CrossMark

ORIGINAL ARTICLE

Context Matters for Social-Emotional Learning: Examining Variation in Program Impact by Dimensions of School Climate

**Meghan P. McCormick[1] · Elise Cappella[2] ·
Erin E. O'Connor[3] · Sandee G. McClowry[2]**

Abstract This paper examines whether three dimensions of school climate—leadership, accountability, and safety/respect—moderated the impacts of the *INSIGHTS* program on students' social-emotional, behavioral, and academic outcomes. Twenty-two urban schools and $N = 435$ low-income racial/ethnic minority students were enrolled in the study and received intervention services across the course of 2 years, in both kindergarten and first grade. Intervention effects on math and reading achievement were larger for students enrolled in schools with lower overall levels of leadership, accountability, and safety/respect at baseline. Program impacts on disruptive behaviors were greater in schools with lower levels of accountability at baseline; impacts on sustained attention were greater in schools with lower levels of safety/respect at baseline. Implications for Social-Emotional Learning program implementation, replication, and scale-up are discussed.

Keywords Social-emotional learning · Academic achievement · School climate · School context

Introduction

Social-emotional skills (e.g., attention, behavioral and emotional regulation, conflict resolution, social skills) are critical for academic success (Greenberg et al. 2003; Raver

et al. 2011). Because of their wide-ranging impact, there is growing political and consumer support for teaching social-emotional skills during early elementary school. For example, Paul Tough's recent book *Why Children Succeed* (2012), which identifies qualities such as perseverance and self-control as critical antecedents of achievement, earnings, and overall well-being, spent 12 weeks on the *New York Times* best sellers' list. Support from education stakeholders has encouraged the development and expansion of school-based Social-Emotional Learning (SEL) programs, designed to improve not only students' social-emotional skills but also their academic development. Recently, a bipartisan group of lawmakers introduced the Academic, Social, and Emotional Learning Act to the 113th Congress to expand the availability of evidence-based programs that teach students social-emotional skills. Advocates have been particularly interested in implementing SEL programs in low-income urban settings where students are more likely to start school with lower levels of social-emotional and academic skills than their more affluent peers (Kahn 2013; Raver 2002).

Despite recent interest in their expansion, however, inconsistent evidence exists that SEL programs improve students' academic achievement over and above typical educational practice (SRCDC 2010). One possible constraint to understanding mixed findings is limited information on how program effects differ across school settings. It could be that SEL programs are highly effective in some types of schools and less so in others, thus confounding overall understanding of program efficacy. Moreover, although some work has considered how demographic characteristics—like school poverty—differentiate SEL program impacts on student outcomes, fewer studies have examined the moderating role of the school-level social processes (e.g., social norms, structures of

✉ Meghan P. McCormick
meghan.mccormick@mdrc.org

[1] MDRC, 16 East 34th St., New York, NY 10016, USA

[2] Department of Applied Psychology, New York University, New York, NY, USA

[3] Department of Teaching and Learning, New York University, New York, NY, USA

🍎 Springer

relationships) within which SEL interventions are embedded (Tseng and Seidman 2007). Because one goal of SEL programs is to improve the quality of interactions among individuals in schools and within classrooms (Durlak et al. 2011), school-level social processes are important to consider when examining SEL program impacts on student academic and social-emotional outcomes.

School climate reflects the norms, goals, values, interpersonal relationships, teaching and learning practices, and organizational structures of the school (NSCC 2007; Thapa et al. 2013). It is a useful construct for understanding social processes at the school-level. A prevention research perspective suggests that schools with the poorest climates have the most to gain from intervention that explicitly targets social interactions (e.g., Cicchetti and Aber 1998; Van Lier et al. 2004). Contrasting work, however, argues that SEL programs will be most effective in settings where extant norms *already* support positive academic and social-behavioral development (Aber et al. 1998; Hughes et al. 2005).

In view of this lack of clarity, the current study examines whether key dimensions of school climate—leadership, accountability, and safety/respect (see Nathanson et al. 2013a, b)—moderate impacts of one SEL program— *INSIGHTS into Children's Temperament*—on low-income urban kindergarten and first grade students' math and reading achievement, sustained attention, and behavior problems. A randomized trial using intent-to-treat analyses identified empirical support for *INSIGHTS* on these four student outcomes (see O'Connor et al. 2014). It is unclear, however, whether students in different schools benefited similarly. Learning about variation in *INSIGHTS*' impacts can inform targeting of programs, and allocation of funds to schools that have the most to gain from implementation of an SEL program.

Social-Emotional Learning Programs and Student Academic Achievement

Young children who successfully develop core social-emotional competencies, such as self management, self awareness, social awareness, relationship skills, and responsible decision making, are most likely to successfully navigate the transition to elementary school (Rimm-Kaufman et al. 2007). Children raised in poverty are at risk for exhibiting emotional and social difficulties at the start of elementary school (Cooper et al. 2011). Childhood poverty is associated with racial/ethnic minority status, particularly in urban neighborhoods (Kumanyika and Grier 2006). Given education reform efforts in urban centers, efforts to expand programs that support the social-emotional and academic development of low-income minority children are growing (Durlak et al. 2011; Kahn 2013).

Termed Social-Emotional Learning (SEL) programs, these school-based interventions aim to enhance an interrelated set of cognitive, emotional, and behavioral skills regarded as foundational for academic performance (Zins and Elias 2006). 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 and Elias 2006). Such competencies promote children's engagement in instructional activities and the classroom setting that, in turn, are expected to enhance academic achievement (Eisenberg et al. 2010). Although most SEL programs employ classroom-based curricula to directly target students' social-emotional skills, a host of other interventions use multi-level program delivery models to provide services in school and family settings (Greenberg et al. 2003).

Universal SEL programs, tested in low-income pre-K and elementary schools, have been successful in improving students' social-emotional skills (e.g., 4Rs, Jones et al. 2011; CSRC Raver et al. 2011; Incredible Years, Webster-Stratton et al. 2008). Other studies have shown that SEL program can benefit overall classroom quality (4Rs, Brown et al. 2010; Cappella et al. 2012). For example, a recent study by Hagelskamp et al. (2013) evaluating the efficacy of the RULER intervention, which aims to improve students' emotional literacy, identified positive effects of the intervention on classroom emotional support, organization, and instructional support 2 years post-intervention.

However, it is less clear whether SEL programs improve academic achievement. While a large-scale meta-analysis by Durlak et al. (2011; $N = 213$ studies) found small overall effects of SEL programs on academic performance in elementary school (average E.S. $= 0.27$), a 2010 report by the Institute for Education Science showed no positive impacts on student achievement for students in third to fifth grade (SCDRC 2010).

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 proximal short-term effects. As such, it may be that longer-term follow-up is needed to examine impacts on academic outcomes. In addition, even given correlations between social-emotional skills and achievement, it is possible that effects of interventions aimed at enhancing social-emotional development do not "spillover" to improve achievement as measured with the assessments included in the SCDRC study. For example, a number of the local programs tested in the broader SCDRC study did identify academic impacts for measures not included in the larger study (e.g., 4Rs: Jones et al. 2011; Positive Action:

Snyder et al. 2009). One understudied possibility for mixed findings is that most impact studies of SEL programs examine average program impacts across *all* schools recruited to participate in the research study. In the context of a randomized control trial, this design—called an "intent to treat model"—is considered the gold standard for determining whether an intervention works (Shadish et al. 2002). Yet, findings from such a study provide little information about the contexts and implementation conditions under which an SEL program may have been most effective in boosting students' academic, social-emotional and behavioral outcomes.

Bioecological systems theory (Bronfenbrenner and Morris 1998) reminds us that SEL programs are embedded in larger contexts, notably classrooms and schools, which have direct, indirect, and interactive influences on children's development. Theoretically then, the school context will influence the extent to which that program effectively promotes children's outcomes. Empirically testing such a theory is critical for determining whether there are settings where SEL program impacts are most prominent, as well as the types of schools that should be targeted for future program implementation (Supplee et al. 2013).

School Settings and Effects of SEL Programs

Although past research is limited, some evaluation studies have considered how school and classroom settings moderate impacts of SEL programs on student outcomes (Aber et al. 1998; Bierman et al. 2010; Hughes et al. 2005). For example, Bierman et al. (2010) examined school poverty as a moderator in the evaluation of the Fast Track version of PATHS, an SEL program for low-income children in first through third grade. The study found that impacts of PATHS on academic engagement were stronger in schools with lower percentages of students in poverty. In discussing this finding, the authors hypothesized that the PATHS program could not be properly implemented in school contexts with higher percentages of students raised in poverty.

In addition to measures of economic disadvantage, other salient setting-level conditions are likely to affect setting-level variation. Indeed, Tseng and Seidman (2007) argue that, over and above financial resources and organization of resources, social processes that take place within settings (e.g., social norms, relationships, and interactions) play a key role in influencing both individuals and contexts. School climate operationalizes social processes by measuring the social and organizational structure of the school in safety, teaching and learning, interpersonal relationships, and institutional environment (Cohen et al. 2009).

A wide body of research has consistently linked school climate to student academic, behavioral, and mental health outcomes (Brand et al. 2008; Espelage et al. 2014). Positive school climate is associated with higher grade point averages, standardized test scores, reading levels, academic writing, and school adjustment (Brand et al. 2008; Garrison 2004). Within low-income urban neighborhoods, school climate may be particularly salient. For example, a recent study by McCoy et al. (2013) found that higher levels of neighborhood crime—which may occur in low-income urban settings—predicted decreases in schools' social-emotional learning and physical/emotional safety, two salient dimensions of climate. McCoy et al. found that greater levels of social-emotional learning, physical/emotional safety and academic rigor predicted greater school-level achievement over time. Given these links, and their possible importance for high-need schools, policymakers have begun to put increased emphasis on measuring school climate as a part of accountability and assessment efforts (e.g., IES Safe and Supportive Schools grant program).

Some researchers have also considered the contribution of distinct dimensions of school climate to students' achievement. For example, Bryk et al. (2010) drew on extensive longitudinal survey and administrative data in Chicago to identify school-level factors that predicted student achievement. Five components of climate, termed "essential supports" were found: school leadership, parent and community ties, professional capacity of the faculty, school learning climate, and instructional guidance (Bryk et al. 2010). The New York City (NYC) Department of Education condensed the five supports and used a teacher-reported survey to collect information on three dimensions—leadership, accountability, and safety/respect. In the NYC framework, leadership represents the extent to which school leaders provide instructional support and engage in trusting relationships with staff. Accountability describes teacher perceptions of high academic standards for student work at their school. Finally, safety/respect is the extent to which teachers feel their school provides students and staff with physical and emotional safety. In a validation study, Nathanson et al. (2013a, b) found associations between these dimensions and elementary school-level math and reading achievement.

Although the research on school climate is growing, few studies have examined climate as a moderator of SEL program impacts on student achievement. In schools characterized by higher levels of leadership, accountability, and safety/respect, SEL programs may be less efficacious because teachers are already receiving the relational and institutional supports they need to be successful in enhancing children's academic, social-emotional, and behavioral development. Alternatively, SEL interventions might be more efficacious in schools with positive climates if programs can only be successfully implemented if sufficient contextual supports are in place (Hughes et al.

2005). With little empirical agreement about the school-level conditions that best support the efficacy of SEL programs for improving student outcomes, it is critical to consider the moderating role of school climate when examining intervention effects. Findings can inform future intervention design, development, and implementation (Supplee et al. 2013).

Current Study: Focus on *INSIGHTS* into Children's Temperament

This study will examine moderated program impacts for one particular SEL program—*INSIGHTS into Children's Temperament*—a comprehensive intervention with teacher, parent, and classroom programs. *INSIGHTS* provides teachers and parents with a temperament framework for supporting the individual differences of children. Temperament is an individual's consistent reaction style of responding to people, events, and other environmental stimuli, particularly those involving stress or change (McClowry 2014). Temperament is biologically based, multidimensional, and relatively stable through childhood (Rothbart and Bates 2006). Key to temperament theory is the concept of goodness of fit, or notion that it is important for a child's temperament to be in consonance with the demands, expectations, and opportunities of the child's environment (Chess and Thomas 1984). Although temperament itself should not be targeted by intervention, the environment can be modified to improve goodness of fit.

Using this framework, *INSIGHTS* helps parents and teachers recognize a child's temperament and respond with warmth and discipline strategies that support adaptive social-emotional and behavioral outcomes (McClowry et al. 2005; McClowry et al. 2010). Primary grade students also participate in classroom curricula designed to enhance empathy for individuals with different temperaments and to use problem-solving techniques when confronted with daily dilemmas (see O'Connor et al. 2014 for more information). There is empirical evidence to support *INSIGHTS'* theory of change. As discussed earlier, O'Connor et al. (2014) found that *INSIGHTS* improved low-income racial/ethnic minority students' math and reading achievement, and there was correlational evidence that gains in sustained attention, and reductions in behavior problems mediated these impacts. These analyses, however, did not consider heterogeneity of program impacts across settings.

The rich data from a large randomized trial of the *INSIGHTS* program, coupled with school-level administrative information and teacher reports of school climate, provide a unique opportunity to explicitly test the school-level conditions under which SEL program impacts on student achievement, sustained attention, and behavior problems were strongest. Using a sample of kindergarten and first grade students from low-income urban elementary schools, this study will examine whether school leadership, accountability, and safety/respect moderated impacts of *INSIGHTS* on children's math and reading achievement, sustained attention, and behaviors. Findings will elucidate whether critical dimensions of school climate explain heterogeneity of SEL program impacts in low-income urban elementary schools.

Method

Three cohorts of urban elementary schools entered the *INSIGHTS* study over three consecutive years; each cohort participated for 2 years of intervention and data collection. Kindergarten classrooms participated during Year 1; first grade classrooms took part in Year 2.

Data for the current study come from multiple sources. Teacher reports of school climate were drawn from the New York City Department of Education teacher survey data collected from 2008 to 2010. School demographic characteristics come from publicly available administrative records (NYCDOE 2014). Information on individual students and *INSIGHTS'* implementation comes from a variety of sources including parent and teacher reports and observations.

Participants and Setting

This study took place in 22 public elementary schools in New York City, composed of majority low-income students. One hundred and twenty teachers and 435 students enrolled in the study. Most teachers were female (94.2 %). The teachers identified as Hispanic or Latino (11.9 %), black or African American (56.4 %), white (24.3 %), and mixed race/other (7 %). Most classrooms were led by one teacher. Some classrooms that included children with individualized education plans had two teachers. All teachers reported having earned a bachelor's degree; ninety-six percent had a master's degree. All classrooms were regular education, with an average of 16.57 students ($SD = 3.54$).

Ninety-one percent of children were age five or six when they enrolled in the study ($M = 5.38$ $SD = 0.61$). Half (52 %) of the children were male. Eighty-seven percent of children qualified for free or reduced lunch. Seventy-five percent of children were black, non-Hispanic, 16 % were Hispanic, non-black, and 9 % were biracial. Most parent participants were the children's biological mothers (84 %). Approximately 28 % of adult respondents had less than a high school degree; 26 % had at least a high school degree or GED; 24 % had at least some college experience; and the remaining 22 % had graduated from a 2 or 4 year college.

Children enrolled in the study were similar in demographic characteristics to the other students at the schools who were invited to the study but did not participate. 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 ($M = 0.80$, $SD = 0.16$). Schools had an average attendance rate of 86.26 % ($SD = 0.19$) and averaged 465 students ($SD = 158.46$).

Measures

Data used in this study were multi-informant and longitudinal. School climate and demographics were assessed in the spring prior to program implementation. Time 1 (T1) data were collected in the winter of the kindergarten year prior to 10 weeks of intervention. Time 2 (T2) data were collected following intervention in the spring of kindergarten. Time 3 (T3) data were collected in the fall of first grade prior to 10 weeks of intervention. Time 4 (T4) data were collected after the first grade intervention, followed by Time 5 (T5) data in late spring. Treatment was measured as an indicator in all analyses (1 = INSIGHTS; 0 = attention-control).

Outcome Variables

Reading 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 (WJ-III; Woodcock et al. 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. Possible scores range from 0 to 76 on the Letter Word ID, and from 0 to 64 on the Applied Problems test. The WJ-III is a nationally normed and widely used achievement test with demonstrated internal consistency. Subscales have internal consistencies ranging from 0.80 to 0.90. In this study, average reliability across the five time points for the Letter-Word ID subtest was 0.84; average reliability for the Applied Problems subtest was 0.88.

Child sustained attention was measured with the Attention Sustained subtest from the Leiter International Performance Scale (Roid and Miller 1997). Children were shown a page with pictures 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. Children were given a limited amount of time to perform four trials (30 s for the first three trials and 60 s for the fourth) but were not scored on speed. The number of incorrect responses was subtracted from the number of correct responses for an overall score. The task has demonstrated high internal consistency and validity (Roid and Miller 1997).

Child behavior problems were measured with the 36-item Sutter–Eyberg Student Behavior Inventory, the teacher-report version of the Eyberg Child Behavior Inventory (Eyberg and Pincus 1999). On a frequency scale ranging from 1 to 7 (1 = *never*, 4 = *sometimes*; 7 = *always*), teachers reported how often each consented child engaged in a range of problematic behaviors. A mean score was calculated, with possible scores ranging from 1 to 7. Querido and Eyberg (2003) showed validity evidence for the measure. The average Cronbach's α in this study was 0.97 across time points.

Moderators

The school climate moderators—leadership, accountability, and safety/respect—were all measured using aggregated reports from all teachers in the school. In other words, within each school, all teachers' perceptions were averaged, by school climate dimension, for an overall school score. Within each school, 92 to 100 % of teachers reported on their perceptions of their school (Nathanson et al. 2013a, b). See Table 1 for a list of the teacher survey items. The school climate dimensions used in this study—leadership, accountability, and safety/respect—have shown initial evidence of reliability and validity (see Nathanson et al. 2013a, b; Rockoff and Speroni 2008). In prior work, researchers have found moderate to large correlations between dimensions of school climate assessed in large administrative surveys (e.g., Bryk et al. 2010; Nathanson et al. 2013a, b; Zullig et al. 2011). In this paper, correlations between dimensions were moderate (leadership and accountability, r = .57; leadership and safety/respect = 0.51; accountability and safety/respect = 0.48), providing some empirical basis to examine dimensions separately in analytic models. Individual dimensions are explored in more detail below.

Leadership was measured using aggregated teacher reports of the quality of instructional leadership provided by principals and other administrative staff at their school. Using a four-point scale (1 = strongly disagree, 4 = strongly agree), five items measured the extent to which teachers reported that their principal respects teaching and learning standards, communicates a clear vision for the school, and tracks academic progress. Higher levels of principal instructional leadership indicate that teachers trust and respect their principal, and view her as very involved in classroom instruction. Because school surveys are used for public reporting purposes, four point scales were subsequently reweighted on a ten-point scale (see more information in

Table 1 Teacher survey questions included in school climate constructs

Construct	Question List: *How much do you agree with the following statements?*
Leadership	School leaders communicate a clear vision for this school
Reliability = 0.94	School leaders let staff know what is expected of them
	School leaders encourage open communication on important school issues
	School leaders give me regular and helpful feedback about my teaching
	School leaders visit classrooms to observe the quality of teaching at this school
Accountability	My school has high expectations for all students
Reliability = 0.93	Teachers in this school set high standards for student work in their classes
	My school has clear measures of progress for student achievement throughout the year
	Teachers in this school use student achievement data to improve instructional decisions
Safety and respect	Order and discipline are maintained at my school
Reliability = 0.90	I can get the help we need at my school to address student behavior and discipline problems
	I am safe at my school
	Most students at my school treat teachers with respect
	Most parents teach teachers at this school with respect

See all school surveys from 2006 to 2014 at http://schools.nyc.gov/Accountability/tools/survey/default.htm

Nathanson et al. 2013a, b). NYC DOE officials were then able to more transparently assign schools a number of total school climate "points" that were included in an overall total possible score of 100. Thus, possible scores reported in this study range from 1 to 10 where 1 is a low score and 10 is a high score. The mean of the five items was taken to calculate an overall leadership score. Past analyses revealed high to moderate levels of reliability and internal consistency for the construct (Bryk et al. 2010). Analyses by the Consortium for Chicago School Research and the Research Alliance for NYC Schools have shown validity evidence for the measure (Byrk et al. 2010; Nathanson et al. 2013a, b). In this paper, reliability for the construct ranged from 0.92 to 0.94.

Accountability was measured by aggregating teacher perceptions on the extent to which the school had high standards for student work (Nathanson et al. 2013a, b). Using a four-point scale (1 = strongly disagree, 4 = strongly agree), four items were used to ask teachers to identify whether their school measures student progress, focuses on improving student performance, and has goals for student academic progress. As described above, individual items were reweighted on a 1–10 point scale. The mean of the four items was taken to calculate an overall accountability score and then aggregated across all teacher reports in the school. Developed by Childress et al. (2011), the construct has shown evidence of reliability and concurrent and predictive validity (Bryk et al. 2010; Nathanson et al. 2013a, b). In this paper, reliability for the construct ranged from 0.91 to 0.93 across the three study years.

Safety and respect measured the extent to which teachers felt that their school provided students and themselves with physical and emotional safety. Using a four-point scale (1 = strongly disagree, 4 = strongly agree), five items assessed teachers' perceptions of how order and discipline were maintained, whether supports were provided for behavior and discipline problems, and whether students and parents had respect for teachers at their school. As described above, individual items were rescaled on a 1–10 point scale. The mean of the five items was used to calculate an overall safety and respect score. Scores were then aggregated across all teachers to calculate a school score. This measure has evidence of reliability for the 2008–2010 time period, when the data for this paper were collected (Nathanson et al. 2013a, b; Rockoff and Speroni 2008). Rockoff and Speroni (2008) also showed concurrent validity for the school safety measure, linking it to overall rates of student suspensions and the NYCDOE's quality review report. Reliability for the measure ranged from 0.89 to 0.90 across the three study years.

Covariates

We controlled for student and school-level characteristics in order to improve the precision of moderated impact estimates (Bloom et al. 2007).

School Demographic Characteristics School poverty was measured as the percent of students in the school who were eligible for free/reduced price lunch. School racial/ethnic composition was assessed as the percentage of black students at the school and the percentage of Hispanic students at the school. Average daily attendance was also included as a covariate.

Child Demographic Characteristics Parent-reported child-level characteristics included ethnicity (dummy coded for child black and child Hispanic; white is the referent), gender (female = 1; male = 0), and child free/reduced price lunch eligibility (eligible = 1, not eligible = 0).

Child temperament was measured with the School-Age Temperament Inventory (SATI; McClowry 2002). Child temperament will be used as a covariate in this paper. The SATI is a 38 item 5-point Likert-type scale (1 = *never*, 3 = *half of the time*, 5 = *always*) that was standardized with a racially/ethnically and socioeconomically diverse sample of 883 parents reporting on their children. The instrument has four dimensions derived from principal factor analysis: negative reactivity (12 items; intensity with which the child expresses negative affect), task persistence (11 items; degree of self-direction a child exhibits in fulfilling responsibilities), withdrawal (9 items; child's initial response to new people/situations), and activity (6 items; large motor activity) (McClowry 2002). Cronbach's α's for the SATI (completed at study enrollment) were activity: α = 0.77; withdrawal: α = 0.81; task persistence: α = 0.85; negative reactivity: α = 0.87.

Procedure

Twenty-three elementary schools (recruited from schools with >80 % of students eligible for free/reduced price lunch) made a two-year commitment to participate in the study. Prior to randomization, one school withdrew from the study during a principal transition. Teachers were recruited in small group or individual meetings. Each cohort began with recruitment of the kindergarten teachers in September. First grade teachers were recruited from the same schools at the beginning of the following year. In all, 96 % of the kindergarten and first grade teachers consented to participate; there was no teacher attrition across time. All schools maintained the same principal throughout the duration of the study. Teachers completed the reports on students and received $50 in gift cards for supplies to thank them for their time.

Parents from the participating kindergarten teachers' classrooms were recruited in September and October. Recruitment of parents took place at school and over the phone. Parents reported on demographic characteristics and child temperament as part of a larger questionnaire, and received a $20 gift card to thank them for their time. After a parent consented, child assent was acquired. Due to resource limitations and concerns about teacher burden, recruitment at each school stopped after at least four students in each classroom enrolled in the study. However, the number of students in each class who enrolled ranged from four to ten. Based on Chi square tests, there were no significant differences between children enrolled in the study

and the school as a whole in terms of gender, race/ethnicity, and free/reduced price lunch eligibility.

Trained data collectors, blind to study condition and procedures, conducted individual child assessments with all children participating in the study at each of the five data time points. Data collectors were trained by an outside consultant on the Woodcock-Johnson and the Leiter-R during a 1-day training session each year. A graduate assistant conducted a mock assessment in the lab and observed all data collectors in the field before actual data collection began.

Random Assignment

Schools were used as the unit of random assignment to limit possible contamination effects which could threaten the internal validity of the study (Shadish et al. 2002). After baseline data were collected in kindergarten, a random numbers table was used to randomly assign schools to *INSIGHTS* or a supplemental reading program, referred to hereafter as the attention-control group. Eleven schools were randomized to *INSIGHTS*; the remaining eleven schools were assigned to the attention-control condition. Half of the children were in the *INSIGHTS* program ($N = 225$); the remaining child participants ($N = 210$) were in the attention-control. Similarly, approximately half of teachers ($N = 57$) participated in the *INSIGHTS* program; the remaining teachers ($N = 63$) were in the attention-control group.

Independent samples t-tests showed that children in *INSIGHTS* evidenced lower overall scores on reading achievement than their peers in the attention-control at baseline ($t(433) = 3.12$, $p < .01$). Chi square analyses also revealed there were more Hispanic children enrolled in *INSIGHTS* schools, relative to attention-control schools. Statistical modeling will adjust for these pretreatment differences. There were no pretreatment differences between the *INSIGHTS* and attention-control schools in terms of school climate, the moderators for this paper.

Intervention Procedures

Teachers and parents in schools assigned to *INSIGHTS* attended 10 weekly 2-h facilitated parallel sessions based on a structured curriculum that included didactic content and professionally produced vignettes as well as handouts and group activities. Teachers and parents were given assignments to apply the program content between sessions. Parents received $20 and teachers received professional development credit and $40 gift cards for each session attended.

During the same 10 weeks, the classroom program was delivered in 45-min lessons to all students in the classrooms

of participating teachers. Curriculum materials included puppets, workbooks, flash cards, and videotaped vignettes. Teachers were engaged in the child sessions, especially when students practiced resolving dilemmas. No make-up sessions were conducted, although teachers were asked to use the program materials with students who missed a session.

Facilitator Training

Facilitators were screened for skills/experiences prior to training. The eight facilitators had diverse racial/ethnic backgrounds and were graduate students in Psychology, Education, and Educational Theater. Facilitators attended a semester-long course to learn the theory and research underlying the program prior to training. New facilitators were then trained by the program developer and experienced staff to use the intervention materials. Each facilitator conducted the full intervention (teacher, parent, and child/classroom) in the schools to which s/he was assigned.

Intervention Fidelity

Facilitators followed scripts, used material checklists, documented sessions, and received ongoing training and supervision. Deviations or clinical concerns were discussed weekly in meetings with the program developer. Parent and teacher sessions were videotaped and reviewed for content and facilitation effectiveness. Fidelity coding was conducted by an experienced clinician who assessed that 94 % of the curriculum was covered in the teacher sessions and 92 % of the curriculum was covered in the parent sessions.

INSIGHTS Dosage

The average number of teacher sessions attended was 9.44 ($SD = 0.91$). The majority of teachers attended all sessions (70.6 %), and another 26.5 % attended eight or nine sessions. The average number of classroom sessions attended by the participating children was 8.30 ($SD = 2.25$). Thirty-two percent of children were present for all classrooms sessions and 46.3 % were present for eight or nine sessions. Participation in the program varied little across schools for teachers and students. The average number of parent sessions was 5.93 (SD = 4.15). There was variation in parent participation across schools ranging from 23 % of parents attending more than 80 % of sessions to 66 % attending less than 80 % of 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-h workshops in which reading coaches provided reading materials and presented strategies to enhance early literacy skills. Parents received $20 and teachers received $40 for classroom resources for each workshop. Twenty-four percent of children who were enrolled in the supplemental reading program participated in the full 10 sessions; an additional 19 % took part in eight or nine sessions. Thirty percent of parents and 83 % of teachers attended both sessions. Reading program facilitators had weekly meetings with the project manager to ensure that all components of the program were being implemented.

Analytic Approach

Missing Data Analysis

There were no missing school-level data. However, for the child-level variables, there was 0–20 % 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, child's gender, child age, child free-lunch eligibility, child behavior problems, child sustained attention, child math achievement, child reading achievement, parent gender, parent age, parent ethnicity, parent education, parent marital status, and parent work status. Analyses revealed that there were no substantial differences in rates of missingness between students by treatment status or achievement outcomes. However, students with more behavior problems, and those who had lower levels of parental education and parents who were not married were most likely to be missing outcome data. Missingness was thus dependent on several demographic characteristics (Little and Rubin 2002).

As such, the data were assumed to be Missing at Random (MAR), and a multiple imputation method (MI) was employed. Twenty separate datasets were imputed by chained equations, using STATA MICE in STATA version 12 (Little and Rubin 2002). Multiple imputation replaces missing values with predictions based on all the other information observed in the study. Multiple imputation accounts for uncertainty about missing data by imputing several values for each missing value, generating multiple datasets. In this paper, STATA ran each set of analyses 20 times and aggregated the findings across the imputed datasets.

Descriptive Statistics

Descriptive statistics on school and child level variables were calculated to describe individual and school level characteristics, assess the extent to which random assignment had accurately created similar treatment and control

groups, and determine variation in moderators and outcomes across schools and treatment conditions.

Growth Curve Modeling

In terms of examining moderated impacts for data, repeated measures (Level 1) were nested in children (Level 2), who were nested in schools (Level 3). As such, three-level individual growth modeling was used to examine change over four waves of data for each outcome (Singer and Willett 2003). All models were fitted with STATA 12. Maximum likelihood estimation was employed in all models. The unconditional means model for this analysis is as follows:

$$\text{Outcome}_{tij} = \gamma_{000} + u_{0ij} + \upsilon_{0j} + \varepsilon_{tij} \qquad (1)$$

In the unconditional model, the subscript t refers to repeated measures collected from child i (level-2 units) over time t in school j. The outcome scores for student i at time t are modeled as a function of: (a) a grand mean outcome score for all children (γ_{000}), (b) deviations in an individual's outcome mean around the grand mean (u_{0ij}),(c) deviations in the school level of the outcome (υ_{0j}), and (d) a time-specific residual term (ε_{tij}).

Unconditional means models were run for each of the child outcomes to determine whether there was significant between-individual and between-school variation in these predictors. Then, intraclass (ICC) correlations were computed. ICCs indicated that 22.07 % of the variation in math achievement, 30.88 % of the variation in reading achievement, 22.66 % of the variation in sustained attention, and 54.30 % of the variation in behavior problems, occurred across students. ICCs at the school level indicated some clustering for the outcomes at the school level (math achievement = 3.12 %; reading achievement = 7.34 %; sustained attention = 5.14 %; behavior problems = 11.98 %). Although school-level clustering was small, we included a random intercept at Level 3 for two reasons: (a) we aimed to estimate coefficients at Level 3 and (b) the study was randomized at Level 3 (Singer and Willett 2003).

Next, a growth model was fitted to examine children's outcome scores across time, regardless of treatment status. Time was centered at the last assessment so that the intercept would represent the average level of the outcome at the final intervention follow-up point (T5).

We then examined different model iterations to ascertain whether it was important to allow slopes to vary randomly across students, and to determine if it was necessary to allow the random intercept and random slope to covary. Results of the likelihood ratio tests comparing nested models indicated that the following model fit was best across the four outcomes:

$$\begin{aligned}
\text{Outcome}_{tij} = {} & \gamma_{000} + \gamma_{100}(\textit{Assessmentpo}\text{int}-4)_{tij} \\
& + \gamma_{010}\textit{T1SustainedAttention}_{ij} \\
& + \gamma_{020}\textit{T1BehaviorProb}_{ij} \\
& + \gamma_{030}\textit{T1ReadingAchieve}_{ij} \\
& + \gamma_{040}\textit{T1MathAchieve}_{ij} + u_{0ij} + u_{tij} + \upsilon_{0j} + \varepsilon_{tij}
\end{aligned}$$
$$(2)$$

As illustrated, this final model includes a random student-level intercept, a random school-level intercept, a random slope for students' outcome scores (u_{tij}), and the provision that the student level random intercept (u_{0ij}) and slope (u_{tij}) were permitted to covary (Corr(u_{0ij}, u_{tij}) = $\rho_{u_0 u_t}$). We also considered whether it was an appropriate to fit quadratic and cubic trends. Non-linear effects, however, were not statistically significant in any models.

Next, the predictor for treatment condition was added along with these student-level covariates: (a) child female, (b) child black, (c) child Hispanic, (d) negative reactivity (continuous: 1–5), (e) task persistence (continuous: 1–5), (f) activity (continuous: 1–5), (g) withdrawal (continuous: 1–5). As already noted, T1 levels of the outcomes were also included as Level 2 covariates. School level variables were entered at Level 3 and included three dimensions of school climate (leadership, accountability, and safety/respect) and several covariates—% black, % Hispanic, daily average attendance (%), and size (number of students). In order to accurately estimate the effect of Level 2 and 3 predictors on the Level 1 outcomes, all continuous predictors at Level 2 and 3 were centered around their grand mean. Categorical variables were not centered, as they were coded dichotomously and were time-invariant. The main effects analysis, illustrated below, included predictors for Treatment (γ_{050}), a number of predictors for the covariates described above (γ_{060} and γ_{004}), as well as the interaction between Treatment and time (γ_{150}).

$$\begin{aligned}
\text{Outcome}_{tij} = {} & \gamma_{000} + \gamma_{001}\textit{Leadership}_{j} + \gamma_{002}\textit{Accountability}_{j} \\
& + \gamma_{003}\textit{Safety/Respect}_{j} + \gamma_{004}\textit{SchoolCovars}_{j} \\
& + \gamma_{010}\textit{T1SustainedAttention}_{ij} \\
& + \gamma_{020}\textit{T1BehaviorProb}_{ij} \\
& + \gamma_{030}\textit{T1ReadingAchieve}_{ij} \\
& + \gamma_{040}\textit{T1MathAchieve}_{ij} \\
& + \gamma_{050}\textit{Treatment}_{ij} + \gamma_{060}\textit{StudentCovars}_{ij} \\
& + \gamma_{100}(\textit{Assessmentpoint}-4)_{tij} \\
& + \gamma_{150}\textit{Treatment}_{ij} * (\textit{Assessmentpoint}-4)_{tij} \\
& + u_{0ij} + u_{tij} + \upsilon_{0j} + \varepsilon_{tij} \qquad (3)
\end{aligned}$$

Moderation Analysis

Cross-level interactions between treatment and the moderators of interest (leadership, accountability, safety/respect) were then added. The model below shows the analysis for leadership. This model was then repeated for accountability and safety/respect:

$$
\begin{aligned}
Y_{tij} = {} & \gamma_{000} + \gamma_{001}Leadership_j + \gamma_{002}Accountability_j \\
& + \gamma_{003}Safety/Respect_j + \gamma_{004}SchoolCovars_j \\
& + \gamma_{010}T1SustainedAttention_{ij} + \gamma_{020}T1BehaviorProb_{ij} \\
& + \gamma_{030}T1ReadingAchieve_{ij} \\
& + \gamma_{040}T1MathAchieve_{ij} + \gamma_{050}Treatment_{ij} \\
& + \gamma_{060}StudentCovars_{ij} + \gamma_{100}(Assessmentpoint-4)_{tij} \\
& + \gamma_{150}Treatment_{ij} * (Assessmentpoint-4)_{tij} \\
& + \gamma_{051}Leadership_j * (Treatment)_{ij} + u_{0ij} + u_{tij} + \upsilon_{0j} + \varepsilon_{tij}
\end{aligned}
\tag{4}
$$

Significant effects on the coefficient γ_{051} would demonstrate that the impact of *INSIGHTS* on student outcomes at the final time point varied depending on the pretreatment level of leadership.

The final models examined whether growth in the outcomes varied by pretreatment levels of leadership, accountability, and safety/respect. Thus, three-way interactions between the school climate dimension of interest, Treatment, and Time were added to the previous set of models, as illustrated for the leadership dimension below.

$$
\begin{aligned}
Outcome_{tij} = {} & \gamma_{000} + \gamma_{001}Leadership_j + \gamma_{002}Accountability_j \\
& + \gamma_{003}Safety/Respect_j + \gamma_{004}SchoolCovars_j \\
& + \gamma_{010}T1SustainedAttention_{ij} \\
& + \gamma_{020}T1BehaviorProb_{ij} \\
& + \gamma_{030}T1ReadingAchieve_{ij} \\
& + \gamma_{040}T1MathAchieve_{ij} + \gamma_{050}Treatment_{ij} \\
& + \gamma_{060}StudentCovars_{ij}\gamma_{100}(Assessmentpoint-4)_{tij} \\
& + \gamma_{150}Treatment_{ij} * (Assessmentpoint-4)_{tij} \\
& + \gamma_{051}Leadership_j * Treatment_{ij} \\
& + \gamma_{151}Leadership_j * Treatment_{ij} \\
& * (Assessmentpoint-4)_{tij} + u_{0ij} + u_{tij} + \upsilon_{0j} + \varepsilon_{tij}
\end{aligned}
\tag{5}
$$

In the simpler model, significant Treatment × Time interactions indicate that growth in the outcomes varies between the treatment and control group. In extending this reasoning to the moderated effects by school climate dimensions, a significant Treatment × Time × School Climate Dimension effect (i.e., coefficient γ_{151}) would demonstrate that dimensions of school climate differentiated growth in student outcomes experienced by students

enrolled in *INSIGHTS*, relative to students in the attention-control condition.

For statistically significant moderated impacts, effect sizes were calculated using procedures developed by Feingold (2009) for growth modeling. Because main impacts of *INSIGHTS* reported in O'Connor et al. (2014) were on growth in the outcomes, we expected moderated impacts in this paper to be most evident in the parameters for Time × Treatment × School Climate Dimension. After running analyses, we used a series of Wald tests to determine whether coefficients were significantly different from one another across models.

Results

Descriptive Findings

Descriptive statistics are presented in Table 2. In general, child scores on sustained attention, math achievement, reading achievement, and behavioral problems increased over time. Dimensions of school climate assessed at baseline varied across schools (Leadership T × $M = 6.60$, $SD = 1.12$; Leadership Control $M = 6.78$, $SD = 1.14$; Accountability T × $M = 7.58$, $SD = 0.83$; Accountability Control $M = 7.58$, $SD = 0.91$; Safety/respect T × $M = 5.85$, $SD = 0.69$, Safety/respect Control $M = 6.04$, $SD = 0.75$). Whereas teachers generally had high perceptions of school accountability and moderate perceptions of leadership across treatment and control, their assessments of safety/respect were lower. The differences in the means between accountability, leadership and safety/respect were statistically significant ($F(2, 19) = 5.31$, $p < .01$). Importantly, variation in dimensions of school climate was similar when comparing *INSIGHTS* schools with attention-control schools.

Leadership Moderated Impacts

Results from the leadership model (see Table 3) revealed significant Treatment × Time × Leadership effects on math ($\gamma = -0.26$, $SE = 0.11$, $p = .03$; E.S. = 0.16) and reading achievement ($\gamma = -0.49$, $SE = 0.19$, $p = .03$; E.S. = 0.21). Growth in math and reading achievement was faster for students enrolled in *INSIGHTS* attending schools with lower baseline levels of leadership, relative to students enrolled in *INSIGHTS* attending schools with higher levels of leadership. Models examining moderated effects on growth in sustained attention and behavior problems were not statistically significant. Effects for Treatment × Leadership were not statistically significant for any of the outcomes.

Table 2 Descriptive statistics across study time points for school and child level variables of interest

Variable	Pre-test (T1)				Post-test (T2)				Post-test (T3)			
	Treatment		Control		Treatment		Control		Treatment		Control	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Child-level variables												
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
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
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
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
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	–	–	–	–	–	–	–	–	–
Elig. free/red. lunch	0.84	–	0.83	–	–	–	–	–	–	–	–	–
School-level variables					–	–	–	–	–	–	–	–
% Eligible free/red. lunch	0.91	0.22	0.87	0.23	–	–	–	–	–	–	–	–
% Black	0.83	0.12	0.75	0.16	–	–	–	–	–	–	–	–
% Hispanic	0.49	0.27	0.39	0.27	–	–	–	–	–	–	–	–
Average attendance	0.89	0.14	0.86	0.20	–	–	–	–	–	–	–	–
School size	488.70	133.70	506.23	189.66	–	–	–	–	–	–	–	–
Leadership	6.60	1.12	6.78	1.14	–	–	–	–	–	–	–	–
Accountability	7.58	0.83	7.58	0.91	–	–	–	–	–	–	–	–
Safety and respect	5.85	0.69	6.04	0.75	–	–	–	–	–	–	–	–

Variable	Post-test (T4)				Post-test (T5)			
	Treatment		Control		Treatment		Control	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Child-level variables								
Math achievement	22.37	4.05	22.45	3.54	23.34	4.62	23.17	3.86
Reading achievement	31.15	8.75	32.86	8.46	33.54	9.07	33.57	8.11
Sustained attention	60.37	8.21	59.39	8.73	61.44	9.03	60.54	9.54
Behavior problems	2.29	1.21	2.35	1.27	2.28	1.36	2.46	1.42
Negative reactivity	–	–	–	–	–	–	–	–
Task persistence	–	–	–	–	–	–	–	–
Activity	–	–	–	–	–	–	–	–
Withdrawal	–	–	–	–	–	–	–	–
Child female	–	–	–	–	–	–	–	–
Child black	–	–	–	–	–	–	–	–
Child hispanic	–	–	–	–	–	–	–	–
Elig. free/red. lunch	–	–	–	–	–	–	–	–
School-level variables	–	–	–	–	–	–	–	–
% Eligible free/red. lunch	–	–	–	–	–	–	–	–
% Black	–	–	–	–	–	–	–	–
% Hispanic	–	–	–	–	–	–	–	–
Average attendance	–	–	–	–	–	–	–	–
School size	–	–	–	–	–	–	–	–

Table 2 continued

Variable	Post-test (T4)				Post-test (T5)			
	Treatment		Control		Treatment		Control	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Leadership	–	–	–	–	–	–	–	–
Accountability	–	–	–	–	–	–	–	–
Safety and respect	–	–	–	–	–	–	–	–

Table 3 Model summary for individual growth models examining sustained attention, behavior problems, math achievement, and reading achievement, moderated by school leadership

Fixed effects	Math achievement		Reading achievement		Sustained attention		Behavior problems	
	γ	SE	γ	SE	γ	SE	γ	SE
Intercept	23.57**	0.64	32.60**	1.37	58.49**	1.57	2.48**	0.18
Child female	−0.34	0.31	−0.45	0.64	1.27†	0.68	−0.21*	0.09
Child black	−0.53	0.47	0.99	0.98	0.33	1.04	0.11	0.13
Child hispanic	−0.82†	0.48	0.18	1.01	0.20	1.08	−0.21	0.13
Elig. free/reduced lunch	−1.09*	0.49	−2.22*	1.03	−0.26	1.01	0.05	0.13
Math achievement, T1	0.26**	0.03	0.23**	0.07	0.09	0.08	−0.01	0.01
Reading achievement, T1	0.13**	0.02	0.49**	0.04	0.15**	0.04	0.01	0.01
Behavior problems, T1	−0.20	0.13	−0.59*	0.27	−0.76**	0.29	0.49**	0.03
Sustained attention, T1	0.03*	0.01	0.04†	0.02	0.17**	0.03	0.01†	0.01
Task persistent	−0.13	0.24	0.34	0.50	0.25	0.54	−0.03	0.07
Negative reactivity	−0.59*	0.25	−0.42	0.51	0.26	0.55	0.13†	0.07
Withdrawn	0.32	0.21	−0.62	0.43	−0.94*	0.46	0.01	0.06
Activity	0.32	0.20	0.93*	0.42	0.14	0.44	−0.04	0.06
School % black	−0.49	1.91	−1.74	4.31	−7.95	5.92	−0.04	0.56
School % hispanic	−1.04	0.72	−2.42	1.67	3.98	2.46	−0.03	0.21
School average attendance	0.44	1.43	2.49	3.15	5.21	4.19	−0.15	0.41
School size	−0.01	0.01	−0.01	0.01	−0.01	0.01	0.01	0.01
School leadership	0.21	0.40	0.91	0.91	0.23	1.25	−0.11	0.12
Treatment	−0.02	0.33	1.65*	0.75	1.65†	1.00	−0.19*	0.08
Treatment × leadership	−0.42	0.45	−1.65	1.04	0.25	1.43	0.17	0.14
Time	2.10**	0.09	3.86**	0.16	3.13**	0.19	0.01	0.02
Time × treatment × leadership	−0.26*	0.11	−0.49*	0.19	−0.06	0.24	−0.01	0.03
Random effects	Variance	SE	Variance	SE	Variance	SE	Variance	SE
Child intercept	4.75**	0.61	31.19**	3.34	22.11**	2.91	0.57**	0.06
Child slope	0.61**	0.23	1.11**	0.39	4.61**	1.19	0.07**	0.12
Corr., slope and intercept	0.03	0.26	5.88**	1.20	−2.21	1.40	0.12**	0.02
School intercept	0.01	0.01	0.39	0.74	2.53	1.37	0.01	0.01
Residual variation	10.69**	0.56	40.24**	1.72	49.53**	2.58	0.64	0.03

** $p < .01$; * $p < .05$; † $p < .1$

Accountability Moderated Impacts

Results from the accountability model (see Table 4) revealed significant Treatment × Time × Accountability effects on math ($\gamma = -0.42$, $SE = 0.14$, $p = .02$; E.S. = 0.25) and reading achievement ($\gamma = -0.70$, $SE = 0.26$, $p < .03$; E.S. = 0.29). Growth in math and reading achievement was faster for Treatment students

Table 4 Model summary for individual growth models examining sustained attention, behavior problems, math achievement, and reading achievement, moderated by accountability

Fixed effects	Math achievement		Reading achievement		Sustained attention		Behavior problems	
	γ	SE	γ	SE	γ	SE	γ	SE
Intercept	23.56**	0.63	32.58**	1.37	58.40**	1.56	2.41**	0.18
Child female	−0.32	0.30	−0.39	0.64	1.26†	0.68	−0.20*	0.08
Child black	−0.51	0.47	1.04	0.98	0.34	1.04	0.10	0.13
Child hispanic	0.81†	0.48	0.22	1.01	0.19	1.07	−0.22†	0.13
Elig. free/reduced lunch	−1.12*	0.49	−2.32*	1.02	−0.23	1.10	0.07	0.13
Math achievement, T1	0.26**	0.03	0.23**	0.07	0.09	0.08	−0.01	0.01
Reading achievement, T1	0.13**	0.02	0.49**	0.04	0.15**	0.15	0.01	0.01
Behavior problems, T1	−0.20	0.13	−0.65*	0.27	−0.76**	0.29	0.49**	0.04
Sustained attention, T1	0.03*	0.01	0.04†	0.03	0.17**	0.03	0.01†	0.01
Task persistent	−0.13	0.24	0.27	0.51	0.26	0.54	−0.04	0.07
Negative reactivity	−0.58*	0.25	−0.43	0.52	0.26	0.55	0.13†	0.07
Withdrawn	−0.31	0.21	−0.61	0.41	−0.95*	0.45	0.01	0.06
Activity	0.31	0.19	0.93*	0.42	0.14	0.44	−0.04	0.05
School % black	0.01	1.58	0.04	3.72	−6.81	5.14	−0.25	0.44
School % hispanic	−0.97	0.82	−1.12	1.98	3.23	2.78	0.24	0.23
School average attendance	0.87	1.20	0.71	2.07	−5.57	3.47	−0.17	0.33
School size	0.01	0.01	−0.01	0.01	−0.01	0.01	0.01	0.01
School accountability	0.01	0.33	−0.20	0.77	0.32	1.03	−0.23**	0.09
Treatment	0.01	0.34	1.70*	0.79	1.65	1.07	−0.18*	0.09
Treatment × accountability	−0.37	0.48	−0.43	1.15	−0.03	1.51	0.32*	0.14
Time	2.11**	0.08	3.88**	0.16	3.14**	0.20	0.01	0.01
Time × treatment × accountability	−0.42**	0.14	−0.70**	0.26	−0.21	0.33	−0.01	0.01
Random effects	Variance	SE	Variance	SE	Variance	SE	Variance	SE
Child intercept	4.74**	0.61	31.36**	3.56	22.11**	2.91	0.57**	0.06
Child slope	0.59**	0.23	1.13**	0.33	4.59**	1.19	0.07**	0.02
Corr., slope and intercept	0.02	0.27	5.96**	1.18	−2.19	1.40	0.13**	0.02
School intercept	0.01	0.01	0.55	0.74	2.58†	1.38	0.01	0.01
Residual variation	10.70**	0.56	40.16**	1.64	49.54**	2.57	0.64**	0.03

** $p < .01$; * $p < .05$; † $p < .1$

enrolled in schools with lower baseline levels of accountability, relative to Treatment students in schools with higher accountability. There was a significant Treatment × Accountability effect for behavior problems ($\gamma = 0.32$, $SE = 0.14$, $p = .04$; E.S. = 0.27). Treatment students in low accountability schools evidenced fewer behavior problems than control group members in low accountability schools (see Fig. 1). Coefficients for Treatment × Time × Accountability predicting sustained attention and behavior problems were not statistically significant. The effect for Treatment × Accountability was similarly non-significant.

Safety and Respect Moderated Impacts

Results from the safety/respect model (see Table 5) revealed significant Treatment × Time × Safety/respect effects on math ($\gamma = -0.49$, $SE = 0.16$, $p = .02$; E.S. = 0.29) and reading achievement ($\gamma = -0.86$, $SE = 0.29$, $p < .01$; E.S. = 0.36) as well as sustained attention ($\gamma = -1.33$, $SE = 0.36$, $p < .01$; E.S. = 0.31). Growth in treatment students' math and reading achievement and sustained attention was faster in schools with lower baseline levels of safety/respect. In addition, there was a significant Treatment × Safety/respect effect for
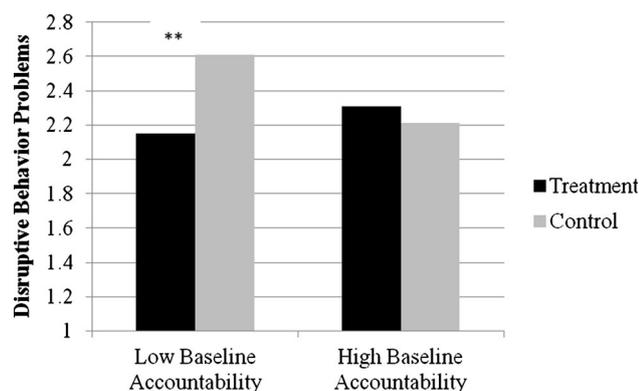
Fig. 1 INSIGHTS impacts on behavior problems, moderated by accountability

reading achievement ($\gamma = -3.24$, $SE = 1.46$, $p = .03$; E.S. $= 0.45$). As illustrated in Fig. 2, this finding demonstrates that reading achievement at the final time point was larger for treatment students in low safety-respect schools, relative to control students in low-safety respect schools. The coefficients for Treatment $\times$ Time $\times$ Safety/Respect and Treatment $\times$ Safety/Respect were both non-significant in the model predicting behavior problems. Treatment $\times$ Safety/Respect was non-significant in the models predicting math achievement and sustained attention.

Model Comparisons

Wald tests revealed that the moderated slope effects on math achievement were not statistically different in the models examining Accountability and Safety/Respect ($\chi^2(2) = 2.11$, $p = .27$). All other statistically significant moderated effects were significantly different from one another.

Discussion

Results of this paper advance theory and research on the role of school settings in understanding SEL program impact heterogeneity on outcomes for low-income urban elementary school students. Students in schools with low levels of leadership, accountability, or safety/respect in the year prior to the study appeared to benefit most from the *INSIGHTS* intervention in kindergarten and first grade. These dimensions of school climate moderated *INSIGHTS'* impacts on social-emotional, behavioral, and academic outcomes, although there was variation in these patterns across dimensions. Effect sizes were consistent with average impacts on academic performance identified in Durlak et al.'s (2011) meta-analysis of SEL programs (average E.S. $= 0.27$).

Moderated Program Impacts

INSIGHTS' program impacts on student math and reading achievement were larger for schools with lower baseline levels of leadership, accountability, and safety/respect. Results can be compared to the contextual moderation analysis of the school-based prevention program the Good Behavior Game (GBG), a behavioral intervention. In a trial of the program implemented in low-income urban elementary schools, researchers identified the strongest program impacts for boys enrolled in classrooms with higher levels of overall aggression (Kellam et al. 2008). Taken together, evidence indicates that school-based programs implemented in settings that have some contextual risk are more likely to improve the individual-level outcomes of the students embedded within them. In the current study, it may be that in schools where social processes are less positive and supportive on average, an SEL program can provide a compensatory structure that engenders a setting more conducive to improving students' math and reading achievement (Bierman et al. 2014; Liew 2012; Reynolds et al. 2011). For example, schools that have less positive school climates may exhibit lower quality interactions between students, teachers, peers, and staff (Lee 2012; Thapa et al. 2013). In addition, in schools with lower levels of school climate, students are more likely to have low levels of behavioral and emotional support (Bradshaw et al. 2012). Such schools may be in more need of a program that aims to improve the emotional support and organization of school and classroom contexts, and students' individual social-emotional and behavioral skills.

Interestingly, moderated program effects for accountability and safety/respect were larger than they were for leadership. In explaining the difference in magnitude of this moderated effect, it is important to remember that SEL programs are typically delivered to students in classrooms either by using an outside facilitator or training teachers to deliver curricula (McCormick et al. 2015; Durlak et al. 2011). Although there are exceptions (e.g., Roderick 2013), *INSIGHTS* and many other SEL programs using a classroom curriculum *do not* directly integrate principals and administrators into programming. There are thus fewer opportunities for SEL programs to provide compensatory inputs for the types of activities included in school leadership.

In contrast, the *INSIGHTS* program does directly target some of the key interactions between teachers and students that may be of lower quality in schools with less safety/respect (Bryk et al. 2010; Cohen et al. 2009; Nathanson et al. 2013a, b). One specific goal of SEL programs in general is to enhance student–teacher and peer relationship quality and increase levels of respect and emotional support conferred to individual students (CASEL 2014; Durlak

Table 5 Model summary for individual growth models examining sustained attention, behavior problems, math achievement, and reading achievement, moderated by safety and respect

| Fixed effects	Math achievement		Reading achievement		Sustained attention		Behavior problems	
	γ	SE	γ	SE	γ	SE	γ	SE
Between-child estimates								
Intercept	22.79**	0.51	30.10**	1.17	57.86**	1.27	2.46**	0.15
Child female	−0.35	0.30	−0.34	0.64	1.32*	0.67	−0.20*	0.08
Child black	−0.70	0.45	0.95	0.95	0.23	0.99	0.15	0.12
Child hispanic	−1.04*	0.46	−0.04	0.97	0.02	1.02	−0.16	0.13
Elig. free/reduced lunch	−0.04	0.03	−0.01	0.07	0.06	0.08	0.01	0.01
Math achievement, T1	0.28**	0.03	0.25**	0.07	0.11	0.07	−0.01	0.01
Reading achievement, T1	0.13**	0.02	0.49**	0.04	0.14**	0.05	0.01	0.01
Behavior problems, T1	−0.15	0.13	−0.43	0.27	−0.66*	0.28	0.50**	0.03
Sustained attention, T1	0.03*	0.01	−0.04†	0.04	0.18**	0.03	0.01	0.01
Task persistent	−0.13	0.24	0.39	0.51	0.33	0.54	−0.02	0.07
Negative reactivity	−0.57*	0.25	0.53	0.52	0.17	0.54	0.12†	0.07
Withdrawn	−0.34†	0.20	−0.65	0.42	−0.74†	0.44	0.01	0.05
Activity	0.29	0.20	0.92*	0.41	0.17	0.44	−0.03	0.05
% Black	−1.45	1.90	−5.98	4.77	−14.08*	5.62	−0.11	0.59
% Hispanic	−0.98	0.73	−2.89	1.84	2.50	2.17	−0.14	0.22
Average attendance	0.49	1.39	5.77†	3.34	−0.36	3.85	−0.10	0.42
School size	0.01	0.01	−0.01†	0.01	−0.01	0.01	−0.01	0.01
Safety and respect	0.51	0.47	2.59*	1.15	2.89*	1.35	−0.06	0.14
Treatment	−0.14	0.36	−1.27	0.90	2.77**	1.06	−0.17	0.11
Treatment × safety/respect	−0.78	0.59	−3.24*	1.46	−1.51	1.65	0.03	0.18
Time	1.98**	0.09	3.71**	0.16	2.88**	0.20	0.01	0.02
Time × treatment × safety/respect	−0.49**	0.16	−0.86**	0.29	−1.33**	0.36	0.01	0.04
Random effects	Variance	SE	Variance	SE	Variance	SE	Variance	SE
Intercept	4.83**	0.60	32.55**	3.37	22.14**	2.86	0.57**	0.06
Slope	0.57**	0.22	1.06**	0.36	3.85**	1.12	0.07**	0.02
Corr., slope and intercept	0.02	0.26	5.86**	1.17	−2.54*	1.32	0.13**	0.02
School intercept	0.01	0.01	0.74	0.80	1.61	1.07	0.01	0.01
Residual variation	10.62**	0.54	39.92**	1.67	49.99**	2.55	0.64**	0.03

** $p < .01$; * $p < .05$; † $p < .1$

et al. 2011). In schools with higher levels of safety/respect prior to program implementation, however, members of the school may already be engaged in supportive, respectful relationships (Higgins et al. 2012). An SEL program is thus less likely to have a compensatory effect in a context where high-quality social processes are already in place.

Similarly, in schools that began the intervention with the lowest levels of accountability, teachers and staff may put less focus on specific structures and policies designed to support student achievement. In such schools, teachers may need even more support from principals than they are already receiving. Given established links between school accountability, or academic press, and aggregate levels of student achievement, students in those schools are likely to

exhibit lower overall achievement prior to the implementation of intervention (Meece et al. 2006). As such, in line with a prevention science approach, these students will have more room for academic improvement relative to students attending schools with a more direct focus on student achievement (Durlak et al. 2011; Greenberg et al. 2003). Although few studies have considered contextual risk moderators, the moderated impacts for accountability in this paper mirror prior work showing larger impacts for *individual* students who begin an intervention with lower academic and social-emotional skills (e.g., Bierman et al. 2010; Jones et al. 2011).

Results also revealed that accountability moderated the impact of *INSIGHTS* on behavior problems, and safety/
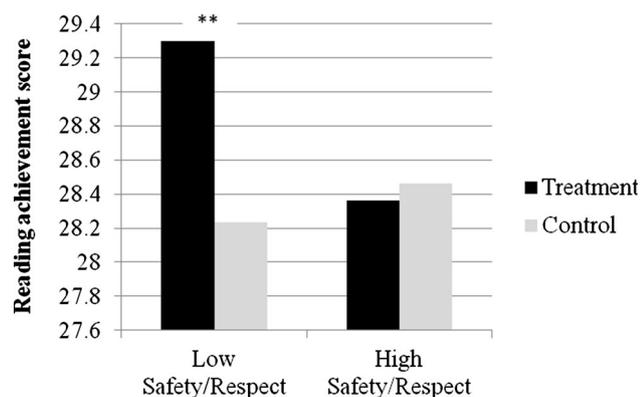
Fig. 2 INSIGHTS impacts on reading achievement, moderated by safety/respect

respect moderated the impact of *INSIGHTS* on sustained attention such that gains were larger in schools with lower baseline levels of these school climate dimensions. Findings are critical for informing targeted intervention. In settings that are less focused on directly improving academic standards, the culture may also be less supportive of students' behavioral engagement and regulation (Higgins et al. 2012). Thus, SEL programs' direct focus on improving behaviors, self-control, and engagement may be more likely to be compensatory in a setting with lower baseline accountability. Similarly, given that children are less likely to develop important attentional capacities in settings with lower levels of physical and emotional safety (see McCoy 2013 for a review), SEL programs may be most likely to improve the sustained attention of students in school settings with lower baseline levels of safety/respect.

The consistent findings in this study are notable given that they are directly in line with theories from prevention science (Cicchetti and Aber 1998). Yet, they run somewhat counter to previous work showing that the relational climate of the school setting needs to be supportive at the start in order for school-based programming to benefit students' outcomes (e.g., Hughes et al. 2005). Previous evaluations where this has been the case, however, typically used students' individual levels of aggression and negative behaviors to operationalize measures of contextual risk (see Aber et al. 1998; Hughes et al. 2005). While such an approach is an appropriate way to measure descriptive norms of aggression, it does not necessarily address the social processes at the setting level—or interactions between key members of the setting—that SEL programs like *INSIGHTS* are designed to target. In future work assessing SEL program moderation, it may be important to use a variety of methods to describe processes—both norms based on individual behaviors and perceived levels

of acceptability about individual behaviors as well as social interactions between individuals—at the setting level.

Limitations and Directions for Future Research

Even given the randomized design of this study, there are a number of limitations. First, the dimensions of school climate are moderately correlated and reported only by teachers. There is thus a concern that the individual dimensions might not uniquely explain differential impacts. Rather, the dimensions together may capture some global measure of school quality, indicating that lower quality schools stand to gain more in terms of student achievement, attention, and behaviors than higher quality schools. Notably, however, we did examine the moderated impact models using the "overall school quality" scale as a combination of the dimensions. For both math and reading outcomes, we found a trend-level interaction effect supporting the hypothesis that schools with lower levels of school quality have the most to gain from the intervention. However, the interaction effects were not statistically significant at the .05 level, and the effect size was smaller than the moderated impacts for leadership, accountability, and safety/respect.

A second limitation is that only teacher reports were used to measure school climate. Although research suggests that teacher reports of climate are the most reliable for distinguishing between-school differences (see Nathanson et al. 2013a, b), alternate perspectives, or observed measures, may be valid as well. School climate also included perspectives from K to fifth grade teachers, although the intervention only took place in K and first grade. The construct thus represents a broader understanding of climate than might be reported only by the teachers in the study. Relatedly, the analyses combined aggregated reports of climate with individual level outcomes. The aggregated assessments are from a different reporter than the individual-level outcomes. However, work by Raghunathan et al. (2003) suggests that estimates combining aggregated and individual data could be biased.

Next, power was limited at Level 3. As such, we were unable to operationalize treatment at the level where randomization occurred. An additional possible limitation is the fact that schools enrolled in this study did have moderate to high levels of climate on two of the assessed dimensions—accountability and leadership. Therefore, it may be important to consider the existence of a threshold effect, wherein a minimum level of supportive climate is needed in order to implement an SEL program that would be likely to impact student outcomes. The final limitation is that findings are generalizable only to urban public

elementary schools. Future work should consider larger and more varied samples of students and schools.

Implications for Research, Policy, and Practice

This paper is one of the first to consider the role of school climate in understanding moderated impacts of an SEL program on student achievement, sustained attention, and behavior problems. The major lesson from this work is that context matters. Across student outcomes, program impacts were generally larger, and sometimes driven by, schools that had lower levels of leadership, accountability and safety/respect prior to implementation of the intervention. Although there are nuanced reasons to explain heterogeneity of effects, future evaluators of SEL programs should build on this work to determine whether such moderated impacts are replicated across diverse implementation settings. Similar to Bierman et al. (2010), it may be important to consider varied cities and types of school settings, while explicitly collecting data on school climate to examine future impact variation. Moreover, it is likely important to consider both norms and social processes at the setting level (e.g., Aber et al. 1998; Hughes et al. 2005) when examining variation in SEL program impacts.

Community psychologists and practitioners interested in improving school-based programs can incorporate components that directly aim to improve social processes in schools. There are educators and practitioners who are faced with difficult choices related to intervention when challenged by underachievement in their schools. One possible response, however, should be to consider whether there are supportive social processes extant in the school. If social interactions are of low quality, a social-emotional learning program may be an appropriate intervention to implement. In schools with less positive contexts, significant impacts on achievement outcomes may be more likely.

Perhaps the biggest lesson from this study, however, is the need for policymakers to expand and implement SEL programs in a variety of settings across the country, especially in under-resourced schools. Importantly, policymakers are paying increased attention to the role of school climate in student learning (Weissbourd et al. 2013). Indeed, the US Department of Education, the Institute for Educational Sciences and President Obama's Bully Prevention Partnership endorse school climate renewal as a strategy for increasing student learning and achievement, and enhancing school connectedness. It may now be possible to combine efforts to implement SEL programs in high-need settings and target improvements in school climate. Findings from this paper suggest the importance of considering the overall climate and characteristics of the school before allocating resources to SEL programs.

Implementing such a strategy may actually be quite feasible in some larger urban areas where there are administrative surveys and outside quality reviews that provide information about school climate (Coburn et al. 2013).

References

Aber, J. L., Jones, S. M., Brown, J. L., Chaudry, N., & Samples, F. (1998). Resolving conflict creatively: Evaluating the developmental effects of a school-based violence prevention program in neighborhood and classroom context. *Development and Psychopathology, 10*(2), 187–213.

Bierman, K. L., Coie, J. D., Dodge, K. A., Greenberg, M. T., Lochman, J. E., McMahon, R. J., & Pinderhughes, E. (2010). The effects of a multiyear universal social–emotional learning program: The role of student and school characteristics. *Journal of Consulting and Clinical Psychology, 78*(2), 156.

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.

Bloom, H. S., Richburg-Hayes, L., & Black, A. R. (2007). Using covariates to improve precision for studies that randomize schools to evaluate educational interventions. *Educational Evaluation and Policy Analysis, 29*(1), 30–59.

Bradshaw, C. P., Waasdorp, T. E., & Leaf, P. J. (2012). Effects of school-wide positive behavioral interventions and supports on child behavior problems. *Pediatrics, 130*(5), e1136–e1145.

Brand, S., Felner, R. D., Seitsinger, A., Burns, A., & Bolton, N. (2008). A large scale study of the assessment of the social environment of middle and secondary schools: The validity and utility of teachers' ratings of school climate, cultural pluralism, and safety problems for understanding school effects and school improvement. *Journal of School Psychology, 46*(5), 507–535.

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

Bryk, A. S., Sebring, P. B., Allensworth, E., Luppescu, S., & Easton, J. (2010). *Organizing schools for improvement: Lessons from Chicago*. Chicago: University of Chicago Press.

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.

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

Childress, S., Higgins, M., Ishimaru, A., & Takahashi, S. (2011). Accountability and organizational learning in New York City reforms. In J. O'Day, C. Bitter, & L. Gomez (Eds.), *Education reform in New York City: Ambitious change in the nation's most complex school system*. Cambridge: Harvard Education Press.

Cicchetti, D., & Aber, J. L. (1998). Contextualism and developmental psychopathology. *Development and Psychopathology, 10*(02), 137–141.

Coburn, C., Penuel, W., & Geil, K. (2013). Research-practice partnerships: A strategy for leveraging research for educational improvement in school districts. In *Working paper*. New York: The William T. Grant Foundation.

Cohen, J., McCabe, E. M., Michelli, N. M., & Pickeral, T. (2009). School climate: Research, policy, practice, and teacher education. *Teachers College Record, 111*, 180–213.

Cooper, C. E., Osborne, C. A., Beck, A. N., & McLanahan, S. S. (2011). Partnership instability, school readiness, and gender disparities. *Sociology of Education, 84*(3), 246–259.

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.

Espelage, D. L., Low, S., Rao, M. A., Hong, J. S., & Little, T. D. (2014). Family violence, bullying, fighting, and substance use among adolescents: A longitudinal mediational model. *Journal of Research on Adolescence, 24*(2), 337–349.

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

Feingold, A. (2009). Effect sizes for growth-modeling analysis for controlled clinical trials in the same metric as for classical analysis. *Psychological Methods, 14*(1), 43.

Garrison, W. M. (2004). Profiles of classroom practices in US public schools. *School Effectiveness and School Improvement, 15*(3–4), 377–406.

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.

Higgins, M., Ishimaru, A., Holcombe, R., & Fowler, A. (2012). Examining organizational learning in schools: The role of psychological safety, experimentation, and leadership that reinforces learning. *Journal of Educational Change, 13*(1), 67–94.

Hughes, J. N., Cavell, T. A., Meehan, B. T., Zhang, D., & Collie, C. (2005). Adverse school context moderates the outcomes of selective interventions for aggressive children. *Journal of Consulting and Clinical Psychology, 73*(4), 731–746.

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.

Kahn, J. (2013). Can emotional intelligence be taught? New York Times. Retrieved July 2, 2014. http://www.nytimes.com/2013/09/15.

Kellam, S. G., Brown, C. H., Poduska, J. M., Ialongo, N. S., Wang, W., Toyinbo, P., & Wilcox, H. C. (2008). Effects of a universal classroom behavior management program in first and second grades on young adult behavioral, psychiatric, and social outcomes. *Drug and Alcohol Dependence, 95*, S5–S28.

Kumanyika, S. K., & Grier, S. (2006). Targeting interventions for ethnic minority and low-income populations. *The Future of Children, 16*(1), 187–207.

Lee, J. (2012). The effects of the teacher–student relationship and academic press on student engagement and academic performance. *International Journal of Educational Research, 53*, 330–340.

Liew, J. (2012). Effortful control, executive functions, and education: Bringing self-regulatory and social-emotional competencies to the table. *Child Development Perspectives, 6*(2), 105–111.

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

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

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

McClowry, S. G., Snow, D. L., & Tamis-LeMonda, C. S. (2005). An evaluation of the effects of INSIGHTS on the behavior of inner city primary school children. *Journal of Primary Prevention, 26*, 567–584.

McClowry, S. G., Snow, D. L., Tamis-LeMonda, C. S., & Rodriguez, E. T. (2010). Testing the efficacy of INSIGHTS on student disruptive behavior, classroom management, and student competence in inner city primary grades. *School Mental Health, 2*, 23–35.

McCormick, M. P., O'Connor, E. E., Cappella, E., & McClowry, S. G. (2015). Getting a good start in school: Effects of INSIGHTS on children with high maintenance temperaments. *Early Childhood Research Quarterly, 30*(A), 128–139.

McCoy, D. C. (2013). Early violence exposure and self-regulatory development: A bioecological systems perspective. *Human Development, 56*(4), 254–273.

McCoy, D. C., Roy, A. L., & Sirkman, G. M. (2013). Neighborhood crime and school climate as predictors of elementary school academic quality: A cross-lagged panel analysis. *American Journal of Community Psychology, 52*(1–2), 128–140.

Meece, J. L., Anderman, E. M., & Anderman, L. H. (2006). Classroom goal structure, student motivation, and academic achievement. *Annual Review of Psychology, 57*, 487–503.

Nathanson, L., Cole, R., Kemple, J. J., Lent, J., McCormick, M., & Segeritz, M. (2013a). New York City School Survey 2008–2010: Assessing the reliability and validity of a progress report measure. Technical Report. Research Alliance for New York City Schools.

Nathanson, L., McCormick, M., & Kemple, J. (2013b). Strengthening assessments of school climate: Lessons from the NYC school survey. New York: The Research Alliance for New York City Schools. Retrieved from http://files.eric.ed.gov/fulltext/ED543180.pdf.

National School Climate Council. (2007). The school climate challenge: Narrowing the gap between school climate research and school climate policy, practice guidelines and teacher education policy. Retrieved from http://www.schoolclimate.org/climate/advocacy.

New York City Department of Education. (2014). Learning environment survey data. Retrieved July 3, 2014 from http://schools.nyc.gov/Accountability/tools/survey/default.htm.

O'Connor, E. E., Cappella, E., McCormick, M. P., & McClowry, S. G. (2014). An examination of the efficacy of INSIGHTS in enhancing the academic learning context. *Journal of Educational Psychology, 106*(4), 1156–1169.

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.

Raghunathan, T. E., Diehr, P. K., & Cheadle, A. D. (2003). Combining aggregate and individual level data to estimate an individual level correlation coefficient. *Journal of Educational and Behavioral Statistics, 28*(1), 1–19.

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.

Reynolds, A. J., Temple, J. A., Ou, S. R., Arteaga, I. A., & White, B. A. (2011). School-based early childhood education and age-28 well-being: Effects by timing, dosage, and subgroups. *Science, 333*(6040), 360–364.

Rimm-Kaufman, S. E., Fan, X., Chiu, Y. J., & You, W. (2007). The contribution of the responsive classroom approach on children's academic achievement: Results from a three year longitudinal study. *Journal of School Psychology, 45*(4), 401–421.

Rockoff, J., & Speroni, C. (2008). Reliability, consistency, and validity of the NYC environmental surveys: A preliminary analysis. In *Working paper*. New York: Columbia University Graduate School of Business.

Roderick, T. (2013). Brief on the smart school leaders program. Retrieved July 3, 2014 from http://www.morningsidecenter.org/smart-school-leaders.

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

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

Shadish, W. R., Cook, T. D., & Campbell, Donald T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. New York: Houghton Mifflin.

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. Oxford: Oxford University Press.

Snyder, F., Flay, B., Vuchinich, S., Acock, A., Washburn, I., Beets, M., & Li, K. K. (2009). Impact of a social-emotional and character development program on school-level indicators of academic achievement, absenteeism, and disciplinary outcomes: A matched-pair, cluster-randomized, controlled trial. *Journal of Research on Educational Effectiveness, 3*(1), 26–55.

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.

Supplee, L. H., Kelly, B. C., MacKinnon, D. M., & Barofsky, M. Y. (2013). Introduction to the special issue: Subgroup analysis in prevention and intervention research. *Prevention Science, 14*, 107–110.

Thapa, A., Cohen, J., Guffey, S., & Higgins-D'Alessandro, A. (2013). A review of school climate research. *Review of Educational Research, 83*(3), 357–385.

Tough, P. (2012). *How children succeed*. New York: Random House.

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

Van Lier, P. A., Muthén, B. O., van der Sar, R. M., & Crijnen, A. A. (2004). Preventing disruptive behavior in elementary schoolchildren: Impact of a universal classroom-based intervention. *Journal of Consulting and Clinical Psychology, 72*(3), 467.

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.

Weissbourd, R., Bouffard, S. M., & Jones, S. M. (2013). School climate and moral and social development. *School Climate Practices for Implementation and Sustainability, 30*, 1–5.

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

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.

Zullig, K. J., Huebner, E. S., & Patton, J. M. (2011). Relationships among school climate domains and school satisfaction. *Psychology in the Schools, 48*(2), 133–145.