

Long-Term Effects of Social–Emotional Learning on Receipt of Special Education and Grade Retention: Evidence From a Randomized Trial of *INSIGHTS*

Meghan P. McCormick

MDRC

Robin Neuhaus

E. Parham Horn

Erin E. O'Connor

New York University

Hope S. White 

SUNY Buffalo

Samantha Harding

Elise Cappella

Sandee McClowry

New York University

Social–Emotional Learning (SEL) programs are school-based preventive interventions that aim to improve children’s social–emotional skills and behavioral development. Although meta-analytic research has shown that SEL programs can improve academic and behavioral outcomes in the short term, few studies have examined program effects on receipt of special education services and grade retention in the longer term. Using an experimental design, the current study leveraged administrative data available through students’ school records (N = 1,634) to examine the impacts of one SEL program implemented in kindergarten and first grade on receipt of special education and grade retention in fifth grade. The study further considered whether impacts varied for low- versus high-income students. Findings revealed no difference between treatment and control group students in grade retention. However, treatment group students were less likely to ever receive special education services by the end of fifth grade, with low-income students appearing to drive this effect. Implications are discussed.

Keywords: *social–emotional learning, intervention, special education, grade retention*

Social–Emotional Learning (SEL) programs are school-based preventive interventions that aim to improve children’s social–emotional skills and behavioral development (Jones & Doolittle, 2017). Meta-analytic research has shown that SEL programs tend to achieve these aims in the short term (Durlak, Weissberg, Dymnicki, Taylor, & Schellinger, 2011). Further work (e.g., Bierman et al., 2010; Brackett, Rivers, Reyes, & Salovey, 2012; Diamond & Lee, 2011) demonstrates that SEL programs can improve children’s academic skills, although that evidence base is more mixed (Durlak et al., 2011; Social and Character Development Research Consortium, 2010). Few studies, however, have been able to track intervention study participants across time in order to understand the long-term impacts of SEL programming. Indeed, a recent meta-analysis by Taylor, Oberle, Durlak, and

Weissberg (2017) found that of 82 SEL program impact studies with at least 6 months of follow-up data, only 6 studies considered effects 4 or more years postintervention.

To date, evaluators have generally tested how SEL programs affect children’s developmental outcomes, perhaps in line with theory hypothesizing that improvements in skills resulting from intervention will beget gains in later skills (Heckman, 2006; Magnuson & Duncan, 2016). A subset of rigorous evaluations has examined effects of SEL programs on receipt of special education (SPED) services and grade retention (Taylor et al., 2017), two outcomes that policy makers are increasingly interested in due to their cost implications for districts. In addition, recent work has further theorized that receipt of SPED services and grade retention may be particularly



important outcomes for determining children's future educational trajectories. For example, as theorized by Bailey, Duncan, Odgers, and Yu (2017), some interventions can have long-term outcomes because they may "equip a child with the right skills or capacities at the right time to avoid imminent risks (. . .) or seize emerging opportunities" (p. 7). In general, SEL programs aim to change student behaviors by supporting children in recognizing/managing emotions, setting and achieving goals, appreciating the perspectives of others, establishing and maintaining positive relationships, making good decisions, and handling interpersonal situations constructively (Elias et al., 1997). In changing these behaviors, it is also possible that SEL programs may qualitatively shift the likelihood that children would be referred to receive SPED services or be retained in grade because behavioral issues would be less likely confounded with need for SPED services or grade repetition. Yet, few studies have examined effects of SEL programs on receipt of SPED services and grade retention partly because longitudinal follow-up on a larger sample is needed to detect impacts on these outcomes.

It is also critical to consider variation in impacts for the children most likely to benefit from SEL programs in the longer term. Work rooted in prevention science asserts that children at heightened risk for social-emotional and/or academic difficulties may exhibit the largest improvements in outcomes after participating in an SEL program (Cicchetti & Aber, 1998; Jones, Brown, & Lawrence Aber, 2011; Rimm-Kaufman et al., 2014). At the same time, theories of differential susceptibility assert that some children are more likely to be affected by environmental contexts and supports than others (Belsky, Bakermans-Kranenburg, & Van Ijzendoorn, 2007). Early behavioral issues are associated with heightened risk for SPED services and/or grade repetition (Blair, 2001; Xia & Kirby, 2009). Given that a much higher proportion of children from low-income backgrounds are reported to have behavioral problems (Qi & Kaiser, 2003), and that children with behavioral and academic problems may demonstrate the greatest gains from SEL programming (Rimm-Kaufman et al., 2014), it is plausible that low-income children who participate in SEL interventions will be less likely to receive SPED services and be retained in grade, in particular. Yet, few longitudinal studies have had diverse enough samples to examine how impacts of SEL programs vary based on individual characteristics and experiences.

Leveraging longitudinal data from a randomized control trial of one SEL program implemented at the transition to elementary school—*INSIGHTS into Children's Temperament*—we add to the literature by examining intervention impacts on grade retention and receipt of SPED services at the end of fifth grade. We then test for variation in program impacts for low-versus high-income children.

Effects of SEL Programs on Grade Retention and Receipt of Special Education

The transition to formal schooling—defined in this article as the start of kindergarten—presents multiple social demands, such as exhibiting nonaggressive and well-regulated behavior (Thompson & Raikes, 2007), paying attention and mastering core academic skills (Eisenberg, Valiente, & Eggum, 2010), and partaking in social interactions (Kellam et al., 2008). Participation in SEL programs can support young children to develop core skills of self-management, self-awareness, social awareness, responsible decision making, and relationship skills (Collaborative for Academic, Social, and Emotional Learning, 2012), which in turn should bolster children's ability to successfully engage in these behaviors and navigate the transition to formal schooling. This is an important foundation because when students do not develop critical social and behavioral readiness skills at the transition to elementary school, the implications for schools can be even more costly with regard to greater likelihood of grade retention, receipt of academic support services, and suspensions/expulsions (Bettencourt, Gross, & Ho, 2016).

Evidence-based SEL programs differ in their approaches, but tend to use teacher professional development (e.g., training sometimes paired with coaching) and a classroom-based curriculum (which may vary in the extent to which they are scripted) to support teachers in using strategies that help students develop social-emotional skills (Low, Smolkowski, & Cook, 2016). Teachers are generally the primary delivery mechanism, often integrating SEL content throughout instructional time in addition to program components that are more manualized. Teachers may support students to recognize and respond to different emotions (Preschool PATHS; Bierman et al., 2010), use evidence-based strategies to respond to challenging behaviors (Incredible Years; Webster-Stratton, Jamila Reid, & Stoolmiller, 2008), or use information on children's unique temperaments to apply differentiated sets of strategies for responding to individual child behaviors (i.e., *INSIGHTS*; McClowry, 2014).

There is robust evidence demonstrating that children who participate in SEL programs exhibit fewer disruptive behavior problems and better behavioral regulation postintervention than children in a comparison or control group (e.g., Durlak et al., 2011; Taylor et al., 2017). Developmental cascades theory would suggest that interventions like SEL programs implemented at a sensitive time like the transition to kindergarten (Heckman, 2006; Masten, Long, Kuo, McCormick, & Desjardins, 2009; Reynolds & Temple, 2006) can lead to positive cascading effects by targeting competence in domains that are linked to functioning across other domains over time (Masten & Cicchetti, 2010). For instance, SEL interventions can trigger such a cascade by targeting social-emotional competencies (e.g., self-regulation) which

are linked to improved functioning across multiple domains (e.g., behavior, academic skills; Durlak et al., 2011). Given the increased likelihood for teachers to recommend children with behavioral and emotion regulation problems for retention or SPED services (Bettencourt et al., 2016; Davoudzadeh, McTernan, & Grimm, 2015), by supporting student social-emotional competencies, SEL interventions may lead to a lower likelihood of referral to SPED or grade retention due to the positive cascading effects triggered by the intervention.

Receipt of Special Education Services

Schools generally begin providing SPED services to students in kindergarten (Levine, 2016). There are a variety of reasons why children might be referred to services and given an Individualized Education Plan, but disruptive behavior problems are one key reason for SPED referral at the transition to formal schooling (Carlson et al., 2009; McIntyre, Eckert, Fiese, Reed, & Wildenger, 2010). Low-income students are also more likely than wealthier students to receive SPED services (either via pull out from the classroom, referral to a SPED contained classroom, or enrollment at a school specifically for SPED students) due to behavioral issues (Coutinho & Oswald, 2000; C. O'Connor & Fernandez, 2006). Although SPED services are critical to supporting the positive development of children who need them (Morgan & Sideridis, 2006), some children are referred to SPED simply due to difficulties in supporting their early behavioral regulation (Morgan et al., 2015). SPED services might be less relevant for these students because they mostly require behavioral supports in the classroom during early schooling, rather than a more comprehensive set of services across the full elementary school experience.

Although children who receive SPED services do for the most part require intervention supports, it is also true that these children are more likely to exhibit poor academic and social-emotional outcomes in middle childhood and early adolescence, and are more likely to stay in SPED throughout schooling rather than be integrated in general education classrooms (Harry & Klingner, 2014). SPED services are also costly: The *Individuals with Disabilities Education Act* allocated approximately \$4,700 per student receiving SPED services for the 2013–2014 academic year (Snyder, de Brey, & Dillow, 2016), over and above the typical cost of general education.

There is some empirical evidence to suggest that SEL programs implemented in early elementary school can reduce receipt of SPED services. In their long-term follow-up of the Good Behavior Game paired with an academically enhanced curriculum implemented in first grade, Bradshaw, Zmuda, Kellam, and Ialongo (2009) found that low-income African American children assigned to the intervention were 6 percentage points less likely to receive SPED services 11 years postintervention, relative to the control group. Even a small reduction in the percentage of students receiving

SPED services can represent a likely cost savings for districts.

Grade Retention

In a similar way, participation in SEL programming could change whether children are retained in grade. Although some studies have found positive short-term impacts of grade retention on behavioral and psychosocial outcomes (e.g., Wu, West, & Hughes, 2010), the majority of work considering longer term follow-up periods has generally found grade retention to have negative effects on students (Hong & Yu, 2007; Hwang & Cappella, 2018). For instance, using a regression discontinuity design based on a test score cutoff, Mariano, Martorell, and Tsai (2018) found that grade retention reduced high school credit accumulation and the likelihood of taking math and English Regents exams. In addition, Hughes, West, Kim, and Bauer (2018) used a propensity score analysis and found that retention in Grades 1 to 5 increased the likelihood of high school dropout. Similar to SPED services, grade retention in itself is expensive. According to a conservative estimate of a 1% annual retention rate, nearly five hundred thousand children are retained in the United States each year (Eide & Goldhaber, 2005; Warren & Saliba, 2012). Given that in 2013–2014, the average per pupil cost of an additional year in public school amounted to \$12,509 (\$24,116 in New York; U.S. Department of Education, National Center for Education Statistics, 2017), the estimated annual cost of retention is approximately \$6.6 billion per year.

When children display enhanced social-emotional skills in the classroom as a result of SEL intervention, it is possible that these enhanced competencies will reduce the likelihood that they will be retained in grade due to behavioral issues. Indeed, recent work by Mattison et al. (2018) has found that kindergarten students whose teachers rate them to be high on externalizing behavior problems are more likely to be retained between first and fifth grades. Some quasi-experimental work has found that participation in SEL programs can reduce the probability that students will be retained in grade. For example, Hawkins, Catalano, Kosterman, Abbott, and Hill (1999) found that children who received the Seattle Social Development program (a combination of teacher training, parenting classes, and social competence training for children in Grades 1–6) were 8 percentage points less likely to repeat a grade by age 18 years, relative to a matched (but not randomly assigned) control group.

However, the number of evaluation studies that have considered receipt of SPED services and grade retention as outcomes are limited. In addition, the extant work in this domain examines interventions implemented in the early 1980s and 1990s, prior to the zeitgeist movements in the early 2000s to not only expand the availability and quality of SEL programs but also to test their efficacy using rigorous randomized trials. Questions thus remain about whether the field is likely

to see long-term effects of contemporary SEL programs on receipt of SPED services and grade retention.

Variation in Impacts on Receipt of SPED for Low-Income Children

SEL programs may have larger impacts on the SPED receipt and grade retention of lower income students. Indeed, lower income children are more likely than affluent students to receive SPED for behavioral issues in the early grades (Harry & Klingner, 2014), a core targeted outcome of SEL programming. This overrepresentation may be a function of lower income children exhibiting more behavior problems, or else may indicate teachers' heightened frustration with behavior problems in classrooms and schools serving primarily lower income children (Bettencourt et al., 2016). Similarly, whereas grade retention can actually be beneficial for higher income children perhaps due to redshirting practices and/or because teachers in schools with more resources are better attuned to the needs of these students and observe benefits in retention (Fortner & Jenkins, 2017), additional research finds that lower income and racial/ethnic minority students are at heightened risk for grade retention due to early behavior problems (Davoudzadeh et al., 2015). If an SEL program does succeed in achieving its target of improving students' behavioral regulation, this may be more effective in reducing receipt of SPED and grade retention for lower income students who are at heightened risk for these outcomes specifically because of early behavior issues at the transition to elementary school. Few studies to date, however, have had diverse enough samples to explicitly test this hypothesis.

INSIGHTS Into Children's Temperament

INSIGHTS is a comprehensive SEL program that supports children's ability to self-regulate by enhancing their attention and behavior management. Details about the intervention components are included in Appendix A. A prior study found that the program reduced children's disruptive behavior problems and improved sustained attention by the end of first grade (E. E. O'Connor, Cappella, McCormick, & McClowry, 2014). This study did not consider program effects on SPED receipt and grade retention because the follow-up was not long enough to observe sufficient change in these outcomes to detect statistically significant program impacts. Indeed, in experimental studies of SEL programs it is rare to have longitudinal data to examine outcomes many years after the end of the intervention. Longitudinal data are likely necessary when estimating impacts on receipt of SPED services and grade retention because the changes in these outcomes take time to manifest themselves. However, administrative data on these outcomes were available as children who participated in study schools moved from kindergarten through fifth grade, making it possible to estimate

impacts on receipt of SPED and grade retention in a longer term follow-up study.

Although there are other examples of prior studies that considered long-term effects of SEL programs on receipt of SPED services (Bradshaw et al., 2009) and grade retention (Hawkins et al., 1999), the current effort adds to the extant literature in four key ways. First, the prior experimental and quasi-experimental studies that examined long-term effects on receipt of SPED (Bradshaw et al., 2009) and grade retention (Hawkins et al., 1999) considered interventions that were implemented in the early 1990s. Given the proliferation of SEL programs and practices in schools across the past 15 years, the studies from the 1990s likely generated a much larger service contrast in provision of SEL activities than would be expected today. More up to date analyses may provide more realistic impact estimates for current programs operating today.

Second, both the Bradshaw et al. (2009) and Hawkins et al. (1999) studies were conducted in within-group, high-need samples. Accordingly, the generalizability of these findings may be limited to the populations and contexts under which they were examined. In contrast, the current study adds to the literature because it includes information on a sample from a different city and contains sufficient diversity in the student population to consider variation in impacts by students' family income. Third, the interventions studied in prior work were fairly intensive and required multiple days of time from teachers to be trained on the interventions, and a larger amount of resources to support implementation. The *INSIGHTS* intervention, in contrast, is comparatively easier to implement and requires less intensive commitment from teachers, students, and schools. Accordingly, the current study adds to the literature by estimating impacts of a less intensive SEL program (perhaps more scalable in future work) for potentially reducing SPED receipt and grade retention.

Finally, more generally, given the documented difficulty in replicating effects across intervention studies noted by Stroebe and Strack (2014) and described by many others in the social sciences as the "replicability crisis," it is critical to consider replicating impacts across different interventions and across time. Examining similar sets of outcomes across different intervention studies is important for building the evidence base on the range of curricula and SEL programs that can have lasting impacts on SPED receipt and grade retention. The current study is well-poised to add to the literature by determining whether different types of models—including less intensive SEL programs—have impacts on receipt of SPED and grade retention, testing effects in more contemporary data, and examining variation in impacts using longitudinal data.

The Current Study

To this end, the current article will first estimate the impacts of one SEL program for kindergarten and first-grade

students—*INSIGHTS into Children's Temperament*—on receipt of SPED services and grade retention at the end of fifth grade. Second, the article will consider how impacts of *INSIGHTS* vary for children from lower income families. This study will build evidence about how one particular SEL program does or does not affect the extent to which children access different educational experiences that they otherwise would not have been exposed to in the absence of the intervention.

Method

Between 2008 and 2012, a total of 22 elementary schools from three New York City school districts were randomly assigned to participate in the *INSIGHTS* intervention or to an attention-control condition. The current study uses administrative data for the students enrolled in study schools in the fall of kindergarten, which was the point at which schools were randomly assigned to study conditions. The study uses administrative data on children's demographic characteristics, subsequent school locations, receipt of SPED, and grade retention between kindergarten and fifth grade.

Participants and Setting

Participants in this study included all students ($N = 1,634$) who were enrolled in kindergarten—the time of random assignment—in an intervention or control school during the years when the randomized controlled trial took place.¹ See demographic characteristics of the study schools and students in Table 1. Children included in the current study sample were representative of the broader total school population.²

Recruitment and Randomization

Schools were initially recruited for participation in the intervention study between 2008 and 2010. The recruitment strategies were approved by university and school system research boards. Principals serving primarily low-income students in three urban school districts were the first to be contacted.³ Team members explained the purpose of the study and its related logistics including randomization to one of two intervention conditions: *INSIGHTS* or the attention-control condition which was an after-school reading program (see further details below). Twenty-three schools were invited and initially agreed to participate. Prior to randomization, one school withdrew from the study due to a principal transition. Teachers were initially recruited in small group or individual meetings. Ninety-six percent of the kindergarten and first-grade teachers at participating schools consented to participate. All teachers agreed to participate for the duration of the study.

In the original study, a subsample of students within the participating schools ($N = 435$) received parental consent for

data collection (see E. E. O'Connor et al., 2014, for more details). In contrast, the current study leverages de-identified data on *all* students who were enrolled in participating schools and thus represents a larger sample than the group of students who consented for research activities. This is possible because all children in participating schools received the intervention delivered by teachers and facilitators, regardless of whether they had consented to participate in data collection. The only difference for the consented subsample is that parents in *INSIGHTS* schools received the opportunity to participate in the parent program. In the current study, students were identified as members of the treatment group based on their enrollment in kindergarten—the point of random assignment—in one of the schools assigned to *INSIGHTS*.

After baseline data were collected in kindergarten, a random numbers table was used to randomize schools to *INSIGHTS* or the attention-control group. Schools were used as the unit of random assignment to limit possible contamination effects that could threaten the internal validity of the study. Eleven schools were randomized to *INSIGHTS*, and 11 schools were assigned to the attention-control group. There were the same number of treatment and control group schools within each participating district. Chi-square tests were used to examine baseline equivalence between the *INSIGHTS* ($N = 800$ students, 11 schools) and attention-control groups ($N = 843$ students, 11 schools) on observed characteristics available on study schools and students enrolled in participating schools. We used independent samples *t* tests to compare baseline characteristics between students and schools assigned to *INSIGHTS* versus the attention-control group. Similar to findings in E. E. O'Connor et al. (2014), preliminary analyses revealed no statistically significant differences between schools assigned to the *INSIGHTS* program versus schools assigned to the attention-control group. Similarly, there were no differences between the group of students assigned to the *INSIGHTS* group versus the attention-control group. See Table 1 for more detailed information establishing baseline equivalence.

Data Collection

The data for this study were obtained from the New York City Department of Education through the Research Alliance for New York City Schools. The New York City Department of Education collects and records administrative data on all students at the beginning (October) and end (June) of each school year. The research team received administrative data through fifth grade for all students who were enrolled in kindergarten in one of the schools participating in the study.

Measures

Receipt of Special Education Services. In the administrative data, receipt of SPED services (excluding services related to

TABLE 1

Baseline Descriptive Statistic for INSIGHTS and Attention-Control Groups

	Attention-control		INSIGHTS		
Characteristic	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	Difference
School characteristics					
Annual attendance rate	88%	—	86%	—	0.02
Female	51%	—	49%	—	0.02
Black	76%	—	79%	—	−0.03
Hispanic	39%	—	43%	—	−0.04
White	11%	—	9%	—	0.02
Other race	13%	—	12%	—	0.01
Eligible for free or reduced-price lunch	72%	—	75%	—	−0.03
Dual language learner	8%	—	6%	—	0.02
From recent immigrant family	6%	—	5%	—	0.01
Prekindergarten attender	59%	—	62%	—	−0.03
% students scoring average on ELA state test	53%	—	49%	—	0.04
% students scoring average on math state test	63%	—	60%	—	0.03
Number of students in school	505.07	190.09	493.21	138.62	11.86
Sample size	11		11		
Student characteristics for full sample					
Baseline age (years)	5.17	0.72	5.19	0.66	−0.02
Female	48%	—	48%	—	0
Black	73%	—	77%	—	−0.04
Hispanic	35%	—	38%	—	−0.03
White	12%	—	10%	—	0.02
Other race	15%	—	13%	—	0.02
Eligibility for free or reduced price lunch	72%	—	75%	—	−0.03
Dual language learner	9%	—	11%	—	−0.02
From recent immigrant family	4%	—	6%	—	−0.02
Prekindergarten attender	62%	—	66%	—	−0.04
Received special education (SPED) services at baseline	1%	—	1%	—	0
Sample size	843		800		
Student characteristics for nonattrited analytic sample					
Baseline age (years)	5.24	0.71	5.18	0.69	0.06
Female	49%	—	50%	—	−0.01
Black	75%	—	79%	—	−0.04
Hispanic	36%	—	39%	—	−0.03
White	11%	—	9%	—	0.02
Other race	14%	—	12%	—	0.02
Eligibility for free or reduced price lunch	77%	—	79%	—	−0.02
Dual language learner	10%	—	9%	—	0.01
From recent immigrant family	5%	—	6%	—	−0.01
Prekindergarten attender	65%	—	69%	—	−0.04
Received SPED services at baseline	1%	—	1%	—	0
Sample size	639		624		

Note. We used independent-samples *t* tests to examine whether the treatment and control groups were significantly different from one another. None of the group-based differences in the final column were statistically significant.

physical disabilities, vision, and/or hearing problems) was indicated for each grade level for each study participant. Receipt of SPED also included enrollment in a stand-alone SPED school. One limitation of these data is that they did

not provide details on the reason for receipt of SPED. Accordingly, receipt of SPED services could be related to a host of issues (e.g., behavioral, academic, speech supports, etc.). We do know, however, that receipt of any language

support services for English language learners is not included in this variable. In addition, receipt of SPED was highly correlated across years, such that students who received SPED services in first, second, third, or fourth grades tended to continue to receive SPED services in each year thereafter.

We first created grade-specific dummy variables to describe whether the student received SPED services within that grade. Once a variable was created for each grade, another dummy variable was coded to identify students who received SPED services during any grade from first grade (the year that intervention ended) to fifth grade, for example, ever referred for SPED services (1 = received SPED services between first and fifth grade; 0 = never received SPED services between first and fifth grade).

Grade Retention. A dummy variable was created to indicate whether a student was ever retained between kindergarten and fifth grade (1 = retained anytime between kindergarten and fifth grade; 0 = never retained). If the student's actual grade level was behind the expected grade level in any grade between kindergarten and fifth grade, then the student was coded as retained. For the small number of cases where values appeared problematic (e.g., if the expected grade level was behind in one year but then ahead of expected grade in a future year), we cross-referenced values with the school discharge variable to determine grade retention categorization.

Demographic Characteristics. At public school enrollment, parents and guardians reported on their children's demographic characteristics—race, ethnicity, gender, eligibility for free or reduced price lunch (used in the current study to describe a student as low-income), immigration status, birthdate, and whether their child spoke a language other than English. These variables were used as covariates in the analysis. Child age on September 1 of the kindergarten year was also calculated using the birthdate and included as a covariate in analytic models.

School-Level Characteristics. We included a set of covariates at the school level as well in our predictive models that match the set of school-level covariates we used in our short-term follow-up study. We used publicly available administrative data to capture the school size (number of students enrolled in the year prior to the intervention implementation), school attendance rate, the percentage of students in the school who were proficient on the state ELA test in the prior year, and the percentage of students in the school proficient on the state math test in the prior year. We further adjusted for study cohort at the school level including dummy variables for Cohort 2 and Cohort 3, with Cohort 1 as the reference group. We used some additional school-level demographic characteristics to examine baseline equivalence at the school level. These were school-level

versions of the student-level demographic characteristics (percent female, percent Black, percent Hispanic, percent eligible for free/reduced price lunch, percent immigrant, percent dual language learner, and percent who attended a district PreK program).

INSIGHTS Intervention Procedures and Attention-Control

Participating schools were divided into three cohorts (Cohort 1: Fall 2008–Spring 2010; Cohort 2: Fall 2009–Spring 2011; Cohort 3: Fall 2010–Spring 2012) and agreed to implement the *INSIGHTS* or attention-control program in kindergarten and first grade. More specific information about the intervention and attention-control conditions is located in Appendix A.

Analytic Approach

Missing Data Analysis. As noted above, the total analytic sample size for this study is $N = 1,634$ students. Of the total number of students, 1,263 (77%) remained in the sample through the end of the fifth grade, while 371 (23%) are missing outcome data because they attrited from the school district before the end of the fifth grade. Follow-up analyses revealed that students in the treatment and attention-control group attrited at similar rates (24% of the baseline attention-control group attrited and 22% of the treatment group attrited). In Table 1, we further illustrate how the baseline characteristics are similar across the nonattrited students in the treatment and attention-control groups and further detail on attrition is included in Appendix B. According to the standards from the What Works Clearinghouse (2017), this study constitutes a low-attrition RCT, meaning that overall attrition and differential attrition indicate a tolerable level of bias for the estimated intervention effect. Using the assumption that data were Missing at Random, the team used a multiple data imputation method in STATA MICE and imputed 100 data sets in order to generate complete data on all covariates, which had minimal levels of missingness (Graham, Olchowski, & Gilreath, 2007).

Impact Analysis. Because school was the original unit of random assignment and most students stayed in the same elementary school that they enrolled in during kindergarten, we expected that student outcomes would not be independent at follow-up. In order to determine the most appropriate fit to the data, we considered a number of different models and compared their log likelihoods and fit statistics (Akaike information criterion [AIC] and Bayesian information criterion [BIC]). This approach for model selection has been recommended by Scott, Simonoff, and Marx (2013). For both sets of outcomes—receipt of SPED services and grade retention through fifth grade—we first ran unconditional logit models with clustered standard errors for schools. Then, we

TABLE 2

Descriptive Statistics for Grade Retention and Special Education (SPED) Referral Through Fifth Grade Unadjusted

Outcome variable	<i>INSIGHTS</i>	Attention-control	Difference	Significant difference
Full sample				
Ever retained in grade	25.86%	24.61%	1.25%	
Ever received SPED services	6.95%	10.34%	-3.39%	*
Low-income sample				
Ever retained in grade	22.89%	21.79%	1.10%	
Ever received SPED services	8.87%	13.46%	-4.59%	*
Higher income sample				
Ever retained in grade	10.43%	9.52%	0.91%	
Ever received SPED services	6.86%	5.80%	1.06%	

Note. $N = 1,263$ students with available outcome data, $N = 973$ students in the low-income group, $N = 290$ in the high-income group.

* $p < .05$. ** $p < .01$.

compared that model fit to a mixed-effects logistic regression (Fitzmaurice, Laird, & Ware, 2004) which included random effects for the school that participated in the original trial. After comparing the log likelihoods of the models and the AIC/BIC statistics across these specifications (these statistics are illustrated in Model 1 in Tables 3 and 4 for the mixed-effects logistic regressions and in Appendix D Tables D1 and D2 for the logistic regressions with clustered standard errors for study school), we found that the mixed-effects logistic regression was a better fit to the data for both outcomes. We then considered a number of covariance structures, including the unstructured, independent, exchangeable, and identity structures. Results were almost identical regardless across approaches, so we retained the unstructured covariance structure for our main set of analyses. The baseline random intercepts model for examining treatment impact is illustrated in Equation 1.

$$\text{logit}\{\Pr(y_{ij} = 1 | \chi_{ij}, \zeta_j)\} = \beta_1 + \beta_2 \text{INSIGHTS}_j + \zeta_j \quad (1)$$

where y_{ij} refers to the outcome for student i in school j ; ζ_j is a random intercept for school, and β_2 is the treatment impact for students who participated in the study in school j . It is important to note here that assignment to *INSIGHTS* is operationalized as a school-level variable. The full set of nested models including all Level-1 and Level-2 covariates is included in Appendix C. After estimating the main effects of *INSIGHTS* on the study outcomes, we then considered how impacts varied for low-income students by creating cross-level interactions between assigned to *INSIGHTS* and the dummy variable for low income.

All findings are reported as odds ratios. For statistically significant main effects and moderated impacts, we then calculated probabilities for treatment and control group members. Differences in the probabilities can be interpreted as percentage differences and facilitate interpretation and comparisons with prior research. Given the relatively small number of

schools in the study and potential for imbalance in findings, we also conducted a series of robustness checks and tested models using logistic regression with clustered standard errors for study school. We further considered how robust the impacts were when we group-mean centered the Level-1 covariates.

Results

Descriptive Statistics for Outcome Variables

Frequencies for the overall sample, and for both low-income and higher income groups by intervention status are reported in Table 2. Independent-samples t tests revealed that students in the *INSIGHTS* group were significantly less likely to receive SPED services through fifth grade: mean difference = -3.39; $t(1,261) = 14.56$, $p < .05$. In addition, low-income children enrolled in the *INSIGHTS* group were also less likely to receive SPED services through the end of fifth grade relative to low-income children enrolled in the attention-control condition: mean difference = -4.59; $t(1,261) = 24.86$, $p < .05$.

Impacts of INSIGHTS on Grade Retention and Receipt of SPED Services

Impact results examining main effects are presented in Table 3 (grade retention) and Table 4 (receipt of SPED services). As shown here, there were no main effects of *INSIGHTS* on grade retention through fifth grade (odds ratio [OR] = 0.79, $p = .33$). However, there was a statistically significant impact on receipt of SPED services such that students enrolled in *INSIGHTS* schools were less likely to receive SPED services through fifth grade than students in the attention-control condition ($OR = 0.85$, $p = .04$). We used the conditional odds ratio to calculate the probability of receiving SPED services by fifth grade for a student in the *INSIGHTS* group versus the control group, who was a member of the reference group on all covariates. Reference group students in the *INSIGHTS* group had a 7% probability of

TABLE 3
Impacts of *INSIGHTS* on Grade Retention

	Model 1			Model 2			Model 3			Model 4		
Fixed effects	OR	Log(odds)	SE	OR	Log(odds)	SE	OR	Log(odds)	SE	OR	Log(odds)	SE
Intercept	0.27	-1.31**	0.15	0.19	-1.67**	0.21	0.33	-1.10**	0.24	0.34	-1.08**	0.22
Treatment (<i>INSIGHTS</i>)	—	—	—	0.77	-0.26	0.29	0.79	-0.23	0.22	0.79	-0.24	0.19
Student-level covariates												
Baseline age	—	—	—	—	—	—	0.61	-0.49**	0.11	0.63	-0.46**	0.09
Female	—	—	—	—	—	—	0.82	-0.20	0.15	0.79	-0.24	0.16
Black	—	—	—	—	—	—	1.07	0.07	0.36	1.01	0.01	0.35
Hispanic	—	—	—	—	—	—	1.31	0.27	0.49	1.32	0.28	0.49
Low SES	—	—	—	—	—	—	2.01	0.70*	0.33	2.04	0.71*	0.31
Dual language learner	—	—	—	—	—	—	0.63	-0.46*	0.19	0.59	-0.53*	0.17
From recent immigrant family	—	—	—	—	—	—	0.72	-0.33	0.27	0.75	-0.29	0.29
Prekindergarten attender	—	—	—	—	—	—	0.61	-0.49**	0.11	0.58	-0.54**	0.09
School-level covariates												
Cohort 2	—	—	—	—	—	—	—	—	—	1.21	0.19	0.25
Cohort 3	—	—	—	—	—	—	—	—	—	1.22	0.20	0.23
School size	—	—	—	—	—	—	—	—	—	1.33	0.29	0.28
School % competent	—	—	—	—	—	—	—	—	—	0.87	-0.14*	0.06
ELA test												
School % competent math test	—	—	—	—	—	—	—	—	—	0.81	-0.21*	0.05
School average attendance	—	—	—	—	—	—	—	—	—	0.85	-0.16	0.15
Variance components	Variance		SE	Variance		SE	Variance		SE	Variance		SE
Baseline study school	0.08*		0.03	0.08*		0.04	0.06		0.04	0.06		0.04
Fit statistics												
AIC	351.48			342.47			328.47			318.46		
BIC	358.89			362.57			351.74			334.42		
Log likelihood	-173.64			-171.24			-167.84			-161.84		

Note. OR = odds ratio; SE = standard error; SES = socioeconomic status; AIC = Akaike information criterion; BIC = Bayesian information criterion. We used a series of models in this impact analysis to examine the sensitivity of our results to the inclusion of covariates. Model 1 in this table is an unconditional model that includes no predictors. Model 2 examines the impacts of *INSIGHTS* on the outcome adjusting for no covariates. Next, Model 3 tests how sensitive the impact results are to the inclusion of a block of student-level covariates. Finally, Model 4 further tests how sensitive the impact results are to the inclusion of blocks of both student- and school-level covariates.

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

receiving SPED services by fifth grade, relative to a 12% probability for students in the attention-control condition.

We found that null effects of *INSIGHTS* on grade retention through fifth grade did not vary by student income. However, we did find a statistically significant interaction between assignment to *INSIGHTS* and low-income in predicting receipt of SPED services through fifth grade ($OR = 0.39$, $p = .04$; see Figure 1). The finding demonstrates that children who were low-income and in *INSIGHTS* schools were less likely than low-income students attending schools in the attention-control group to receive SPED services between kindergarten and fifth grade. Using the odds ratio estimate, we calculated the probabilities for the two groups

in the adjusted models, finding that low-income students in the *INSIGHTS* group (in the reference group for all covariates) had a 9% probability of receiving SPED services by fifth grade, relative to the low-income students in the control group who had a 15% probability of receiving SPED services by fifth grade. Further probing demonstrated that there was no difference in receipt of SPED for higher income students assigned to *INSIGHTS* versus the attention-control.

Robustness Checks

We used a series of robustness checks to examine how sensitive the impact analysis results were to our model

TABLE 4
Impacts of INSIGHTS on Receipt of Special Education Services

	Model 1			Model 2			Model 3			Model 4		
Fixed effects	OR	Log(odds)	SE	OR	Log(odds)	SE	OR	Log(odds)	SE	OR	Log(odds)	SE
Intercept	0.10	−2.31**	0.23	0.09	−2.41**	0.01	0.03	−3.51**	0.02	0.01	−4.61**	0.01
Treatment (<i>INSIGHTS</i>)	—	—	—	0.83	−0.19	0.12	0.81	−0.21*	0.09	0.85	−0.16*	0.07
Student-level covariates												
Baseline age	—	—	—	—	—	—	1.12	0.11	0.24	1.18	0.17	0.22
Female	—	—	—	—	—	—	0.44	−0.82**	0.14	0.42	−0.87**	0.08
Black	—	—	—	—	—	—	0.17	−1.77**	0.11	0.13	−2.04**	0.18
Hispanic	—	—	—	—	—	—	0.31	−1.17**	0.21	0.34	−1.08**	0.08
Low SES	—	—	—	—	—	—	0.28	−1.27*	0.56	0.24	−1.43**	0.56
Dual language learner	—	—	—	—	—	—	2.01	0.70	0.42	1.95	0.67	0.38
From recent immigrant family	—	—	—	—	—	—	0.63	−0.46	0.28	0.58	−0.54	0.31
Prekindergarten attender	—	—	—	—	—	—	0.89	−0.12	0.16	0.82	−0.20	0.16
School-level covariates												
Cohort 2	—	—	—	—	—	—	—	—	—	1.03	0.03	0.43
Cohort 3	—	—	—	—	—	—	—	—	—	1.11	0.10	0.53
School size	—	—	—	—	—	—	—	—	—	1.31	0.27*	0.11
School % competent ELA test	—	—	—	—	—	—	—	—	—	0.95	−0.05	0.21
School % competent math test	—	—	—	—	—	—	—	—	—	0.91	−0.09	0.33
School average attendance	—	—	—	—	—	—	—	—	—	1.11	0.10	0.29
Variance components	Variance		SE	Variance		SE	Variance		SE	Variance		SE
Study school	0.09**		0.03	0.09*		0.03	0.09*		0.02	0.06*		0.02
Fit statistics												
AIC	210.65			206.65			178.56			165.86		
BIC	217.67			211.76			186.43			180.32		
Log likelihood	−103.03			−101.91			−94.56			−93.45		

Note. OR = odds ratio; SE = standard error; SES = socioeconomic status; AIC = Akaike information criterion; BIC = Bayesian information criterion. We used a series of models in this impact analysis to examine the sensitivity of our results to the inclusion of covariates. Model 1 in this table is an unconditional model that includes no predictors. Model 2 examines the impacts of *INSIGHTS* on the outcome adjusting for no covariates. Next, Model 3 tests how sensitive the impact results are to the inclusion of a block of student-level covariates. Finally, Model 4 further tests how sensitive the impact results are to the inclusion of blocks of both student- and school-level covariates.

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

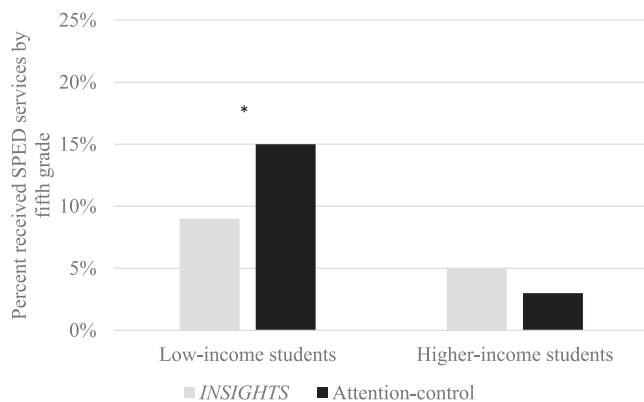


FIGURE 1. Impacts of *INSIGHTS* on receipt of SPED services by fifth grade.

* $p < .05$. ** $p < .01$.

specification. In general, we found that the bulk of our results were robust across different modeling assumptions. All robustness checks are described thoroughly in Appendix D.

Discussion

This study aimed to leverage existing administrative data on all students enrolled in schools that participated in a randomized control trial of the *INSIGHTS* program to determine whether there were long-term effects of the intervention on receipt of SPED services and grade retention from kindergarten through fifth grade—and whether outcomes varied for lower income versus higher income students. Findings revealed that students in the *INSIGHTS* group were 5 percentage points less likely than comparison group members to receive SPED services by fifth grade. However, this effect was sensitive to model selection and did not fully replicate in a model using clustered standard errors for study school. Even so, we further found that *INSIGHTS* reduced the likelihood of receiving SPED services for *low-income students* by 6 percentage points. This impact was not sensitive to model selection and we are thus more confident that the intervention reduced receipt of SPED services within the sample of low-income students. There were no main or moderated effects of *INSIGHTS* in reducing grade retention by fifth grade.

Impacts of *INSIGHTS* on receipt of SPED services are somewhat aligned with prior research. For example, Bradshaw et al. (2009) found that assignment to the Good Behavior Game reduced the likelihood of placement in a SPED classroom by 6 percentage points with is a similar effect size to the one found in this study, although the follow-up period was about twice as long in that prior study, so the research designs are not directly comparable. Work by Rimm-Kaufman, Pianta, and Cox (2000) has found that kindergarten teachers working in schools serving primarily low-income families are more likely to report significant behavioral problems for children at kindergarten entry. Lower income students who exhibit behavior problems in early educational settings are more likely than higher income students with equivalent behaviors to be referred to SPED (C. O'Connor & Fernandez, 2006; Shavelson & Towne, 2002). An earlier study demonstrated that *INSIGHTS* did reduce children's disruptive behaviors in the short term and improved their literacy and math skills (O'Connor et al., 2014). Importantly, children identified to be at behavioral risk at baseline benefited the most from the intervention in terms of their reduced disruptive behaviors and improved behavioral engagement (McCormick, O'Connor, Cappella, & McClowry, 2015). Positive impacts on students' behaviors in the short term may have later manifested in broader reductions in receipt of SPED services. This result may have been particularly true for the lower income students in the sample with a higher likelihood of receiving SPED services due to challenging behaviors (Qi & Kaiser, 2003; Xia & Kirby, 2009).

In contrast, the grade retention finding for this study is not fully aligned with prior work evaluating impacts of SEL programs (e.g., Hawkins et al., 1999; Hawkins et al., 2009). For example, in their 6-year follow-up of the Seattle Social Development Program, Hawkins and colleagues (1999) found that the program reduced grade retention by 12.5 percentage points. The null impact found in the current study may reflect the fact that although the strategies taught by the *INSIGHTS* program were intended to be integrated into typical instruction during the school day, the activities were not directly tied to academic content. The intervention itself may not have been sufficiently intense to improve children's academic skills to the point where it would change their likelihood of being retained in grade. SEL programs explicitly target social-emotional skills, which are theorized to then support gains in academic skills (Durlak et al., 2011). However, for children who are risk of being retained in grade based on poor early academic skills, more intensive supports directly focused on academic instruction may be necessary over and above involvement in SEL programming.

In addition, it is important to note that the comparison group in this study did receive a reading intervention. As such, we are comparing *INSIGHTS*—an SEL program—to a more active control group participating in an academically

oriented set of services rather than a business as usual control group. We originally decided to use this approach in order to ensure that any treatment effects we observed were related to the *INSIGHTS* SEL program and not the fact that treatment schools were simply receiving additional attention and professional development supports. The current study thus represents a relatively conservative test of the *INSIGHTS* program and more stringently examines the effect of a program to support social-emotional development relative to a more academically oriented intervention. With respect to the null effect on grade retention, it might have been that the supports from the Read Aloud program were similarly effective in supporting children's grade promotion. Further work examining the effects of *INSIGHTS*—and other SEL studies—on grade retention may consider a less active control group in order to increase the treatment contrast and see how that change does or does not affect impacts on grade retention.

Prior studies that have looked at the longer term effects of SEL programs on student outcomes, albeit from implementation efforts several years ago, have found longer run effects through high school and beyond (Taylor et al., 2017). For example, earlier studies have found effects on outcomes such as high school dropout or graduation (Bradshaw et al., 2009; Felner et al., 1993; Hawkins, Kosterman, Catalano, Hill, & Abbot, 2008), college attendance or completing a college degree (Bradshaw et al., 2009; Hawkins et al., 2008), safe sexual behaviors and STD diagnosis (Hill et al., 2014), criminal involvement or arrests (Eddy, Reid, Stoolmiller, & Fetrow, 2003; Hawkins et al., 2008), and adult mental health outcomes (Hawkins et al., 2008; Jalongo, Poduska, Werthamer, & Kellam, 2001; Riggs & Pentz, 2009). More contemporary evidence (e.g., longer term impacts of the Positive Action program; Duncan et al., 2017) has demonstrated benefits of universal SEL program in reducing behavior problems through eighth grade. Yet, no studies to date have explicitly tested whether impacts of SEL programming on receipt of SPED services did in fact lead to longer term outcomes in high school and beyond. This study represents an initial step in identifying an impact of one SEL program on receipt of SPED services through fifth grade. Even so, additional research in the future would be needed to explicitly test whether reductions in receipt of SPED services in the short term would then link to long-term effects in high school and beyond.

Limitations

Although this study uses an experimental design and multiple years of follow-up data for a large sample, there are a number of limitations. First, the study sample is made up of the total population of students who were enrolled in kindergarten in schools participating in the study during their first year of study implementation. Although we know that

students who consented to the study participated at high rates, we cannot determine the exact level of dosage for the students included in this study. In order to use an intent-to-treat design in this study, we included students in the treatment group if they were enrolled in a school assigned to *INSIGHTS* during the kindergarten year when the intervention was initially implemented. If the child was enrolled in a control group school in kindergarten and then switched to *INSIGHTS*, they remained in the control group in order to maintain the internal validity of the intent to treat design, and vice versa. Review of students' school locations, however, suggests that crossovers from treatment to control group school were minimal. It was more likely that a student would leave an *INSIGHTS* school for a nonstudy school or vice versa. Even so, we are unable to describe the extent to which we observed noncompliance in this study for the full student sample.

Although the administrative data are a core strength of the study, it is still true that understanding these data is complicated, particularly with respect to grade retention. It is possible that there is some measurement error in operationalizing the outcome constructs of interest in this study. In order to limit the possibility of Type I error posed by running too many impact analyses, we restricted this article to consider four total impact models and two total outcomes. It is possible that if we used different operationalizations of the outcomes (e.g., comparing impacts on grade retention at different follow-up time points) we could have found differential effects of the treatment across time. Further work with greater statistical power to conduct multiple comparisons adjustments may consider probing these outcomes further.

We also found that the results of the analysis did not completely replicate across a model using random intercepts for study schools and a logistic regression using clustered standard errors for study school (see Appendix D). However, the findings from the logistic regression were similar in direction and magnitude and there is a fierce debate in the literature about the conceptual difference between findings at the $p < .05$ and $p < .10$ levels (Cristea & Ioannidis, 2018). Moreover, work by Roodman, Nielsen, MacKinnon, and Webb (2019) has shown that clustered standard errors may be too conservative of an approach when there are few clusters of interest. We do interpret the main effects of *INSIGHTS* on receipt of SPED services with caution given lack of exact replication. However, we have more confidence that *INSIGHTS* did reduce receipt of SPED services for low-income students within the sample—an effect that replicated across all models.

Finally, we lack detailed and nuanced data on the potential moderators and mechanisms explaining impacts of *INSIGHTS* on receipt of SPED and why those impacts were larger for lower income children. Furthermore, we were unable to demonstrate that the impacts on receipt of SPED were not accompanied by negative impacts in other domains

within the current study sample. Doing so would have allowed us to be more confident in our assertion that reducing receipt of SPED services is a positive outcome, and not actually indicative of a situation where children who need SPED services are unable to access them. Our past work examining the short-term impacts of the program (E. E. O'Connor et al., 2014) demonstrated that *INSIGHTS* reduced disruptive behaviors and improved behavioral engagement and that those effects were driven by students who had challenging temperaments at baseline (McCormick et al., 2015). In addition, our team has completed a complementary set of analyses (McCormick, O'Connor, Cappella, & McClowry, 2019) that examine the impacts of *INSIGHTS* on academic skills outcomes in third, fourth, and fifth grades for students who were in the original trial ($N = 331$ with complete data at follow-up). Although these results are still under review, we have found positive treatment impacts on language skills in third and fourth grade and no group-based differences on children's math skills. This is only a subsample of the group of students examined in the current study, but findings do provide further evidence that a reduction in receipt of SPED services is not associated with negative outcomes for students in other domains.

Conclusions and Implications

Results from this study demonstrate that a time-limited and fairly intensive SEL intervention that supports teachers in effectively responding to and managing the behaviors of children implemented at the critical transition to elementary school can reduce receipt of SPED services through fifth grade for low-income students. The fields of intervention and education research are currently interested in identifying the mechanisms through which early intervention can affect longer term outcomes. Rather than following a linear process, the current study provides some early evidence for Bailey et al.'s (2017) contention that early programming can provide access to qualitatively different experiences for students—for example, by reducing the likelihood of that they will be referred to SPED services due to a behavioral issue—which may lead to longer term adaptive outcomes. Low-income children who need SPED services should continue to receive them. But SEL programs may be able to support children's behavioral development early on in a way that can prevent them from perhaps unnecessarily receiving SPED programming and better promoting their academic development at the transition to elementary school.

Policy makers can benefit from this work when considering cost implications of early programming, and integrating evidence-based SEL intervention into the early grades. Although some whole-school reforms to support students' social emotional skills can be quite costly, shorter embedded interventions—like *INSIGHTS*—can be less expensive. Moreover, if SEL programs can have even small effects on

outcomes like receipt of SPED services, they can still have significant cost savings for school districts serving low-income students. Although this study did not include an in-depth cost effectiveness component, one could potentially balance the cost of *INSIGHTS* implementation—estimated at about \$5,500/school for 2 years of programming (or about \$50/child)—against the expected cost of a lower income student receiving SPED services over and above regular education (average of \$4,700/year; Snyder et al., 2016). However, this is likely an overestimate in comparing costs as many of the expenses associated with the receipt of SPED are likely driven by a few very high-need children and the students who may benefit from intervention are likely lower need. Future work including a more in-depth cost study focused explicitly on the subsample of students who do not receive SPED services as a result of an SEL program is needed. Finally, the study provides initial evidence that practitioners can be trained to support a diverse set of students' social-emotional development in such a way that can link to adaptive outcomes even after those students transition out of trained teachers' classrooms. Integrating evidence-based SEL programming into teacher training and professional development on a broader scale may thus be warranted.

Appendix A

Information on the INSIGHTS Intervention and Attention-Control Conditions

INSIGHTS Into Children's Temperament. *INSIGHTS* is a comprehensive SEL program that supports children's ability to self-regulate by enhancing their attention and behavior management. The program aims to help teachers learn to recognize the consistent behavioral style that a child exhibits across settings as an expression of temperament. Although temperament is stable, adult responses can be adjusted to match each child's unique temperament. The aim of this effort is to enhance goodness of fit, the consonance of a child's particular temperament to the demands, expectations, and opportunities of the environment (Thomas & Chess, 1977). By accepting and understanding a child's particular temperament, teachers are able to use scaffolding and stretching strategies to better support the child when temperamentally challenging situations are encountered (Dennis, van Aken, Penke, & Wood, 2013; Rothbart, Sheese, Rueda, & Posner, 2011). Children also learn how to identify their own temperaments. The component of *INSIGHTS* directed at children focuses on expanding children's self-regulation and social-emotional skills. Children are introduced to four puppets with different temperaments during 10 weeks of classroom sessions. Utilizing the puppets, the children explore how some situations are easy for certain puppets, while others are more challenging. During the remaining weeks, the children work with the puppets to use problem-solving strategies when confronted with daily

dilemmas. This part of the intervention supports children in identifying strategies matched to their temperament to address the daily challenges they may experience in the classroom.

INSIGHTS Intervention Procedure. Participating schools were divided into three cohorts (Cohort 1: Fall 2008–Spring 2010; Cohort 2: Fall 2009–Spring 2011; Cohort 3: Fall 2010–Spring 2012) and agreed to implement the *INSIGHTS* or attention-control program in kindergarten and first grade. The *INSIGHTS* program was delivered to teachers and children concurrently by facilitators. The goal of the model was that after the end of the 10-week intervention, teachers would continue to implement the model for the duration of the academic year (and into future years). Teachers participated in sessions ten 2-hour sessions during the same 10-week period that children participated in ten 45-minute sessions with facilitators.

Students in study schools received the intervention across two consecutive years while enrolled in kindergarten and first grade. Kindergarten and first-grade teachers received the intervention for 10 weeks during one academic year, when students in the target cohort were enrolled in their classroom. Parents of consented students had an opportunity to participate in 10 weeks of parent intervention as well during their child's kindergarten or first grade year.

Facilitator training. *INSIGHTS* was delivered by a set of facilitators trained and supervised by the program developer. Prior to conducting the intervention in schools, all facilitators attended a graduate-level course to learn the theory and research underlying the *INSIGHTS* intervention. New facilitators were mentored by experienced facilitators and supervised by the developer. Each facilitator conducted the comprehensive intervention in the school to which she or he was assigned.

Program delivery: Focus on teachers. Teachers attended 10 weekly 2-hour sessions based on a structured curriculum that included didactic content and professionally produced vignettes as well as handouts and group activities. Teachers were given assignments based on program content to complete between sessions. When necessary, make-up sessions were offered. 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-minute lessons to all students enrolled in the classrooms of participating kindergarten and first-grade teachers. Teachers practiced lessons with any students who missed a session.

Fidelity. To maintain model 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 (see E. E. O'Connor et al., 2014).

Supervision was focused on challenges related to conducting sessions, implementation logistics, and participant concerns. Teacher sessions were also videotaped and reviewed for coverage of content and effectiveness of facilitation (Hulleman & Cordray, 2009). Fidelity coding revealed that 94% of the curriculum was adequately covered.

INSIGHTS dosage. Dosage was assessed using program facilitators' reports of number of sessions attended by teachers, children, and parents. Because we are leveraging data from all students who were enrolled in study schools, we do not have complete student- and parent-level dosage data for our full sample. However, we do have data for the subsample of students in these schools who were selected to participate in data collection activities and use that information as a way to estimate dosage across the broader group of student participants.

As noted above, kindergarten teachers were invited to participate in 10 *INSIGHTS* sessions during the year that the target cohort of children in the school were enrolled in kindergarten; first-grade teachers were invited to participate in 10 sessions in the year when that target cohort was in first grade. Children were invited to participate during both years. Dosage for teachers and children was high across schools. The average number of teacher sessions attended was 9.44 ($SD = 0.91$) across kindergarten and first grade. Most teachers attended all 10 sessions (70.6%) and another 26.5% attended eight or nine sessions. There were no teacher-level differences in participation by teacher grade. Children attended 8.30 sessions on average ($SD = 2.25$) in each grade. Thirty-two percentage of these children were present for all classroom sessions, and 46.3% were present for 8 or 9 sessions across both kindergarten and first grade.

Parents were invited to participate in 10 *INSIGHTS* sessions when their child was enrolled in either kindergarten or first grade. Parent participation was lower than participation in the teacher and child components of the program. Although both program facilitators and program staff (recruiters and field managers) made all possible efforts to engage parents in *INSIGHTS* sessions, the average number of sessions eligible parents attended across years was 5.93 ($SD = 4.15$). Twenty-five percent of the parents were present for all sessions and 38% were present for 8 sessions or more. This means that only about 10% of treatment group students in the current study—leveraging the larger data set—had parents participate in at least 80% of parent sessions. Follow-up work done by the team using causal modeling to understand the added value of the parent program showed that the effects of the program were similar regardless of whether the parents participated or not (see McCormick, Cappella, O'Connor, & McClowry, 2016). This study showed that the children who were more disadvantaged—and perhaps stood to gain the most from the program—were most likely to have parents who did not participate. These children still benefited substantially from the teacher and child program sessions even in the absence of the parent component.

Attention-Control Condition

A supplemental reading program served as the attention-control condition. The rationale for having an attention-control condition, rather than a business as usual condition, was to test whether it was truly the SEL programming provided by *INSIGHTS* that changed student outcomes, or simply having more attention and opportunities for professional development. In addition, the program provided the schools in low-income neighborhoods with additional literacy-related resources and allowed for a conservative estimate of intervention effects. There was no overlap in content covered in the supplemental reading program and the *INSIGHTS* intervention.

Teachers in attention-control schools attended two separate workshops, each 2 hours long, in which strategies to enhance literacy were presented and reading materials for the children were provided. Teachers received professional development credit and \$40 for classroom resources for each workshop. Reading program facilitators conducted the weekly reading sessions directly with students across the 10-week program period. Fidelity checklists revealed that 95% of topics were covered.

Appendix B

Missing Data Analysis

The total analytic sample size for this study is $N = 1,634$ students. Of the total number of students, 1,263 (77%) remained in the sample through the end of the fifth grade, while 371 (23%) are missing outcome data because they attrited from the school district before the end of the fifth grade. Chi-square tests indicated that those who left the school district by fifth grade were less likely to be eligible for free or reduced price lunch in both the treatment ($\chi^2 = 14.63, p < .01$) and attention-control ($\chi^2 = 16.14, p < .01$) groups. Children who left the district were also less likely to have attended a public preschool program in both the treatment ($\chi^2 = 33.28, p < .01$) and attention-control groups ($\chi^2 = 39.43, p < .01$). Children enrolled in *INSIGHTS* schools attrited at the same rate as children enrolled in attention-control schools. Within the nonattrited sample, there were $N = 639$ children in the attention-control group (76% of the baseline attention-control group) and $N = 624$ children in the *INSIGHTS* group (78% of the baseline treatment group). We used a cross-tab analysis with a chi-square test ($\chi^2 = 32.87, p = .56$) to determine that the rate of attrition was not statistically different across these two groups. According to the standards from the What Works Clearinghouse (2017), this study constitutes a low-attrition RCT, meaning that overall attrition and differential attrition indicate a tolerable level of bias for the estimated intervention effect. Given these moderately low levels of attrition and minimal differential attrition between the treatment and control groups, we decided

not to impute outcome data. The students in the treatment group are similar to the control group students that did not attrit and are included in the impact analysis. See Table 1 in the main text for further details on the nonattrited sample.

Covariate data had minimal levels of missingness given that these variables were collected when parents enrolled their children in public school. However, we did compare the students missing covariate data from those who had complete covariate data. We found that 14.45% of the sample with nonmissing outcome data was missing one or more covariates. Students missing covariate data were more likely to be White, to not have attended PreK, to be in the second cohort of the study, and to be in the control group. Using the assumption that data were Missing at Random, the team used a multiple data imputation method in STATA MICE and imputed 100 data sets in order to generate complete data on all covariates (Graham et al., 2007). Final parameter estimates were generated by calculating the mean across estimates from all 100 data sets. All predictive analyses presented in this article use the imputed data. A set of robustness checks demonstrated that the results from the multiple imputation procedure mirror the results using a listwise deletion approach with a smaller sample.

Appendix C

Model Specifications

The first impact model we examined in this study (Model 2 in Tables 3 and 4 in the main text) is as follows. This model includes only an intercept (β_1) and a Level-2 indicator variable for assignment to the *INSIGHTS* treatment (β_2). The model includes a random intercept for study school (ζ_j) and a residual error term (ϵ_{ij}).

$$\text{logit}\{\Pr(y_{ij} = 1 | \chi_{ij}, \zeta_j)\} = \beta_1 + \beta_2 \text{INSIGHTS}_j + \zeta_j + \epsilon_{ij}$$

Next, we added a series of Level-1 covariates to this model to improve the precision of the impact estimates. Results are illustrated in the panel labeled Model 3 in Tables 3 and 4.

$$\begin{aligned} \text{logit}\{\Pr(y_{ij} = 1 | \chi_{ij}, \zeta_j)\} = & \beta_1 + \beta_2 \text{INSIGHTS}_j + \\ & \beta_3 \text{BASELINECHILDAGE}_{ij} + \\ & \beta_4 \text{FEMALE}_{ij} + \beta_5 \text{BLACK}_{ij} + \\ & \beta_6 \text{HISPANIC}_{ij} + \beta_7 \text{LOW} - \text{SES}_{ij} + \\ & \beta_8 \text{DLL}_{ij} + \beta_9 \text{IMMIGRANT}_{ij} + \\ & \beta_{10} \text{PREKATTEND}_{ij} + \zeta_j + \epsilon_{ij} \end{aligned}$$

Finally, we added a series of Level-2 covariates to this model to further improve the precision of the impact estimates and adjust for any small group-based differences in school-level demographic characteristics. Results are illustrated in the panel labeled Model 4 in Tables 3 and 4.

$$\begin{aligned} \text{logit}\{\Pr(y_{ij} = 1 | \chi_{ij}, \zeta_j)\} = & \beta_1 + \beta_2 \text{INSIGHTS}_j + \\ & \beta_3 \text{BASELINECHILDAGE}_{ij} + \\ & \beta_4 \text{FEMALE}_{ij} + \beta_5 \text{BLACK}_{ij} + \\ & \beta_6 \text{HISPANIC}_{ij} + \beta_7 \text{LOW} - \text{SES}_{ij} + \\ & \beta_8 \text{DLL}_{ij} + \beta_9 \text{IMMIGRANT}_{ij} + \\ & \beta_{10} \text{PREKATTEND}_{ij} + \\ & \beta_{11} \text{COHORT2}_j + \\ & \beta_{12} \text{COHORT3}_j + \\ & \beta_{13} \text{SCHOOLSIZE}_j + \\ & \beta_{14} \% \text{ELACOMPETENT}_j + \\ & \beta_{15} \% \text{MATHCOMPETENT}_j + \\ & \beta_{16} \text{ATTENDANCE}_j + \zeta_j + \epsilon_{ij} \end{aligned}$$

Appendix D

Summary of Additional Robustness Checks

Findings from models using logistic regressions with clustered standard errors for study school are included in Appendix D Tables D1 and D2. As illustrated, impact findings appeared fairly similar using this model specification. However, the standard errors on the treatment coefficients and treatment \times low-income interaction terms are a bit larger in these models, relative to the random intercepts model specification. In addition, the main effect of *INSIGHTS* on receipt of SPED services was only marginally significant at $p = .08$, when we used this specification ($OR = 0.82$, $\log(\text{odds}) = -0.20$, $SE = 0.11$, $CI [-0.42, 0.02]$, $p = .08$). The moderated effect for low-income students was robust to this specification at the $p < .05$ level ($OR = 0.33$, $\log(\text{odds}) = -1.11$, $SE = 0.43$, $CI [-1.95, -0.26]$, $p = .03$).

Furthermore, given that we were missing 23% of outcome data, we conducted a bounding exercise in line with recommendations by Lee (2009) using the *Bounds* package in STATA to estimate worst-case bounds on regressions with some observations missing outcome and covariate data. This estimation calculates a confidence interval around an upper and lower bound of the impact estimate. Using this strategy, we found that the $\log(\text{odds})$ of the lower bound of the impact estimate on receipt of SPED services was -0.11 ($OR = 0.89$, $SE(\log(\text{odds})) = 0.06$, $CI [-0.23, 0.01]$). The $\log(\text{odds})$ odds of the upper bound of the impact estimate was $-.20$ ($OR = 0.82$, $SE[\log(\text{odds})] = 0.09$, $CI [-0.38, -0.02]$). As such, the lower bound was no longer statistically significant for the main effect of *INSIGHTS* on receipt of SPED services. Effects for the moderation findings were more consistent with a lower bound of -0.62 ($OR = 0.54$, $SE(\log(\text{odds})) = 0.28$, $CI [-1.18, -0.06]$) and an upper bound of -1.31 ($OR = 0.27$, $SE(\log(\text{odds})) = 0.44$, $CI [-2.19, 0.43]$), which were both statistically significant at the $p < .05$ level.

Next, we tested whether our results changed when we group-mean centered the Level-1 covariates in our analyses. Enders and Tofighi (2007) argue that centering Level-1

variables can improve statistical inference when one is using a two-level random effects model where the variable of interest—assignment to *INSIGHTS* in our case—is operationalized at Level 2. We found that our impact results for both grade retention and receipt of SPED services replicated when using this approach (see Appendix D Tables D3 and D4, respectively). The point estimates for the impacts were consistent and the standard errors and corresponding *p* values were actually a bit smaller in these models than in our original mixed-effects logistic regressions where we did not group-mean center the Level-1 covariates. We retain the uncentered versions of the Level-1 covariates in the presentation of our main set of results in order to maintain simple interpretation of effects.

In addition, we tested to see whether the results from our mixed-effects regression were sensitive to the choice of the covariance structure. As noted above, we found that our results were consistent regardless of whether we tested the mixed-effects regression using an unstructured, independent, exchangeable, or identity structure. Finally, we examined whether the results were consistent when we added school-level measures of the individual-level demographic covariates. We tested a final set of models where we further included covariates at the school-level for percent female, percent Black, percent Hispanic, percent low SES, percent dual language learner, % immigrant, and % students who attended PreK. Impact findings were all robust to the further inclusion of these school-level demographic characteristics.

TABLE D1

Impacts of INSIGHTS on Grade Retention (Logistic Regression With Clustered Standard Errors)

Fixed effects	Model 1			Model 2			Model 3			Model 4		
	OR	Log(odds)	Robust SE	OR	Log(odds)	Robust SE	OR	Log(odds)	Robust SE	OR	Log(odds)	Robust SE
Intercept	0.28	-1.29	0.13	0.32	-1.15	0.19	0.22	-1.52	0.41	0.22	-1.52	0.41
Treatment (<i>INSIGHTS</i>)	—	—	—	0.77	-0.26	0.27	0.79	-0.23	0.26	0.79	-0.24	0.24
Baseline age	—	—	—	—	—	—	0.51	-0.68**	0.16	1.97	0.68**	0.12
Female	—	—	—	—	—	—	0.84	-0.18	0.15	0.84	-0.18	0.16
Black	—	—	—	—	—	—	1.07	0.07	0.22	1.07	0.07	0.27
Hispanic	—	—	—	—	—	—	1.08	0.08	0.20	1.08	0.08	0.31
Low SES	—	—	—	—	—	—	2.41	0.88**	0.17	2.41	0.88**	0.18
Dual language learner	—	—	—	—	—	—	0.63	-0.46*	0.22	0.63	-0.46	0.29
From recent immigrant family	—	—	—	—	—	—	0.91	-0.09	0.29	0.91	-0.09	0.28
Prekindergarten attender	—	—	—	—	—	—	0.36	-1.01**	0.17	0.36	-1.01**	0.15
Cohort 2	—	—	—	—	—	—	—	—	—	1.43	0.36*	0.14
Cohort 3	—	—	—	—	—	—	—	—	—	1.19	0.17	0.21
School size	—	—	—	—	—	—	—	—	—	1.21	0.19	0.27
School % competent ELA test	—	—	—	—	—	—	—	—	—	0.84	-0.17*	0.09
School % competent math test	—	—	—	—	—	—	—	—	—	0.82	-0.20*	0.09
School average attendance	—	—	—	—	—	—	—	—	—	0.94	-0.06	0.11
Fit statistics												
AIC		362.17			361.54			349.64			342.54	
BIC		372.53			371.84			338.53			329.48	
Log likelihood		-184.35			-170.42			-169.04			-159.47	

Note. OR = odds ratio; SE = standard error; SES = socioeconomic status; AIC = Akaike information criterion; BIC = Bayesian information criterion.

p* < .10. *p* < .05. ****p* < .01.

TABLE D2

Impacts of INSIGHTS on Receipt of SPED Services (Logistic Regression With Clustered Standard Errors)

Fixed effects	Model 1			Model 2			Model 3			Model 4		
	OR	Log(odds)	Robust SE	OR	Log(odds)	Robust SE	OR	Log(odds)	Robust SE	OR	Log(odds)	Robust SE
Intercept	0.10	-2.29**	0.19	0.09	-2.38**	0.08	0.11	-2.19**	0.08	0.11	-2.18**	0.08
Treatment (<i>INSIGHTS</i>)	—	—	—	0.83	-0.19 ⁺	0.12	0.81	-0.21 ⁺	0.12	0.82	-0.20 ⁺	0.11
Baseline age	—	—	—	—	—	—	1.21	0.19	0.13	1.20	0.18	0.13
Female	—	—	—	—	—	—	0.29	-1.25**	0.22	0.29	-1.25**	0.25
Black	—	—	—	—	—	—	3.46	1.24**	0.33	3.53	1.26**	0.36
Hispanic	—	—	—	—	—	—	3.60	1.28**	0.47	3.39	1.22*	0.52

(continued)

TABLE D2 (CONTINUED)

Fixed effects	Model 1			Model 2			Model 3			Model 4		
	OR	Log(odds)	Robust SE	OR	Log(odds)	Robust SE	OR	Log(odds)	Robust SE	OR	Log(odds)	Robust SE
Low SES	—	—	—	—	—	—	1.75	0.56**	0.21	1.72	0.54*	0.25
Dual language learner	—	—	—	—	—	—	0.54	−0.61*	0.24	0.53	−0.63*	0.28
From recent immigrant family	—	—	—	—	—	—	0.16	−1.86*	0.91	0.16	−1.82*	0.89
Prekindergarten attender	—	—	—	—	—	—	0.90	−0.11	0.14	0.87	−0.14	0.13
Cohort 2	—	—	—	—	—	—	—	—	—	0.78	−0.25	0.25
Cohort 3	—	—	—	—	—	—	—	—	—	0.84	−0.17	0.31
School size	—	—	—	—	—	—	—	—	—	1.25	0.22*	0.09
School % competent ELA test	—	—	—	—	—	—	—	—	—	1.08	0.08	0.25
School % competent math test	—	—	—	—	—	—	—	—	—	0.90	−0.11	0.23
School average attendance	—	—	—	—	—	—	—	—	—	1.28	0.25	0.21
Fit statistics												
AIC		218.43			217.54			211.57			206.54	
BIC		226.43			210.54			204.65			176.44	
Log likelihood		−107.54			−101.04			−97.74			−96.44	

Note. OR = odds ratio; SE = standard error; SES = socioeconomic status; AIC = Akaike information criterion; BIC = Bayesian information criterion; SPED = special education.

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

TABLE D3

Impacts of INSIGHTS on Grade Retention (Group-Mean Centered L1 Covariates)

Fixed effects	OR	Log(odds)	SE
Intercept	0.24	−1.43**	0.22
Treatment (INSIGHTS)	0.77	−0.26	0.18
Student-level covariates			
Baseline age (centered)	0.61	−0.49**	0.09
Female (centered)	0.74	−0.30**	0.12
Black (centered)	1.12	0.11	0.35
Hispanic (centered)	1.38	0.32	0.49
Low SES (centered)	2.01	0.70**	0.31
Dual language learner (centered)	0.62	−0.48*	0.17
From recent immigrant family (centered)	0.74	−0.30	0.29
Prekindergarten attender (centered)	0.52	−0.65**	0.09
School-level covariates			
Cohort 2	1.22	0.20	0.25
Cohort 3	1.19	0.17	0.23
School size	1.36	0.31	0.28
School % competent ELA test	0.88	−0.13*	0.06
School % competent math test	0.82	−0.20**	0.05
School average attendance	0.86	−0.15	0.15
Variance components		Variance	SE
Baseline study school		0.05	0.04
Fit statistics			
AIC		319.57	
BIC		335.37	
Log likelihood		−159.47	

Note. OR = odds ratio; SE = standard error; SES = socioeconomic status; AIC = Akaike information criterion; BIC = Bayesian information criterion.

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

TABLE D4

Impacts of INSIGHTS on Receipt of SPED Services (Group-Mean Centered L1 Covariates)

Fixed effects	OR	Log(odds)	SE
Intercept	0.11	−2.21**	0.09
Treatment (<i>INSIGHTS</i>)	0.83	−0.19*	0.06
Student-level covariates			
Baseline age (centered)	1.09	0.09	0.25
Female (centered)	0.46	−0.78**	0.09
Black (centered)	0.14	−1.97*	0.91
Hispanic (centered)	0.33	−1.11	0.82
Low SES (centered)	0.27	−1.31*	0.56
Dual language learner (centered)	1.91	0.65	0.46
From recent immigrant family (centered)	0.61	−0.49	0.28
Prekindergarten attender (centered)	0.75	−0.29	0.18
School-level covariates			
Cohort 2	1.13	0.12	0.42
Cohort 3	0.12	−2.12**	0.55
School size	1.33	0.29*	0.12
School % competent ELA test	0.96	−0.04	0.22
School % competent math test	0.92	−0.08	0.32
School average attendance	1.09	0.09	0.27
Variance components		Variance	SE
Study school		0.04*	0.02
Fit statistics			
AIC		166.75	
BIC		182.57	
Log likelihood		−94.54	

Note. OR = odds ratio; SE = standard error; SES = socioeconomic status; AIC = Akaike information criterion; BIC = Bayesian information criterion.

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

Acknowledgments

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

ORCID iD

Hope S. White  <https://orcid.org/0000-0003-0915-7808>

Notes

1. We only include children enrolled in study schools at the time of random assignment in this analysis to maintain the internal validity of the design. These students were exposed to the intervention or comparison condition in kindergarten and first grade, assuming they remained in the school for both consecutive years. However, there were students who are not included in the study sample who did receive the intervention in first grade. We do not include those students in analyses because they enrolled in schools after random assignment.

2. Earlier work from the *INSIGHTS* randomized control trial aimed to use a consented sample to collect data on students and

teachers and estimate the short-term impacts of the program on a host of outcomes at multiple levels (see E. E. O'Connor et al., 2014). That work focused on a subset of students participating in the intervention whose parent consented for them to take part in data collection activities. We limited the number of students enrolled in the study in each classroom (4–10 students per classroom was the average) in order to reduce burden on teachers for direct assessments and teacher reports. The consented sample for the study does not differ from the sample enrolled in the current study, which uses de-identified and thus more complete data.

3. We did aim to enroll schools in the study who had at least 75% of students eligible for free or reduced-price lunch. However, at the time that this study was being conducted in Brooklyn, NY, we observed a massive wave of gentrification where higher income families were moving into traditionally lower income neighborhoods and enrolling their children in neighborhood schools. Accordingly, although we originally sought to conduct the study only within lower income schools, the student sample has sufficient variation in order to consider subgroup effects for children who were and were not eligible for free or reduced-price lunch. Many of the schools that the study was originally conducted in did transition from being made up of primarily free lunch eligible students in 2008–2010 to primarily non-free lunch eligible by 2018. Although

this gentrification process has raised a number of challenges more generally for neighborhoods and schools in Brooklyn, it has allowed us to consider variation in effects across a more diverse group than we otherwise would have been able to do.

References

- Bailey, D., Duncan, G. J., Odgers, C. L., & Yu, W. (2017). Persistence and fadeout in the impacts of child and adolescent interventions. *Journal of Research on Educational Effectiveness*, 10, 7–39.
- Belsky, J., Bakermans-Kranenburg, M. J., & Van Ijzendoorn, M. H. (2007). For better and for worse: Differential susceptibility to environmental influences. *Current Directions in Psychological Science*, 16, 300–304.
- Bettencourt, A., Gross, D., & Ho, G. (2016). *The costly consequences of not being socially and behaviorally ready by kindergarten: Associations with grade retention, receipt of academic support services, and suspensions/expulsions*. Baltimore, MD: Baltimore Education Research Consortium.
- 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, 156–168.
- Blair, C. (2001). The early identification of risk for grade retention among African American children at risk for school difficulty. *Applied Developmental Science*, 5, 37–50.
- 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, 218–224.
- Bradshaw, C. P., Zmuda, J. H., Kellam, S. G., & Ialongo, N. S. (2009). Longitudinal impact of two universal preventive interventions in first grade on educational outcomes in high school. *Journal of Educational Psychology*, 101, 926–937.
- Carlson, E., Daley, T., Bitterman, A., Heinzen, H., Keller, B., Markowitz, J., & Riley, J. (2009). *Early school transitions and the social behavior of children with disabilities: Selected findings from the pre-elementary education longitudinal study*. Rockville, MD: Westat.
- Cicchetti, D., & Aber, J. L. (1998). Contextualism and developmental psychopathology. *Development and Psychopathology*, 10, 137–141.
- Collaborative for Academic, Social, and Emotional Learning. (2012). *2013 CASEL guide: Effective social and emotional learning programs* (Preschool and elementary school ed.). Chicago, IL: Author.
- Coutinho, M. J., & Oswald, D. P. (2000). Disproportionate representation in special education: A synthesis and recommendations. *Journal of Child and Family Studies*, 9, 135–156.
- Cristea, I. A., & Ioannidis, J. P. (2018). *P* values in display items are ubiquitous and almost invariably significant: A survey of top science journals. *PLoS One*, 13, e0197440.
- Davoudzadeh, P., McTernan, M. L., & Grimm, K. J. (2015). Early school readiness predictors of grade retention from kindergarten through eighth grade: A multilevel discrete-time survival analysis approach. *Early Childhood Research Quarterly*, 32, 183–192.
- Denissen, J. J. A., van Aken, M. A. G., Penke, L., & Wood, D. (2013). Self-regulation underlies temperament and personality: An integrative developmental framework. *Child Development Perspectives*, 7, 255–260.
- Diamond, A., & Lee, K. (2011). Interventions shown to aid executive function development in children 4 to 12 years old. *Science*, 333, 959–964.
- Duncan, R., Washburn, I. J., Lewis, K. M., Bavarian, N., DuBois, D. L., Acock, A. C., . . . Flay, B. R. (2017). Can universal SEL programs benefit universally? Effects of Positive Action program on multiple trajectories of social-emotional and misconduct behaviors. *Prevention Science*, 18, 214–224.
- Durlak, J. A., Weissberg, R. P., Dymnicki, A. B., Taylor, R. D., & Schellinger, K. B. (2011). The impact of enhancing students' social and emotional learning: A meta-analysis of school-based universal interventions. *Child Development*, 82, 405–432.
- Eddy, M., Reid, J. B., Stoolmiller, M., & Fetrow, A. (2003). Outcomes during middle school for an elementary school-based preventive intervention for conduct problems: Follow-up results from a randomized trial. *Behavior Therapy*, 34, 535–552.
- Eide, E. R., & Goldhaber, D. D. (2005). Grade retention: What are the costs and benefits? *Journal of Education Finance*, 31, 195–214.
- Eisenberg, N., Valiente, C., & Eggum, N. D. (2010). Self-regulation and school readiness. *Early Education and Development*, 21, 681–698.
- Elias, M. J., Zins, J. E., Weissberg, R. P., Frey, K. S., Greenberg, M. T., Haynes, N. M., . . . Shriver, T. P. (1997). *Promoting social and emotional learning: Guidelines for educators*. Alexandria, VA: ASCD.
- Enders, C. K., & Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: A new look at an old issue. *Psychological Methods*, 12, 121–138.
- Felner, R. D., Brand, S., Adan, A. M., Mulhall, P. F., Flowers, N., Sartain, B., & DuBois, D. L. (1993). Restructuring the ecology of the school as an approach to prevention during school transitions: Longitudinal follow-ups of the School Transitional Environment Project (STEP). *Prevention in Human Services*, 10, 103–136.
- Fitzmaurice, G. M., Laird, N. M., & Ware, J. H. (2004). *Applied longitudinal analysis*. Hoboken, NJ: Wiley.
- Fortner, C. K., & Jenkins, J. M. (2017). Kindergarten redshirt-ing: Motivations and spillovers using census-level data. *Early Childhood Research Quarterly*, 38, 44–56.
- Graham, J. W., Olchowski, A. E., & Gilreath, T. D. (2007). How many imputations are really needed? Some practical clarifications of multiple imputation theory. *Prevention Science*, 8, 206–213.
- Harry, B., & Klingner, J. (2014). *Why are so many minority students in special education? Understanding race and disability in schools* (2nd ed.). New York, NY: Teachers College Press.
- Hawkins, J. D., Catalano, R. F., Kosterman, R., Abbott, R., & Hill, K. G. (1999). Preventing adolescent health-risk behaviors by strengthening protection during childhood. *Archives of Pediatrics & Adolescent Medicine*, 153, 226–234.
- Hawkins, J. D., Kosterman, R., Catalano, R. F., Hill, K. G., & Abbot, R. D. (2008). Effects of social development intervention in childhood 15 years later. *Archives of Pediatrics & Adolescent Medicine*, 162, 1133–1141.

- Hawkins, J. D., Oesterle, S., Brown, E. C., Arthur, M. W., Abbot, R. D., Fagan, A. A., & Catalano, R. F. (2009). Results of a type 2 translational research trial to prevent adolescent drug use and delinquency: A test of Communities That Care. *Archives of Pediatrics & Adolescent Medicine*, 163, 789–798.
- Heckman, J. J. (2006). Skill formation and the economics of investing in disadvantaged children. *Science*, 312, 1900–1902.
- Hill, K. G., Bailey, J. A., Hawkins, J. D., Catalano, R. F., Kosterman, R., Oesterle, S., & Abbott, R. D. (2014). The onset of STI diagnosis through age 30: Results from the Seattle Social Development Project intervention. *Prevention Science*, 15(Suppl. 1), S19–S32.
- Hong, G., & Yu, B. (2007). Early-grade retention and children's reading and math learning in elementary years. *Educational Evaluation and Policy Analysis*, 29, 239–261.
- Hughes, J. N., West, S. G., Kim, H., & Bauer, S. S. (2018). Effect of early grade retention on school completion: A prospective study. *Journal of Educational Psychology*, 110, 974–991.
- Hulleman, C. S., & Cordray, D. S. (2009). Moving from the lab to the field: The role of fidelity and achieved relative intervention strength. *Journal of Research on Educational Effectiveness*, 2, 88–110.
- Hwang, S. H., & Cappella, E. (2018). Rethinking early elementary grade retention: Examining long-term academic and psychosocial outcomes. *Journal of Research on Educational Effectiveness*, 11(4), 559–587.
- Ialongo, N., Poduska, J., Werthamer, L., & Kellam, S. (2001). The distal impact of two-first-grade preventive interventions on conduct problems and disorder in early adolescence. *Journal of Emotional and Behavioral Disorders*, 9, 146–161.
- Jones, S. M., Brown, J. L., & Lawrence Aber, J. (2011). Two-year impacts of a universal school-based social-emotional and literacy intervention: An experiment in translational developmental research. *Child Development*, 82, 533–554.
- Jones, S. M., & Doolittle, E. J. (2017). Social and emotional learning: Introducing the issue. *Future of Children*, 27(1), 3–11.
- 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(Suppl. 1), S5–S28.
- Lee, D. S. (2009). Training, wages, and sample selection: Estimating sharp bounds on treatment effects. *Review of Economic Studies*, 76, 1071–1102.
- Levine, R. (2016). Turning 5: Helping families of preschoolers with disabilities navigate the transition to kindergarten in New York City. *Voices in Urban Education*, 43, 27–36.
- Low, S., Smolkowski, K., & Cook, C. (2016). What constitutes high-quality implementation of SEL programs? A latent class analysis of Second Step® implementation. *Prevention Science*, 17, 981–991.
- Magnuson, K., & Duncan, G. J. (2016). Can early childhood interventions decrease inequality of economic opportunity? *Russell Sage Foundation Journal of the Social Sciences: RSF*, 2, 123–141.
- Mariano, L. T., Martorell, P., & Tsai, T. (2018). *The effects of grade retention on high school outcomes: Evidence from New York City schools* (RAND Working Paper No. 1259). Santa Monica, CA: RAND.
- Masten, A. S., & Cicchetti, D. (2010). Developmental cascades. *Development and Psychopathology*, 22, 491–495.
- Masten, A. S., Long, J. D., Kuo, S. I.-C., McCormick, C. M., & Desjardins, C. D. (2009). Developmental models of strategic intervention. *European Journal of Developmental Science*, 3, 282–291.
- Mattison, A., Raffaele Mendez, L. M., Dedrick, R., Dickinson, S., Wingate, E., & Hanks, C. (2018). Early elementary teacher ratings of behavior as predictors of grade retention: Race, gender, and socioeconomic status as potential moderators. *Psychology in the Schools*, 55, 1171–1187.
- McClowry, S. G. (2014). *Temperament-based elementary classroom management*. Lanham, MD: Rowman & Littlefield.
- 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 in kindergarten and first grade. *Early Childhood Research Quarterly*, 30, 128–139.
- McCormick, M. P., Cappella, E., O'Connor, E. E., Hill, J., & McClowry, S. G. (2016). Does intervention for parents add value to social-emotional learning programs? Evidence from the randomized trial of INSIGHTS. *Journal of Research on Educational Effectiveness*, 9(3), 364–394.
- McCormick, M. P., O'Connor, E. E., Cappella, E., & McClowry, S. G. (2019). *Long-term effects of INSIGHTS on academic achievement: Do impacts vary by school readiness?* Paper presented at the annual Institute of Education Sciences Principal Investigators' meeting. Washington, D.C.
- McIntyre, L. L., Eckert, T. L., Fiese, B. H., Reed, F. D., & Wildenger, L. K. (2010). Family concerns surrounding kindergarten transition: A comparison of students in special and general education. *Early Childhood Education Journal*, 38, 259–263.
- Morgan, P. L., Farkas, G., Hillemeier, M. M., Mattison, R., Maczuga, S., Li, H., & Cook, M. (2015). Minorities are disproportionately underrepresented in special education: Longitudinal evidence across five disability conditions. *Educational Researcher*, 44, 278–292.
- Morgan, P. L., & Sideridis, G. D. (2006). Contrasting the effectiveness of fluency interventions for students with or at risk for learning disabilities: A multilevel random coefficient modeling meta-analysis. *Learning Disabilities: Research and Practice*, 21, 191–210.
- O'Connor, C., & Fernandez, S. D. (2006). Race, class, and disproportionality: Reevaluating the relationship between poverty and special education placement. *Educational Researcher*, 35(6), 6–11.
- O'Connor, E. E., Cappella, E., McCormick, M. P., & McClowry, S. G. (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, 1156–1169. doi:10.1037/a0036615
- Qi, C. H., & Kaiser, A. P. (2003). Behavior problems of preschool children from low-income families: Review of the literature. *Topics in Early Childhood Special Education*, 23, 188–216. doi:10.1177/02711214030230040201
- Reynolds, A. J., & Temple, J. A. (2006). Economic benefits of investments in preschool education. In E. Zigler, W. Gilliam,

- & S. Jones (Eds.), *A vision for universal prekindergarten* (pp. 37–68). New York, NY: Cambridge University Press.
- Riggs, N. R., & Pentz, M. A. (2009). Long term effects of adolescent marijuana use prevention on adult mental health services utilization: The Midwestern prevention project. *Substance Use & Misuse*, 44, 616–631.
- Rimm-Kaufman, S. E., Larsen, R. A., Baroody, A. E., Curby, T. W., Ko, M., Thomas, J. B., . . . DeCoster, J. (2014). Efficacy of the Responsive Classroom approach: Results from a 3-year, longitudinal randomized controlled trial. *American Educational Research Journal*, 51, 567–603.
- Rimm-Kaufman, S. E., Pianta, R. C., & Cox, M. J. (2000). Teachers' judgments of problems in the transition to kindergarten. *Early Childhood Research Quarterly*, 15, 147–166.
- Roodman, D., Nielsen, M. Ø., MacKinnon, J. G., & Webb, M. D. (2019). Fast and wild: Bootstrap inference in Stata using boottest. *Stata Journal*, 19, 4–60.
- Rothbart, M. K., Sheese, B. E., Rueda, M. R., & Posner, M. I. (2011). Developing mechanisms of self-regulation in early life. *Emotion Review*, 3, 207–213.
- Scott, M. A., Simonoff, J. S., & Marx, B. D. (Eds.). (2013). *The Sage handbook of multilevel modeling*. Thousand Oaks, CA: Sage.
- Shavelson, R. J., & Towne, L. (Eds.). (2002). *Scientific research in education* (Committee on Scientific Principles for Education Research, Center for Education, Division of Behavioral and Social Sciences and Education; National Research Council). Washington, DC: National Academies Press.
- Snyder, T. D., de Brey, C., & Dillow, S. A. (2016). *Digest of Education Statistics 2015* (NCES 2016-014). Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.
- 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: U.S. Department of Education.
- Stroebe, W., & Strack, F. (2014). The alleged crisis and the illusion of exact replication. *Perspectives on Psychological Science*, 9, 59–71. doi:10.1177/1745691613514450
- Taylor, R. D., Oberle, E., Durlak, J. A., & Weissberg, R. P. (2017). Promoting positive youth development through school-based social and emotional learning interventions: A meta-analysis of follow-up effects. *Child Development*, 88, 1156–1171.
- Thomas, A., & Chess, S. (1977). *Temperament and development*. New York, NY: Brunner/Mazel.
- Thompson, R. A., & Raikes, H. A. (2007). The social and emotional foundations of school readiness. In D. F. Perry, R. K. Kaufmann, & J. Knitzer (Eds.), *Social and emotional health in early childhood: Building bridges between services and systems* (pp. 13–35). Baltimore, MD: Paul H. Brookes.
- U.S. Department of Education, National Center for Education Statistics. (2017). *The condition of education 2017* (NCES 2017-144, Public School Expenditures). Washington, DC: Author.
- Warren, J. R., & Saliba, J. (2012). First-through eighth-grade retention rates for all 50 states: A new method and initial results. *Educational Researcher*, 41, 320–329.
- What Works Clearinghouse. (2017, October). *Standards Handbook (Version 4.0)*. Retrieved from <http://whatworks.ed.gov>
- Webster-Stratton, C., Jamila 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, 471–488.
- Wu, W., West, S. G., & Hughes, J. N. (2010). Effect of grade retention in first grade on psychosocial outcomes. *Journal of Educational Psychology*, 102, 135–152.
- Xia, N., & Kirby, S. (2009). *Retaining students in grade: A literature review of the effects of retention on students' academic and nonacademic outcomes (RAND Report)*. Santa Monica, CA: RAND.

Authors

MEGHAN P. MCCORMICK is a research associate at MDRC. She received a PhD in applied psychology from New York University.

ROBIN NEUHAUS is a doctoral candidate in teaching and learning at New York University.

E. PARHAM HORN is a doctoral candidate in counseling psychology at New York University.

ERIN E. O'CONNOR is a professor of teaching and learning at New York University. She received an EdD in human development from the Harvard Graduate School of Education.

HOPE S. WHITE is a doctoral candidate in clinical psychology at SUNY Buffalo.

SAMANTHA HARDING is a master's candidate at New York University.

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

SANDEE MCCLOWRY is a professor emeritus at New York University. She received a PhD in nursing from the University of California, San Francisco.