.10
Off-task behaviors	0.16	0.09	0.14	0.09	0.16	0.09	0.00	1.00
Teacher reported reading competence	2.85	0.78	2.58	0.87	2.81	0.86	0.05	0.91
Teacher reported math competence	2.81	0.68	2.75	0.70	2.79	0.68	0.03	1.00
Student teacher conflict	1.92	1.01	1.67	0.94	1.88	1.02	0.04	0.99
Student teacher closeness	4.21	0.86	3.98	0.82	4.24	0.79	−0.03	1.09
Parent involvement total score	2.71	0.56	2.65	0.54	2.72	0.53	−0.02	1.06
Child negative reactivity	2.90	0.85	2.89	0.90	2.93	0.88	−0.04	0.97
Child task persistence	3.74	0.83	3.82	0.76	3.76	0.79	−0.02	1.05
Child withdrawal	2.51	0.90	2.33	0.92	2.49	0.91	0.02	0.99
Child motor activity	2.73	0.93	2.81	0.94	2.78	0.93	−0.05	1.00
Child age at study start	5.84	0.68	5.52	0.66	5.81	0.73	0.04	0.93
Child Black	0.77	0.42	0.75	0.43	0.75	0.38	0.05	1.11
Child biracial	0.03	0.16	0.20	0.14	0.03	0.17	0.00	0.94
Child Hispanic	0.24	0.43	0.21	0.40	0.22	0.39	0.05	1.10
Child male	0.48	0.50	0.48	0.50	0.46	0.47	0.04	1.06
Parent age	35.36	7.91	34.92	7.19	35.11	7.65	0.03	1.03
Parent years of education	13.27	2.72	12.98	2.79	13.11	2.76	0.06	0.99
Parent single	0.49	0.50	0.53	0.50	0.51	0.49	−0.04	1.02
Parent Black	0.75	0.45	0.76	0.44	0.76	0.41	−0.02	1.10
Parent works full-time	0.16	0.31	0.16	0.14	0.17	0.34	−0.03	0.91

TABLE 12

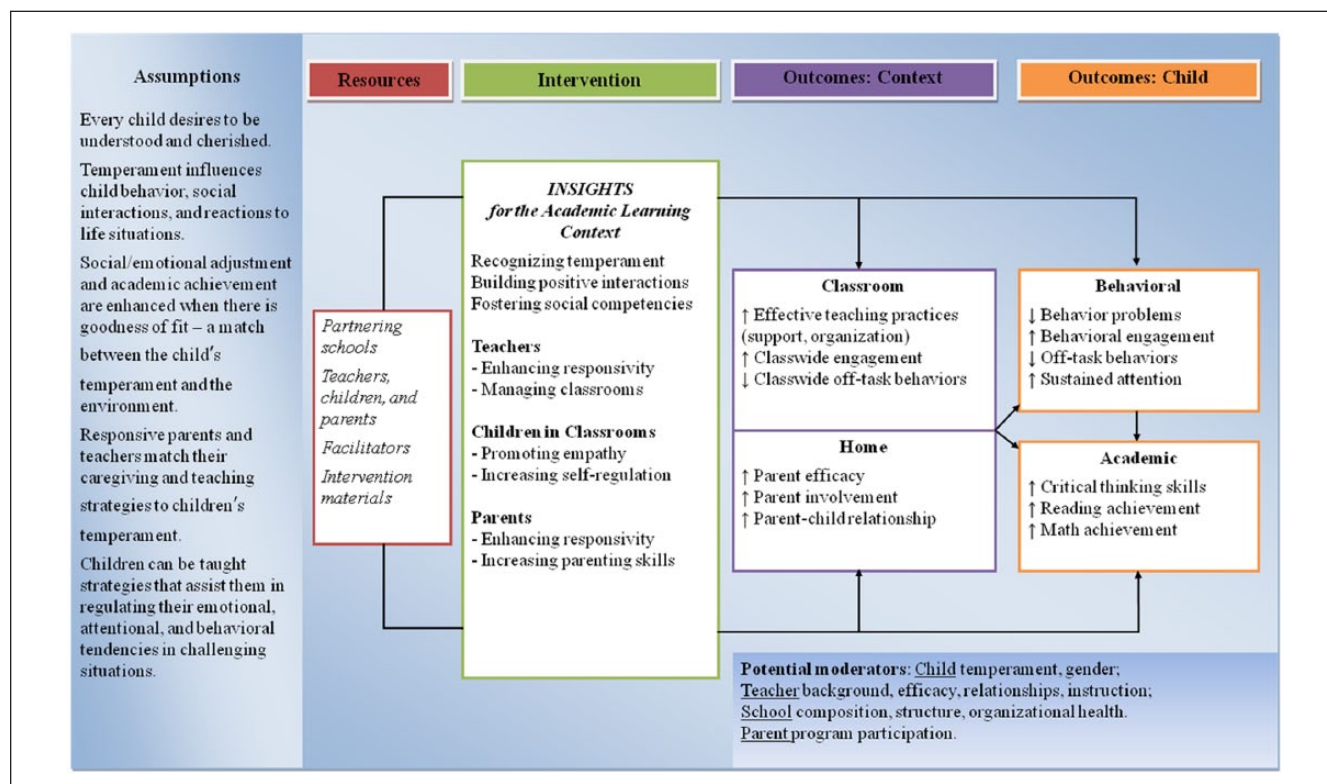*Principal Score Matching Procedure Predicting Achievement Mediated Through Improvements in Classroom Quality*

Fixed effects	Math in first grade		Reading in first grade	
	B	SE	B	SE
Classroom emotional support (postmatching)	3.80*	1.88	−1.27	1.99
Classroom organization (postmatching)	1.84*	0.73	3.03*	1.45

$^\dagger p < .10.$ *$p < .05.$ **$p < .01.$

the immediate success of such programs. This study demonstrates that it is important to measure classroom-level outcomes in addition to student achievement outcomes. Policy makers may consider assessing classroom-level 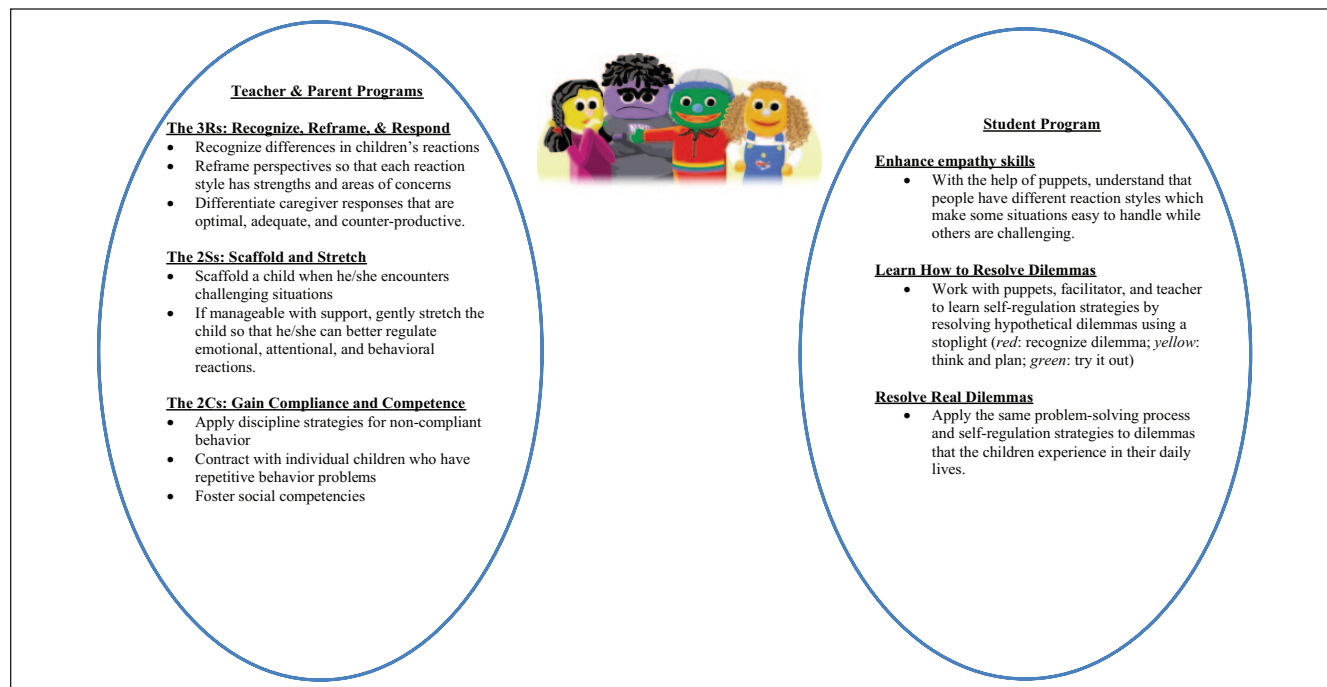outcomes in determining accountability criteria for SEL programs in public schools. Taken together, results suggest the importance of continually monitoring the progress of SEL programs in schools not just on students themselves, but on the classroom settings in which they are embedded.

Appendix A

INSIGHTS Theory of Change

Appendix B

INSIGHTS Curriculum Overview

Funding

Note

1. The *F* statistics for treatment predicting classroom emotional support and organization are as follows: kindergarten emotional support: 5.51, $p = .28$; kindergarten organization: 4.31, $p = .61$; first grade emotional support: 13.41, $p < .01$; first grade organization: 11.92, $p < .05$. An *F* statistic greater than 10 typically is thought to describe a strong instrument.

References

Abry, T., Hulleman, C. S., & Rimm-Kaufman, S. E. (2015). Using indices of fidelity to intervention core components to identify program active ingredients. *American Journal of Evaluation*, *36*(3), 320–338.

Abry, T., Rimm-Kaufman, S. E., Larsen, R. A., & Brewer, A. J. (2013). The influence of fidelity of implementation on teacher–student interaction quality in the context of a randomized controlled trial of the Responsive Classroom approach. *Journal of School Psychology*, *51*(4), 437–453.

Angrist, J. D., Imbens, G. W., & Rubin, D. B. (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association*, *91*(434), 444–455.

Banilower, E. R., Smith, P. S., Weiss, I. R., Malzahn, K. A., Campbell, K. M., & Weis, A. M. (2013). *Report of the 2012 National Survey of Science and Mathematics Education*. Ridgewood, NJ: Horizon Research.

Barnett, W. S. (2011). Effectiveness of early educational intervention. *Science*, *333*(6045), 975–978.

Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, *51*(6), 1173–1182.

Bierman, K. L., Nix, R. L., Heinrichs, B. S., Domitrovich, C. E., Gest, S. D., Welsh, J. A., & Gill, S. (2014). Effects of Head Start REDI on children's outcomes 1 year later in different kindergarten contexts. *Child Development*, *85*(1), 140–159.

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

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.

Bronfenbrenner, U., & Morris, P. A. (1998). The ecology of developmental processes. In W. Damon (Series Ed.) & R. M. Lerner (Vol. Ed.), *Handbook of child psychology: Vol. 1. Theoretical models of human development* (5th ed., pp. 993–1028). New York, NY: John Wiley.

Brown, J. L., Jones, S. M., LaRusso, M. D., & Aber, J. L. (2010). Improving classroom quality: Teacher influences and experimental impacts of the 4rs program. *Journal of Educational Psychology*, *102*(1), 153–167.

Burchinal, M., Vandergrift, N., Pianta, R., & Mashburn, A. (2010). Threshold analysis of association between child care quality and child outcomes for low-income children in pre-kindergarten programs. *Early Childhood Research Quarterly*, *25*(2), 166–176.

Cadima, J., Doumen, S., Verschueren, K., & Buyse, E. (2015). Child engagement in the transition to school: Contributions of self-regulation, teacher–child relationships and classroom climate. *Early Childhood Research Quarterly*, *32*, 1–12.

Cappella, E., Hamre, B. K., Kim, H. Y., Henry, D. B., Frazier, S. L., Atkins, M. S., & Schoenwald, S. K. (2012). Teacher consultation and coaching within mental health practice: Classroom and child effects in urban elementary schools. *Journal of Consulting and Clinical Psychology*, *80*(4), 597–610.

Cappella, E., O'Connor, E. E., McCormick, M. P., Turbeville, A. R., Collins, A., & McClowry, S. G. (in press). Classwide efficacy of INSIGHTS: Observed student behaviors and teacher practices in kindergarten and first grade. *Elementary School Journal*.

Catalano, R. F., Berglund, M. L., Ryan, J. A. M., Lonczak, H. S., & Hawkins, J. D. (2002). Positive youth development in the United States: Research findings on evaluations of positive youth development. *Prevention & Treatment*, *5*(15). Retrieved from http://journals.apa.org/prevention/volume5/pre0050015a.html

Collaborative for Academic and Social Emotional Learning. (2014). *CASEL's Collaborative Districts Initiative (CDI)*. Retrieved from http://static.squarespace.com/static/513f79f9e4b05ce7b70e9673/t/52aa8d55e4b0e46ab623a595/1386909013728/cdifactsheet-20131212.pdf

Conners, C. K., Sitarenios, G., Parker, J. D., & Epstein, J. N. (1998). The revised Conners' Parent Rating Scale (CPRS-R): Factor structure, reliability, and criterion validity. *Journal of Abnormal Child Psychology*, *26*(4), 257–268.

Crosnoe, R., Morrison, F., Burchinal, M., Pianta, R., Keating, D., Friedman, S. L., & Clarke-Stewart, K. A. (2010). Instruction, teacher–student relations, and math achievement trajectories in elementary school. *Journal of Educational Psychology*, *102*(2), 407–417.

Curby, T. W., Rimm-Kaufman, S. E., & Ponitz, C. C. (2009). Teacher–child interactions and children's achievement trajectories across kindergarten and first grade. *Journal of Educational Psychology*, *101*(4), 912–925.

Diamond, A., & Lee, K. (2011). Interventions shown to aid executive function development in children 4–12 years old. *Science*, *333*, 959–964.

DiPerna, J. C., & Elliott, S. N. (2000). *The Academic Competence Evaluation Scales*. San Antonio, TX: Psychological Corporation.

Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., . . . Japel, C. (2007). School readiness and later achievement. *Developmental Psychology*, *43*(6), 1428–1446.

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.

Eisenberg, N., Valiente, C., & Eggum, N. D. (2010). Self-regulation and school readiness. *Early Education and Development*, *21*(5), 681–698.

Enders, C. K. (2013). Dealing with missing data in developmental research. *Child Development Perspectives*, *7*(1), 27–31.

Entwisle, D. R., & Alexander, K. L. (1998). Facilitating the transition to first grade: The nature of transition and research on factors affecting it. *Elementary School Journal*, *98*, 351–364.

Epstein, D., & Miller, R. T. (2011). *Slow off the mark: Elementary school teachers and the crisis in science, technology, engineering, and math education*. Washington, DC: Center for American Progress.

Eyberg, S. M., & Pincus, D. (1999). *Eyberg Child Behavior Inventory and Sutter-Eyberg Student Behavior Inventory–Revised: Professional manual*. Odessa, FL: Psychological Assessment Resources.

Fantuzzo, J., Bulotsky-Shearer, R., McDermott, P., McWayne, C., Frye, D., & Perlman, S. (2007). Investigation of dimensions of social-emotional classroom behavior and school readiness for low-income urban preschool children. S*chool Psychology Review*, *36*, 44–62.

Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. New York, NY: Cambridge University Press.

Gennetian, L. A., Magnuson, K., & Morris, P. A. (2008). From statistical associations to causation: What developmentalists can learn from instrumental variables techniques coupled with experimental data. *Developmental Psychology*, *44*(2), 381–394.

Greenberg, M. T., Weissberg, R. P., O'Brien, M. U., Zins, J. E., Fredericks, L., Resnik, H., & Elias, M. J. (2003). Enhancing school-based prevention and youth development through coordinated social, emotional, and academic learning. *American Psychologist, 58*(6–7), 466–474.

Hagelskamp, C., Brackett, M. A., Rivers, S. E., & Salovey, P. (2013). Improving classroom quality with the ruler approach to social and emotional learning: Proximal and distal outcomes. *American Journal of Community Psychology, 51*(3–4), 530–543.

Hamre, B., Hatfield, B., Pianta, R., & Jamil, F. (2014). Evidence for general and domain-specific elements of teacher–child interactions: Associations with preschool children's development. *Child development*, *85*(3), 1257–1274.

Hamre, B. K., Pianta, R. C., Downer, J. T., DeCoster, J., Mashburn, A. J., Jones, S. M., . . . Hamagami, A. (2013). Teaching through Interactions. *Elementary School Journal*, *113*(4), 461–487.

Hawkins, J. D., Kosterman, R., Catalano, R. F., Hill, K. G., & Abbott, R. D. (2005). Promoting positive adult functioning through social development intervention in childhood: Long-term effects from the Seattle Social Development Project. *Archives of Pediatrics & Adolescent Medicine*, *159*(1), 25–31.

Hill, J. L., Brooks-Gunn, J., & Waldfogel, J. (2003). Sustained effects of high participation in an early intervention for low-birth-weight premature infants. *Developmental Psychology*, *39*(4), 730–744.

Hill, J. L., Waldfogel, J., Brooks-Gunn, J., & Han, W. J. (2005). Maternal employment and child development: A fresh look using newer methods. *Developmental Psychology*, *41*(6), 833–850.

Hill, J., Weiss, C., & Zhai, F. (2011). Challenges with propensity score strategies in a high-dimensional setting and a potential alternative. *Multivariate Behavioral Research*, *46*(3), 477–513.

Hughes, J. N. (2011). Longitudinal effects of teacher and student perceptions of teacher–student relationship qualities on academic adjustment. *Elementary School Journal*, *112*(1), 38–60.

Jennings, P. A., & Greenberg, M. T. (2009). The prosocial classroom: Teacher social and emotional competence in relation to student and classroom outcomes. *Review of Educational Research*, *79*(1), 491–525.

Jeynes, W. (2012). A meta-analysis of the efficacy of different types of parental involvement programs for urban students. *Urban Education*, *47*(4), 706–742.

Jones, S. M., & Bouffard, S. M. (2012). Social and emotional learning in schools: From programs to strategies. *Social Policy Report*, *26*(4), 1–33.

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*, 533–554.

Jones, S. M., Bub, K. L., & Raver, C. C. (2013). Unpacking the Black box of the Chicago School Readiness Project intervention: The mediating roles of teacher–child relationship quality and self-regulation. *Early Education and Development*, *24*(7), 1043–1064.

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

Leak, J., Duncan, G. J., Li, W., Magnuson, K., Schindler, H., & Yoshikawa, H. (2010, November). *Is timing everything? How early childhood education program impacts vary by starting age, program duration and time since the end of the program*. Presentation at the biennial meeting for the Society for Research on Child Development, Montreal, Quebec, Canada.

Manz, P. H., Fantuzzo, J. W., & Power, T. J. (2004). Multidimensional assessment of family involvement among urban elementary students. *Journal of School Psychology*, *42*(6), 461–475.

Mashburn, A. J., Hamre, B. K., Downer, J. T., & Pianta, R. C. (2006). Teacher and classroom characteristics associated with teachers' ratings of prekindergartners' relationships and behaviors. *Journal of Psychoeducational Assessment*, *24*(4), 367–380.

Mashburn, A. J., Pianta, R. C., Hamre, B. K., Downer, J. T., Barbarin, O. A., Bryant, D., . . . Howes, C. (2008). Measures of classroom quality in prekindergarten and children's development of academic, language, and social skills. *Child Development*, *79*(3), 732–749.

McClowry, S. G. (2002). The temperament profiles of school-age children. *Journal of Pediatric Nursing*, *17*(1), 3–10.

McClowry, S. G., Rodriguez, E. T., Tamis-LeMonda, C. S., Spellmann, M. E., Carlson, A., & Snow, D. L. (2013). Teacher/student interactions and classroom behavior: The role of student temperament and gender. *Journal of Research in Childhood Education*, *27*, 283–301.

McCormick, M. P., O'Connor, E. E., Cappella, E., & McClowry, S. G. (2013). Teacher-child relationships and academic achievement in kindergarten and first grade: A multilevel propensity score model approach. *Journal of School Psychology*, *51*(5), 611–624.

Morrissey, T. W., Hutchison, L., & Winsler, A. (2014). Family income, school attendance, and academic achievement in elementary school. *Developmental Psychology*, *50*(3), 741–753.

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

O'Connor, E., & McCartney, K. (2007). Attachment and cognitive skills: An investigation of mediating mechanisms. *Journal of Applied Developmental Psychology*, *28*(5), 458–476.

Page, L. (2012). Principal stratification as a framework for investigating meditational processes in experimental settings. *Journal of Research on Educational Effectiveness*, *5*(3), 215–244.

Perry, K. E., Donohue, K. M., & Weinstein, R. S. (2007). Teaching practices and the promotion of achievement and adjustment in first grade. *Journal of School Psychology*, *45*(3), 269–292.

Pianta, R. C. (2001). *STRS: Student–Teacher Relationship Scale: Professional manual*. Odessa, FL: Psychological Assessment Resources.

Pianta, R. C., Belsky, J., Vandergrift, N., Houts, R., & Morrison, F. J. (2008). Classroom effects on children's achievement trajectories in elementary school. *American Educational Research Journal*, *45*(2), 365–397.

Pianta, R. C., La Paro, K., & Hamre, B. K. (2008). *Classroom Assessment Scoring System (CLASS)*. Baltimore, MD: Brookes.

Ponitz, C. C., McClelland, M. M., Matthews, J. S., & Morrison, F. J. (2009). A structured observation of behavioral self-regulation and its contribution to kindergarten outcomes. *Developmental Psychology*, *45*(3), 605–619.

Querido, J. G., & Eyberg, S. M. (2003). Psychometric properties of the Sutter-Eyberg Student Behavior Inventory–Revised with preschool children. *Behavior Therapy*, *34*, 1–15.

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

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), 3–19.

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.

Razza, R. A., Martin, A., & Brooks-Gunn, J. (2010). Associations among family environment, sustained attention, and school readiness for low-income children. *Developmental Psychology*, *46*(6), 1528–1542.

Reardon, S. F. (2011). The widening academic achievement gap between the rich and the poor: New evidence and possible explanations. In G. J. Duncan & R. J. Murnane (Eds.), *Whither opportunity* (pp. 91–116). New York, NY: Russell Sage.

Reardon, S. F., & Raudenbush, S. W. (2013). Under what assumptions do site-by-treatment instruments identify average causal effects? *Sociological Methods*, *42*(2), 143–163.

Rimm-Kaufman, S. E., Curby, T., Grimm, K., Nathanson, L., & Brock, L. (2009). The contribution of children's self-regulation and classroom quality to children's adaptive behaviors in the kindergarten classroom. *Developmental Psychology*, *45*, 958–972.

Rivers, S. E., Brackett, M. A., Reyes, M. R., Elbertson, N. A., & Salovey, P. (2013). Improving the social and emotional climate of classrooms: A clustered randomized controlled trial testing the RULER Approach. *Prevention Science*, *14*(1), 77–87.

Robinson, J. P., & Lubienski, S. T. (2011). The development of gender achievement gaps in mathematics and reading during elementary and middle school examining direct cognitive assessments and teacher ratings. *American Educational Research Journal*, *48*(2), 268–302.

Roid, G. H., & Miller, L. J. (1997). *Leiter international performance scale-revised: Examiners manual*. Wood Dale, IL: Stoelting.

Roorda, D. L., Koomen, H. M., Spilt, J. L., & Oort, F. J. (2011). The influence of affective teacher–student relationships on students' school engagement and achievement a meta-analytic approach. *Review of Educational Research*, *81*(4), 493–529.

Rubin, D. B., & Little, R. J. (2002). *Statistical analysis with missing data*. Hoboken, NJ: John Wiley.

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, *7*(2), 147–177.

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

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*. Washington, DC: Author.

Spilt, J. L., Hughes, J. N., Wu, J. Y., & Kwok, O. M. (2012). Dynamics of teacher–student relationships: Stability and change across elementary school and the influence on children's academic success. *Child Development*, *83*(4), 1180–1195.

Spivak, A. L., & Farran, D. L. (2014, March). *Young children's social competence: Longitudinal evaluation of the tools of the mind curriculum and teacher and child behaviors in prekindergarten classrooms*. Presentation at the Society for Research on Educational Evaluation, Washington, DC.

Südkamp, A., Kaiser, J., & Möller, J. (2012). Accuracy of teachers' judgments of students' academic achievement: A meta-analysis. *Journal of Educational Psychology*, *104*(3), 743–762.

Tatto, M. T., & Senk, S. (2011). The mathematics education of future primary and secondary teachers: Methods and findings from the Teacher Education and Development Study in Mathematics. *Journal of Teacher Education*, *62*(2), 121–137.

Tseng, V., & Seidman, E. (2007). A systems framework for understanding social settings. *American Journal of Community Psychology, 39*(3–4), 217–228.

Ursache, A., Blair, C., & Raver, C. C. (2012). The promotion of self-regulation as a means of enhancing school readiness and early achievement in children at risk for school failure. *Child Development Perspectives*, *6*(2), 122–128.

Valiente, C., Eisenberg, N., Spinrad, T. L., Haugen, R. G., Thompson, M. S., & Kupfer, A. (2013). Effortful control and impulsivity as concurrent and longitudinal predictors of academic achievement. *Journal of Early Adolescence*, *33*(7), 946–972.

van Lier, P. A., Vitaro, F., Barker, E. D., Brendgen, M., Tremblay, R. E., & Boivin, M. (2012). Peer victimization, poor academic achievement, and the link between childhood externalizing and internalizing problems. *Child Development*, *83*(5), 1775–1788.

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*, *49*(5), 471–488.

Webster-Stratton, C., & Taylor, T. (2001). Nipping early risk factors in the bud: Preventing substance abuse, delinquency, and violence in adolescence through interventions targeted at young children (0–8 years). *Prevention Science*, *2*(3), 165–192.

Wechsler, D. (1989). *WPPSI-R: Wechsler Preschool and Primary Scale of Intelligence*. San Diego, CA: Psychological Corporation.

Woodcock, R. W., & Johnson, M. B. (1989). *WJ-R Tests of Cognitive Ability*. Itasca, IL: Riverside.

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock–Johnson Psychoeducational Battery–Third Edition*. Chicago, IL: Riverside.

Zins, J. E., & Elias, M. J. (2006). Social and emotional learning. In G. G. Bear & K. M. Minke (Eds.), *Children's needs III: Development, prevention, and intervention* (pp. 1–13). Bethesda, MD: National Association of School Psychologists.

Authors

Meghan P. Mccormick is a research associate in the families and children policy area at MDRC. She received a PhD in applied psychology with a concentration in quantitative analysis from New York University.

Elise Cappella is an associate professor of applied psychology at New York University. She received a PhD in clinical and community psychology from the University of California, Berkeley.

Erin E. O'connor is an associate professor of teaching and learning at New York University. She received an EdD in human development and education from the Harvard Graduate School of Education.

Sandee G. Mcclowry is a professor of applied psychology and teaching and learning at New York University. She received a PhD in family nursing theory from the University of California, San Francisco.