Child Development, xxxxx 2011, Volume 00, Number 0, Pages 1–22

Two-Year Impacts of a Universal School-Based Social-Emotional and Literacy Intervention: An Experiment in Translational Developmental Research

Stephanie M. Jones
Harvard University

Joshua L. Brown
Fordham University

J. Lawrence Aber
New York University

This study contributes to ongoing scholarship at the nexus of translational research, education reform, and the developmental and prevention sciences. It reports 2-year experimental impacts of a universal, integrated school-based intervention in social-emotional learning and literacy development on children's social-emotional, behavioral, and academic functioning. The study employed a school-randomized, experimental design with 1,184 children in 18 elementary schools. Children in the intervention schools showed improvements across several domains: self-reports of hostile attributional bias, aggressive interpersonal negotiation strategies, and depression, and teacher reports of attention skills, and aggressive and socially competent behavior. In addition, there were effects of the intervention on children's math and reading achievement for those identified by teachers at baseline at highest behavioral risk. These findings are interpreted in light of developmental cascades theory and lend support to the value of universal, integrated interventions in the elementary school period for promoting children's social-emotional and academic skills.

Over the last two decades, developmental science has made significant progress in understanding children's trajectories toward social-emotional and academic outcomes (e.g., Arnold, 1997; Blair, 2002; Duncan et al., 2007; Miles & Stipek, 2006). At the same time, there has been dramatic growth in the design, implementation, and rigorous evaluation of school-based interventions to promote positive social-emotional development and/or academic achievement (e.g., Brown, Jones, LaRusso, & Aber, 2010; Embry, 2002; Jones, Brown, Hoglund, & Aber, under review; Kellam et al., 2008; Payton et al., 2008; Prothrow-Stith, 2007; Webster-Stratton, Reid, & Stoolmiller, 2008). This present study contributes to ongoing scholarship in the school-based prevention of social-emotional, behavioral, and academic problems by reporting experimental impacts of a novel social-emotional learning and literacy intervention (the 4Rs Program, ''Reading, Writing, Respect and Resolution'') on a cohort of third grade children's social-emotional, behavioral, and academic functioning after two consecutive years of exposure to the intervention.

School-Based Social-Emotional Learning and Violence Prevention Interventions: Theory and Research

The past decade has seen dramatic growth in programs focused on enhancing children's social and emotional skills to reduce aggression and

Second revision submitted to *Child Development* on January 5, 2011.

The original research and analyses presented in this article were supported by grants from IES (# R305LO30003) and W. T. Grant Foundation (#1656) to Lawrence Aber (PI) and from W. T. Grant Foundation (#7520) and NIMH (#1R01MH082085-01A2) to Joshua Brown and Stephanie Jones (PIs). The authors are very grateful to Tom Roderick and his entire staff at the Morningside Center for their talent at and commitment to developing and implementing the 4Rs (Reading, Writing, Respect, and Resolution) and to a rigorous external evaluation of the 4Rs. The authors are also enormously grateful to the many postdoctoral fellows, doctoral students, and research assistants at New York University, Fordham University, and Harvard University who have helped in every phase of the work reported here, from data collection through data analysis. Finally, we are grateful to the many thousands of students, hundreds of teachers, and dozens of schools who have made this work possible.

Correspondence concerning this article should be addressed to Stephanie M. Jones, Harvard Graduate School of Education, 14 Appian Way, Cambridge, MA 02138. Electronic mail may be sent to jonesst@gse.harvard.edu.

 0009-3920/2011/xxxx-xxxx
DOI: 10.1111/j.1467-8624.2010.01560.x

violence and to promote positive interactions (Clayton, Ballif-Spanvill, & Hunsaker, 2001; Payton et al., 2008; Wilson, Lipsey, & Derzon, 2003). This period has also witnessed a growing convergence of developmental science and prevention science in guiding the design and evaluation of interventions aimed at preventing future aggressive and violent behavior in children and youth (Institute of Medicine, 1994; Maggs & Schulenberg, 2001). From developmental science, knowledge has grown about the mechanisms by which exposure to violence affects children's risk for such outcomes (Coie & Dodge, 1998; Dodge, 2006). Our greater understanding of these causal mechanisms has led to improvements in both the design and evaluation of preventive interventions, which increasingly target these mechanisms as their focus of change (e.g., Aber, Brown, Chaudry, Jones, & Samples, 1996; Dodge, 2001; Hudley & Graham, 1993). Indeed, several of these causal mechanisms are central to the 4Rs Program and to the evaluation design of this study (see below). Developmental science is also the source of valuable theoretical perspectives that are aligned with our expectations of how such interventions work in school settings and how their effects are anticipated to unfold over time. For example, our approach to the evaluation of the 4Rs is guided by developmental cascades theory (Masten & Cicchetti, 2010; Masten, Long, Kuo, McCormick, & Desjardins, 2009) which emphasizes an ecological, multi-level lens, focuses attention on change processes in multiple domains, and considers child development in context as dynamic systems (Jones, Brown, & Aber, 2010).

From prevention science, knowledge has grown about the effectiveness of these intervention strategies at reducing children's risk for future aggressive and violent behavior (Aber, Brown, & Jones, 2003; Clayton et al., 2001; Conduct Problems Prevention Research Group (CPPRG), 1999, 2007; Embry, 2002; Prothrow-Stith, 2007; Webster-Stratton & Taylor, 2001). Although the literature on school-based preventive interventions is rich with studies of interventions targeted at subgroups of high-risk children (Vitaro, Brendgen, & Tremblay, 2001), with a few exceptions (e.g., the Good Behavior Game; Kellam et al., 2008), it has only recently expanded in the area of large-scale evaluations of universal interventions implemented with general populations of students (Durlak, 1995; IES, 2003; Payton et al., 2008). Furthermore, among the best studies of universal school-based interventions designed to reduce risk for future aggression and violence, a variety of methodological challenges limit the quality and generalizability of the knowledge base (Hundert et al., 1999). With some exceptions (e.g., CPPRG, 2007; Ialongo et al., 1999; Kellam et al., 1998), these earlier studies rarely employ both experimental random assignment designs and appropriate analytic methods (e.g., multilevel modeling with intervention status modeled at the level of random assignment). Similarly, studies typically fail to employ multi-year, longitudinal designs and to examine changes in both short- and long-term developmental outcomes as they unfold over time.

Integrated Approaches

Historically, school-based interventions have been focused on either academic or social-emotional outcomes. However, consistent with developmental cascades theory and other recent research (e.g., Blair, 2002), and considering the limited time in schools for outside interventions integrating academic and social-emotional learning intervention efforts, and of examining the direct effects of such integrated intervention on both domains of development (e.g., Bierman et al., 2008; Durlak, Weissberg, Dymnicki, Taylor, & Schellinger, 2011; Ialongo et al., 1999). Efforts at such integration are supported by a mounting body of research linking successful school adjustment and performance with the development of social-emotional competence (e.g., Miles & Stipek, 2006; Pianta, Steinberg, & Rollins, 1995; Rogoff, 1990; Wentzel, 1991). At the elementary school level, learning involves successful relationships with teachers and peers requiring skills in problem-solving and conflict resolution (Chen, Rubin, & Li, 1997; Maughan, Pickles, Hagell, Rutter, & Yule, 1996). Research also suggests that children who have difficulty regulating emotions and therefore experience high levels of negative emotional arousal may have trouble concentrating in class and recalling things they have learned (Raver, Garner, & Smith-Donald, 2007). Finally, there is a great deal of evidence that poor academic achievement co-occurs with disruptive and aggressive behavior (DeBaryshe, Patterson, & Capaldi, 1993; Masten et al., 1995; Ollendick, Weist, Borden, & Greene, 1992) and is a predictor of future maladaptive behavior (Kellam, Brown, Rubin, & Ensminger, 1983; Williams & McGee, 1994). Indeed, prior reviews have proposed that shared or correlated risks underlie the development of poor social-emotional skills, aggressive behavior, and academic problems (Hinshaw, 1992; Olweus, 1983).

Recently, interest has grown in developing and testing integrated approaches to building social-emotional and literacy skills both in the preschool period (e.g., the eight studies funded as part of the Interagency Consortium on School Readiness; Bierman et al., 2008; Raver et al., 2009), and in the elementary school years (e.g., Shinn, Walker, & Stoner, 2002; Zins, Weissberg, Wang, & Walberg, 2004). Several recent reviews of school-based preventive interventions targeting social-emotional or academic skills suggest that the ''prevention strategies that directly target one of these problem areas tend to have indirect, positive effects on the other target area'' (Walker & Shinn, 2002, p. 3). For example, in our previous work, we found that exposure to a social-emotional learning and conflict resolution curriculum for 2 years was associated with improvements both in social-emotional learning (Aber et al., 2003) and in scores on standardized tests of math achievement (Brown, Roderick, Lantieri, & Aber, 2004). Other work has shown similar cross-over effects, primarily in the case of impacts in the academic domain that are linked to exposure to a school-based social-emotional learning intervention (e.g., Durlak et al., 2011; Flay & Allred, 2003; Flay, Allred, & Ordway, 2001). Kellam et al. (e.g., Ialongo et al., 1999; Kellam, Rebok, Mayer, Ialongo, & Kalodner, 1994) reported the effects of two independent universal interventions (Good Behavior Game and Mastery Learning) implemented with first grade children, followed by reports in which the two interventions were integrated as they are in the 4Rs Program, with the integrated program resulting in larger effects in each domain than seen previously (Ialongo et al., 1999).

The 4Rs Program Theory of Change

The 4Rs Program is a universal, school-based intervention for grades K-5 focused on social-emotional learning and literacy development that provides a pedagogical link between the teaching of social-emotional competencies and fundamental academic skills, capitalizing on their mutual influence on successful youth development (Hinshaw, 1992). (See Method for a full description of the program.) Building off its predecessor, the Resolving Conflict Creatively Program (RCCP; Aber et al., 1996; Aber, Jones, Brown, Chaudry, & Samples, 1998; Aber et al., 2003), and consistent with both developmental cascades theory (Masten & Cicchetti, 2010) and current notions of translational research, the 4Rs Program can be characterized as a ''developmental intervention'' (National Research Council, 1993), one that focuses on changing underlying processes that lead to aggression and violence when they are still in the formative stage. The theoretical model underlying the social-emotional learning core of the 4Rs Program emphasizes the social-cognitive and social-emotional processes prior research has shown link individual, family, school, and community risk factors to the development of aggressive behavior, and that place children at higher risk for future violence (Coie & Dodge, 1998; Dodge, 2006; Gershoff & Aber, 2006; Huesmann & Guerra, 1997).

The key processes in the social-cognitive domain include: the scope and intensity of children's attributional biases toward interpreting ambiguous social cues as hostile rather than prosocial (e.g., Dodge, Bates, & Pettit, 1990), their normative beliefs about aggression (Huesmann & Guerra, 1997), and the developmental maturity of children's interpersonal negotiation strategies with peers (e.g., Selman, Beardslee, Schultz, Krupa, & Podoresky, 1986). Theory, basic, and applied research suggest that each of these social-cognitive processes is: (a) affected by certain types of experiences (e.g., a history of harsh, punitive, or abusive parenting, exposure to community violence; Dodge, 2006) or a peer environment in which violence is normative (Bierman & Wargo, 1995; Tremblay, Pagani-Kurtz, Masse, Vitaro, & Pihl, 1995), and in turn; (b) increases the probability of aggression and violence by children and youth (Dodge, Laird, Lochman, Zelli, and Conduct Problems Prevention Research Group, 2002; Guerra & Slaby, 1990; Huesmann & Guerra, 1997). These social-cognitive processes constitute a potential causal mechanism linking exposure to ecological risk with later developmental outcomes of aggression and violence. While the body of evidence supporting the linkages described here is largely correlational (Dodge, 2006), basic principles of translational research and prevention science (e.g., Kellam & Langevin, 2003) suggest that if experimental evaluations of interventions that target theoretically and empirically supported causal mechanisms reveal changes in both the proximal mechanisms (e.g., social-cognitive processes) as well as the distal target (aggressive behavior), additional support for the basic developmental theory is provided.

In our earlier work evaluating the RCCP (Aber et al., 1998, 2003), we examined children's trajectories of aggression and violence over the course of elementary school and estimated the effects of participation in RCCP on changes in these trajectories. In contrast to the present experimental

evaluation of the 4Rs Program, the design of the RCCP evaluation was quasi-experimental. We found that children's exposure to teachers who taught more of the SEL curriculum (the core of both RCCP and 4Rs) slowed children's rates of growth in aggression-related cognitions, including their hostile attributional biases (HAB) and aggressive interpersonal negotiation strategies (AINS) over the course of elementary school (Aber et al., 2003).

The key processes in the social-emotional domain include children's aggressive and prosocial fantasies, and their levels of affective symptomatology, including symptoms of depression and attention deficit hyperactivity disorder (ADHD). Child reports of affective symptomatology have been used extensively in prior evaluations of preventive interventions (CPPRG, 1999, 2002, 2004, 2007; Dalberg, Toal & Behrens, 1996; Kellam et al., 1994). Moreover, because of the fear, anxiety, and sadness caused by environments characterized by high concentrations of children who misperceive social interactions and cues, or who hold beliefs that aggression is a normative response to problem-solving situations, the 4Rs Program is expected to reduce children's depressive symptomatology and build their capacity to focus attention. Our earlier work found significant impacts of exposure to RCCP on children's depressive symptoms and aggressive fantasies (Aber et al., 2003).

The literature briefly reviewed here, our prior work, and developmental cascades theory (Jones et al., 2010) inform both short- and longer term expectations for 4Rs Program effects on children's outcomes. Our short-term expectations after 1 year were for program impacts primarily in the social-cognitive and social-emotional domains, and this was partly confirmed (see Jones et al., 2010). The approach of embedding social-emotional learning and conflict resolution lessons in a balanced literacy delivery strategy, and research tying together the social-emotional and academic domains, support our expectation for longer term effects on behavior and academic achievement. Thus, we directed our focus to constructs in four primary domains that either represent direct proximal targets of the intervention, or are known predictors of children's future functioning: children's Social-Cognitive Processes, their Social-Emotional Symptomatology, their Aggressive and Prosocial Behavior, and their Academic Functioning. Results after 1 year of exposure to the 4Rs Program indicate main effects of the intervention on children's social cognitions (i.e., self-reports of HAB [ES = .17]), and social-emotional symptomatology (self-reports of depression [ES = .17]). In addition, there were intervention impacts on children's self-reports of aggressive fantasies, teacher reports of academic skills, and attendance records for those children identified by teachers at baseline at highest behavioral risk (ES = .32–.59; Jones et al., 2010).

This study is the first report of experimental effects of the 4Rs Program on change over 2 years using four repeated assessments of children's social-cognitive, social-emotional, behavioral, and academic outcomes. The primary questions addressed in the article are: (a) What is the experimental impact of the 4Rs Program on 2-year change in third grade children's social-cognitive processes, social-emotional symptomatology, aggressive and socially competent behavior, and academic functioning, controlling for key demographic covariates? (b) Building on recent findings demonstrating significantly stronger intervention impacts for families facing more versus fewer poverty-related risks (e.g., Tolan, Gorman-Smith, & Henry, 2004), is the 2-year impact of the 4Rs Program moderated by child-level demographic baseline covariates including child gender, race/ethnic background, family socioeconomic risk, and community risk? (c) Finally, building on evidence from the 1st year of our 4Rs evaluation (Jones et al., 2010) and other experimental evaluations of universal school-based prevention programs (CPPRG, 2007; Multisite Violence Prevention Project, 2009), is the 2-year impact of the 4Rs Program moderated by children's baseline behavioral risk?

Method

Participants were 1,184 children (49% boys; average age at time 1 = 8.17 years, *SD* = 0.7), and 146 teachers (88% female; average age = 35) in 18 public New York City (NYC) inner-city elementary schools. (This study is one of seven sites participating in the national Social and Character Development Research Program. The report is limited to the set of processes and outcomes most central to the theory of change of the 4Rs Program, specifically. The impact of the 4Rs Program on the broader set of outcomes addressed by the national Social and Character Development Research Program will be presented in future reports.) The children and teachers are part of an ongoing, longitudinal evaluation of a universal, school-wide literacy and social-emotional learning intervention program (4Rs: Reading, Writing, Respect and Resolution) implemented for three consecutive

years in nine intervention (n = 630; 53.2%) and nine control (n = 554; 46.8%) schools. As noted earlier, in this article, we describe impacts of 4Rs on children's developmental trajectories across four repeated time points, and two consecutive years of exposure. Data were gathered from children and their teachers at four repeated time points: Wave 1 baseline, fall (2004) of third grade, Wave 2 spring (2005) of third grade, Wave 3 fall (2005) of fourth grade, and Wave 4 spring (2006) of fourth grade. Because the 4Rs Program was randomized at the school level, children who left a participating school were not followed (n = 59 in Wave 2, 127 in Wave 3, and 48 in Wave 4) and consent was requested for new children enrolling in a participating school (n = 124 at Wave 2, 177 at Wave 3, and 65 at Wave 4; Vuchinich et al., under review). Attrition between waves was minimal (on average 8.6%) and primarily because of student mobility. Parental refusals to continue were rare (e.g., n = 1 in Wave 2).

According to parent-reports at baseline, 53.4% (n = 503) of children lived in a single-parent household, 15.1% (n = 142) of parents were unemployed, 31% (n = 292) of parents had less than a high school diploma or GED, and 61.8% (n = 582) of households were at or below 100% of the federal poverty level. Based on parent-reports at baseline and NYC Department of Education records when parent-reports were missing, children represented diverse racial/ethnic groups; 45.8% (n = 542) were Hispanic/Latino, 41.3% (n = 489) Black/African American, 4.3% (n = 51) non-Hispanic White, and 8.6% (n = 102) represented other racial/ethnic groups (e.g., Asian, Pacific Islander, Native American). Table 1 presents baseline demographic characteristics by intervention and control schools.

Procedure

Forty-one schools representative of the population of NYC elementary schools were identified as

Table 1
Sample Baseline Demographic Characteristics by Intervention and Control Schools

Demographic characteristics	Intervention (n = 630)	Control (n = 554)	Total (N = 1,184)
Child age			
Mean age (in years)	8.17 (0.71)	8.18 (0.67)	8.17 (0.69)
Child gender			
Boys (%)	311 (49.4)	270 (48.7)	581 (49.1)
Girls (%)	319 (50.6)	284 (51.3)	603 (50.9)
Child race/ethnicity			
Non-Hispanic, White (%)	31 (4.9)	20 (3.6)	51 (4.3)
Hispanic/Latino (%)	287 (45.6)	255 (46)	542 (45.8)
Black/African American (%)	261 (41.4)	228 (41.2)	489 (41.3)
Other (%)	51 (8.1)	51 (9.2)	102 (8.6)
Household SES risk index			
No risks (%)	129 (20.5)	100 (18.1)	229 (19.3)
One risk (%)	192 (30.5)	159 (28.7)	351 (29.6)
Two risks (%)	161 (25.6)	188 (33.9)	349 (29.5)
Three risks (%)	120 (19.0)	82 (14.8)	202 (17.1)
Four risks (%)	28 (4.4)	25 (4.5)	53 (4.5)
Mean SES risk (*SD*)	1.56 (1.14)	1.59 (1.08)	1.57 (1.11)
Community risk			
Mean community risk (*SD*)	2.06 (.79)[a]	1.97 (.73)	2.01 (.76)
Child behavioral risk index			
No risks (%)	532 (84.4)	481 (86.8)	1013 (85.6)
One risk (%)	63 (10)	49 (8.8)	112 (9.5)
Two risks (%)	35 (5.6)	24 (4.3)	59 (5)
Mean behavioral risk (*SD*)	0.21 (.53)[a]	0.17 (.48)	0.19 (.51)
Y1–Y2 classroom size	19.16 (4.41)[a]	20.17 (4.34)	19.63 (4.40)
Y1–Y2 teacher burnout	1.94 (.87)[a]	2.01 (.91)	1.97 (.89)
Y1–Y2 teacher experience	5.77 (4.32)[a]	8.39 (5.73)	6.99 (5.19)

Note. [a]Baseline covariate significantly higher for intervention compared to control group: community risk, $t(1182) = 2.20$, $p < .050$; behavioral risk, $t(1182) = 2.41$, $p < .050$; Y1–Y2 (Year 1–Year 2) classroom size, $t(1182) = 3.94$, $p < .001$; Y1–Y2 teacher experience, $t(1182) = 8.92$, $p < .001$. SES = socioeconomic status.

potential participants in the 4Rs evaluation. Twenty-four agreed to the matching and randomization process. Prior to randomization, a pairwise matching procedure was used to maximize demographic similarity of intervention and control schools. An algorithm was used to compute the distance from each school to every other school along 20 demographic and school characteristics drawn primarily from the 2001–2002 administrative databases kept by the city's Department of Education (Jones et al., 2010). The matching procedure produced 12 matched school pairs, and the 9 best matched pairs of schools were initially selected for inclusion in the study with the three remaining pairs reserved as back-ups. The first back-up pair was eliminated from consideration by a Local Instructional Superintendent. Two of our initially selected best matching pairs were dropped (in one case the Principal was previously trained in the RCCP—an original exclusion criterion, and in the other, the Principal was unwilling to proceed with the study if assigned to the control group) and were replaced with our 2nd and 3rd back-up pairs. The first school in each pair was assigned to the intervention or control group based on a randomly generated number, and the second school in the pair was, therefore, assigned to the other group. The two groups did not differ significantly on any of the matching characteristics and eta-square values were minimal. Based on these statistics, the schools can be described as racially and ethnically diverse, composed primarily of students receiving free school lunch, and with attendance rates over 89% and 1-year stability rates ranging from 86% to 95%.

Consent packages (in English and Spanish) were sent home to all parents of third grade children in the 18 participating schools informing them of the study and seeking consent for their child to participate. For the purposes of this article, as children continued to enter the study over time consent rates were calculated at the end of Year 2. The overall consent rate was 64.54% across schools (range = 44%–79%); consent rates did not differ between intervention (65.2%) and control (63.7%) schools. Nonparticipants included children whose parents did not speak English or Spanish well enough to consent and special needs children who could not be interviewed even on an individual basis (e.g., because of autism). Demographic and achievement comparisons between consented and nonconsented children were conducted using school records data (with nonconsented children's records stripped of identifiable information), and while significant differences were found for gender (fewer male participants, Cohen's $d = .15$) and school absences (higher absences for nonparticipants, Cohen's $d = .21$), their effect sizes were small. No differences were found for children's race/ethnic status, receipt of free lunch, number of suspensions, and New York state reading and math test scale scores. Furthermore, there were no differences in strength of association between demographic and achievement variables and children's consent status between intervention and control schools. These results make us confident that we have obtained a representative sample of children from the 18 schools.

With regard to outcomes, at each wave teachers completed questionnaires rating the language and literacy skills, as well as social competence and externalizing problems of each consented child in their class. Teachers were paid the union wage of $36.50 per hour for survey completion at each assessment. At each wave, children also completed questionnaires rating their aggressive and prosocial cognitions, and their internalizing symptoms. Data were collected from the children in small class groups ($n = 5–20$). Questions were read aloud by a research assistant while a second research assistant circulated to monitor children's placement of responses and answer their questions. Children who were either nonconsented or consented but refused assent, worked on an alternative activity with their classroom teacher.

Intervention

The 4Rs Program (Reading, Writing, Respect and Resolution) is a universal, school-based intervention in literacy development and social-emotional learning that integrates a focus on social and emotional development into the language arts curriculum for children in grades K-5. Developed and run by a community-based nonprofit organization called the Morningside Center for Teaching Social Responsibility (MCTSR), the 4Rs Program uses high quality children's literature as a springboard for helping students gain skills and understanding in the areas of handling anger, listening, assertiveness, cooperation, negotiation, mediation, building community, celebrating differences, and countering bias. By highlighting universal themes of conflict, feelings, relationships, and community, the 4Rs curriculum adds social and emotional meaning and skill building to rigorous literacy instruction.

The 4Rs Program has two primary components: (a) a comprehensive 7-unit, 21–35 lesson, literacy-

based curriculum in social-emotional learning; and (b) 25 hr of training followed by ongoing coaching of teachers to support them in teaching the 4Rs curriculum with a minimum of 12 contacts in one school year. Each of the seven curricular units is organized around a specific grade-appropriate children's book. Each unit begins with a comprehensive book reading and discussion, ensuring students understand the primary themes of the story and allowing them to connect the themes to their own lives. This is followed by three to five social-emotional learning skill lessons in which children are able to practice specific skills in the context of a larger discussion of the book. The lessons are designed to engage the children in the learning and practice of social-emotional, conflict resolution, and community building skills relevant to each unit. Each unit also includes additional activities and related readings. The curriculum provided to teachers (i.e., the seven units and accompanying lessons) is detailed in a standardized, grade-specific teaching guide.

Intensive professional development for teachers using the 4Rs curriculum includes a 25-hr introductory training course, and ongoing classroom coaching by a 4Rs staff developer from Morningside Center. Teachers receive Learning Kits with a full set of materials needed to implement the program, including the children's books, the manualized teaching guide for the appropriate grade, and the parent guide. The introductory training is designed to: (a) introduce the teachers to the curricular units, children's books, lessons, and activities; (b) provide opportunity for teachers to practice conflict resolution skills at the adult level through role play and experiential learning; and (c) to inspire them to employ curricular ideas and skills in their own professional and personal lives. Ongoing classroom coaching encompasses class lesson modeling and workshops led by the staff developer, co-planning and teaching of lessons by the teacher and staff developer, and lessons observations and feedback. Staff developers convene regular conferences with teachers either one-on-one or in a grade-specific group. A full cost study for this evaluation has not yet been completed; however, the estimated implementation cost of the 4Rs Program for this evaluation study is approximately $90.00 per child per year.

School-wide implementation of the two primary 4Rs components (curriculum delivery and teacher training and coaching) was systematically tracked and monitored during the course of the study. Implementation data from Year 1 show that teachers in the nine intervention schools received (a) on average 2.4 (SD = 0.33) days of training in the delivery of the 4Rs curriculum, and (b) an average of 38 (SD = 9.6) days of coaching per school. On average, teachers delivered three-quarters of a lesson per week, with the majority closer to the benchmark of one lesson per week. The majority of teachers spent on average between 20 and 25 total hours on 4Rs during Year 1 (~40 min/week). Year 2 implementation data reveal a slight decrease in training days, coaching days, average classroom lessons per week, and amount of time spent on 4Rs per week. Our data indicate that teachers trained in the 1st year of the study who remained in the school the following year were closer to program benchmarks (i.e., on average they implemented one lesson/week and spent ~50 min on 4Rs per week).

Information on the implementation of various social and character development activities was also gathered from all third-, fourth-, and fifth-grade teachers in both intervention and control schools using a measure developed for the broader SACD research program. The percentage of control teachers reporting using any SACD activities in their classroom ranged from 75.6% to 85.6% over the three full years of the study. A greater percentage of intervention compared to control teachers reported using any SACD activity in Year 1 (effect size = .31), as well as specific activities targeting violence prevention/peace promotion (effect size = .50) and social and emotional development (effect size = .52; SACD Research Consortium, 2008).

Measures

All scale scores were computed as the mean across the items for each construct. Basic psychometric properties and mean levels for each construct at each time point are presented by intervention and control groups (see online Supporting Information Table S1). These measures are theoretically based (e.g., from Dodge (1986) and Social Information Processing Theory; Selman (2003) and Developmental-Structural Theory) and have been shown to be psychometrically valid with low-income, multiethnic groups (e.g., Dalberg, Toal & Behren, 1998). Moreover, they have been used by our team in previous work evaluating the RCCP (Aber et al., 1996). The primary constructs and measures are presented by domain, followed by the child baseline covariates and classroom covariates.

Social-Cognitive Processes

Hostile attribution biases were measured using child self-reports on an adaptation of the Home

Interview Questionnaire (Dalhberg, Toal, & Behrens, 1998; Dodge, 1986). A series of six vignettes that depict ambiguous but provocative social scenarios is read aloud while pictorial representations of the scenarios are presented (e.g., a student's milk carton is spilled on another child's back). Following presentation of each vignette, children are presented with four possible causal attributions regarding the intent of the provocateur and are asked to select one causal attribution. Two attributions refer to the provocateur's intent as benign or accidental = 0 (e.g., the milk was spilled accidentally) and two responses describe the provocateur's intent as hostile or purposeful = 1 (e.g., the student was being mean). Cronbach alphas for this scale across the four waves range from .76–.81.

Aggressive interpersonal negotiation strategies were also self-reported using an adaptation of the Home Interview Questionnaire (Dalhberg et al., 1998; Dodge, 1986). Following the presentation of each vignette and causal attributions, children are presented with four response strategies about what they would do next in the scenario and are asked to select one strategy. Three response strategies are nonaggressive or benign = 0 (e.g., act like nothing happened) and one strategy is aggressive or hostile = 1 (e.g., spill paint on the child). Cronbach alphas for this scale across the four waves range from .89–.91.

Normative beliefs about aggression were measured using the Normative Beliefs about Aggression Scale (Huesmann & Guerra, 1997), a measure design to assess children's beliefs about the acceptability of the use of aggression in certain situations. The total normative beliefs scale ranges from (1) low normative beliefs, the use of aggression is ''perfectly OK,'' to (4) high normative beliefs, the use of aggression is ''really wrong'' and includes 12 items. Cronbach alphas for this scale across the four waves range from .85–.92.

Social-Emotional Symptomatology

Child ADHD Symptoms was measured using teacher reports on nine items from the ADHD Symptomatology Scale (Milch, Loney, & Landau, 1982). This questionnaire asks teachers about children's typical behaviors observed within the past 30 days and is rated on a 4-point scale (*never* = 1; *almost always* = 4). The scale includes two subscales: Hyperactivity (e.g., is excitable or impulsive) and Inattentiveness (e.g., has difficulty organizing tasks or activities). Cronbach alphas for this scale across the four waves range from .91–.92.

Depressive symptoms were assessed via self-reports of depressive symptoms on the Diagnostic Interview Schedule for Children Predictive Scales (Lucas et al., 2001). This questionnaire contains six items that ask children about whether they have experienced particular depressive symptoms (e.g., ''Has there been a time when nothing was fun for you and you just weren't interested in anything?'' in the past year [for Wave 1] or since the new year [for Wave 2]). Items are rated on a 2-point scale (*no* = 0; *yes* = 1). Cronbach alphas for this scale across the four waves range from .49–.86.

Aggressive and prosocial fantasies were measured using self-reports on the What I Think instrument (Rosenfeld, Huesmann, Eron, & Torney-Purta, 1982). The Aggressive Fantasies subscale contains six items that ask children about aggressive thoughts that just ''pop into your head'' or ''daydreams'' (e.g., daydreams about hitting or hurting someone; pretending to fight with someone). The Prosocial subscale contains six items that ask children about prosocial thoughts that just ''pop into your head'' or ''daydreams'' (e.g., daydreams about helping other kids, about doing nice things for other kids). Items are rated on a 3-point scale (*no* = 0; *a lot* = 2). Cronbach alphas for these scales across the four waves range from .59–.74.

Aggressive and Prosocial Behavior

Child Aggression was measured using teacher reports of aggression on the Behavioral Assessment System for Children (Reynolds & Kamphaus, 1998). This questionnaire asks teachers about children's typical behaviors observed within the past 30 days, and includes 13 aggressive behavior items (e.g., physically aggressive, threatening or critical of others). Items are rated on a 4-point scale (*never* = 1; *almost always* = 4). Cronbach alphas for this scale across the four waves range from .95–.96.

Child Social Competence was measured using teacher reports on the 18 item Social Competence Scale (CPPRG, 1999), which includes two subscales: Prosocial Behaviors (e.g., is good at understanding other people's feelings) and Emotion Regulation (e.g., can calm down when excited or all wound up). Cronbach alphas for this scale across the four waves were .97.

Academic Functioning

Academic skills were measured from teacher-reports on items adapted from the Early Childhood Longitudinal Study, Kindergarten Cohort of

1998–1999 (ECLS–K), 3rd Grade Assessment. Nine items tap the degree to which a child has acquired and demonstrates the targeted language and literacy skills, knowledge, and behaviors appropriate for third graders (e.g., reads fluently, conveys ideas clearly, composes multiparagraph stories). Items are rated on a 5-point scale (*not yet* = 1; *proficient* = 5). Cronbach alphas for this scale across the four waves range from .96–.97.

Standardized math and reading achievement were measured using children's scaled scores on the New York State standardized assessments of math and reading achievement at the end of third grade for the 2004–2005 school year and at the end of fourth grade for the 2005–2006 school year. *Attendance rate* was obtained from the NYC Department of Education. Attendance rate was assessed as the proportion of full days present in school during children's third and fourth grade school years (2004–2005; 2005–2006).

Baseline Covariates

Household socioeconomic status (SES) risk index was calculated as the sum of four parent-reported, dichotomous demographic characteristics at baseline; single-parent household, less than high school education, poverty at or below 100% of the federal poverty level, and unemployment. For children with missing parent-reported risk indicators, these values were imputed based on child school, gender, and race/ethnicity.

Community risk was measured from parent-reports on the Community Risks and Resources Questionnaire (Forehand et al., 2000). Parents are asked to rate how well seven statements describe their neighborhood (e.g., people roam streets and carry weapons, drugs are sold or used, houses/apartments are in poor condition). Items are rated on a 4-point scale (*not at all* = 1; *a lot* = 4). For children with missing parent-reported community risk indicators, these values were imputed based on child school, gender, and race/ethnicity.

Child behavioral risk was measured from baseline teacher-reports of aggression and conduct problems on the Behavioral Assessment System for Children (Reynolds & Kamphaus, 1998). This questionnaire asks teachers about children's typical behaviors observed within the past 30 days, and includes 13 aggressive behavior items (e.g., physically aggressive, argumentative, threatening or critical of others) and 11 conduct-disordered behavior items (e.g., steals, truancy). Items are rated on a 4-point scale (*never* = 1; *almost always* = 4). Behavioral risk was calculated based on a nationally normed *t*-score average ($t = 63.5$ and 62.9 for aggressive behavior and conduct disorder, respectively). Children were grouped at baseline according to whether they were below the *t*-score cutoff on both behaviors = 0, at or above the cutoff on one behavior = 1, or above the cutoff on both behaviors = 2.

In addition to this set of child baseline covariates, child gender and race/ethnic background were included. To account for potential teacher effects owing to the fact that teachers change between the 2 years of the study (i.e., from third to fourth grade), a set of key teacher/classroom covariates were also included at the child level in the models for teacher-reported outcomes. These included Teacher Burnout (Maslach, Jackson, & Schwab, 1996), Teacher Experience, and Classroom Size. It should be noted that because these constructs reflect the entire period of growth examined in this article, and were included primarily as a way to control for changes in the teachers reporting on children over time, their associations with growth and change in the outcomes of interest are not presented next but are included in the relevant table.

Results

The results will be described in two parts. First, a set of preliminary analyses are presented, followed by results from a series of hierarchical linear growth models employed to address the study's primary questions.

Preliminary Analyses

Descriptive statistics for the primary child outcomes at Wave 1 (baseline, Fall third grade), Wave 2 (Spring third grade), Wave 3 (Fall fourth grade), and Wave 4 (Spring fourth grade) are presented for the sample overall and by intervention group (see online Supporting Information Table S1). As expected, owing to the matching and randomization procedures we employed, there were no differences in mean levels of the outcomes at baseline by intervention group.

A series of independent samples *t*-tests, analyses of variance, and bivariate correlations were employed to examine basic mean differences in Wave 1 (baseline, Fall 2004) measures of the child outcomes by key baseline covariates including: child gender, child race/ethnic background, baseline behavioral risk, household SES risk, and community risk. Findings from these analyses are

consistent with those presented next in which the association between baseline covariates and the growth parameters (intercept and slope) are described, and are therefore not presented in detail here. Complete tables from these analyses are available from the first author upon request.

Main Analyses

Analytic Strategy

As described in the Method section, schools were matched into pairs prior to randomization on 20 characteristics, including such constructs as size, race/ethnic composition, reading achievement, per pupil expenditures. Thus, the data presented here are nested: time (four repeated assessments) is nested in children, children are nested in schools, and schools are nested in their matched pairs. To accommodate these design features, estimates of intervention impact on change in the primary child outcomes from preintervention baseline (Wave 1, Fall 2004) to the fourth time point (Wave 4, Spring 2006) were calculated using a series of three-level hierarchical linear growth models with matched pair fixed effects in HLM 6.02. In these models, Level 1 represents time (i.e., the four repeated assessments of the constructs of interest for each child), Level 2 represents the child, and Level 3 represents schools. All child- and classroom-level covariates were included at Level 2. Level 3 included an intervention dummy as well as eight school pair dummies to represent the school matches. In addition, as indicated in the Introduction, both in response to the growing literature reporting stronger effects of school-based interventions for high-risk children (e.g., CPPRG, 2007) and building on our findings from the 1st year of the study (Jones et al., 2010), we examined a number of cross-level intervention by baseline covariate interactions. In interpreting the results, we adhere to the traditional benchmarks for considering an alpha level of $p < .05$ as statistically significant, but given the nature of the design (resulting in only 8 *df* to estimate the intervention effect), we note effects up to the .10 level, particularly in the case of interactions (McClelland & Judd, 1993). In addition, for both main and interaction effects, we report effect sizes (e.g., McCartney & Rosenthal, 2000). Following procedures adapted from Snijders and Bosker (1999), effect sizes for the growth models were calculated by dividing the relevant intervention estimate by an estimated standard deviation of the outcome (calculated from a fully unconditional three-level model as the square-root of the total variance). Effect sizes for the two-level models employed for the school record outcomes were calculated by dividing the estimate of the intervention effect by the standard deviation of the control group (Jones et al., 2010).

In the following presentation, we describe results from three models: (a) a series of unconditional models designed to determine the most appropriate functional form to represent change in the child- and teacher-reported outcomes (as primarily linear or curvilinear), (b) building from the unconditional models, a set of models that include intervention status and the full set of baseline covariates as predictors of the intercept and the relevant growth parameters, and (c) models in which intervention status is interacted with each baseline covariate. As noted in the Method section, we chose and included a set of classroom covariates to account for changes in the teachers reporting on children between years and for any potential influence of the intervention on teachers and their reports on children, resulting in a very conservative assessment of the impact of intervention on children's outcomes. In addition, we provide additional support for intervention effects on the growth parameters by conducting a set of parallel but simpler models in which the impact of intervention on Wave 4 (Spring 2006) was examined controlling for baseline levels and for the baseline covariates. Finally, because the school records data were only available for two time points (i.e., no preintervention baseline, and once in the Spring of each year), these models were estimated slightly differently as presented next.

Unconditional Models

To determine the most appropriate functional form for each outcome three models were compared: intercept only; intercept-slope; and intercept-slope-quadratic. Model comparisons were conducted using a method similar to the chi-square difference test frequently applied to structural equation models. In this case, the statistical significance of changes in the Deviance statistic between models, given parallel change in degrees of freedom between models, were examined. In all cases, intercept–slope models were significantly better than intercept only models. With respect to child self-report outcomes, in all cases, the intercept–slope–quadratic model was a significantly better fit than the intercept–slope only models. However, despite the better fit of the model, the quadratic

parameter itself was not significant for any child-reported outcomes. In contrast, for teacher-reported outcomes (Academic Skills, Aggression, Social Competence, and ADHD Symptoms), the intercept–slope–quadratic models did not provide a better fit to the data than did the intercept–slope only models. In general, consistent with our prior work (e.g., Aber et al., 2003), children showed growth over the four repeated assessments in negative outcomes (e.g., HAB, AINS, aggressive fantasies) and declines over time in positive outcomes (e.g., prosocial fantasies). In general, children's academic skills increased over the course of each academic year, and their scaled scores on standardized tests of reading and math achievement increased between school years.

Based on this set of models, intervention effects were estimated for the intercept, slope, and quadratic parameters for the child-report outcomes, and for the intercept and slope parameters only for the teacher-reported outcomes. In addition, intervention by baseline covariate interactions were estimated only for the appropriate growth parameters for each reporter/outcome. Finally, because the form of the final models differ between child- and teacher-reported dependent variables, results are reported next first for child self-reported outcomes, followed by teacher-reported outcomes, and then for school record outcomes.

Main Effects of Intervention and Baseline Covariates

Child self-reports. Intervention effects were estimated for three parameters: the intercept, centered at Wave 1 and representing intervention–control differences at the preintervention baseline (Fall 2004), the slope representing intervention–control differences in linear change in the outcome of interest across the four repeated assessments, and the quadratic parameter representing intervention–control differences in acceleration or deceleration of the trajectory over the four repeated assessments. As presented in Table 2, there were no intervention–control differences at the preintervention baseline for any outcome examined (i.e., no significant effects for the intercept parameter). This suggests that our process of pairwise matching and randomizing schools was successful in creating equivalent groups at baseline. There were intervention–control differences in the slope parameter for two of the six child-report outcomes, one with a p-value $< .1$ (HAB, ES = .14, see Figure 1); and the other with a p-value $< .05$ (Depressive Symptoms, ES = .22). The effects are supported by intervention effects for Spring 2006 (Wave 4) levels of HAB, at $p < .05$ (ES = .25), controlling for baseline levels. There was an intervention effect on the quadratic parameter for AINS at $p < .1$ (ES = .08, see Figure 2), which indicates that the paths for the intervention and control groups begin to diverge at the beginning of the 2nd year of exposure to the 4Rs Program. This effect is supported by significant intervention effects on Spring 2006 (Wave 4) levels of AINS (ES = .42), controlling for baseline levels. There were no intervention effects for normative beliefs about aggression, aggressive or prosocial fantasies.

As shown in Table 2, with regard to the impact of the baseline covariates, there were few intercept associations with the exception of gender for AINS, normative beliefs about aggression, and aggressive fantasies (with girls showing lower levels of aggressive strategies and fantasies at baseline), and baseline behavioral risk for higher HAB, AINS, aggressive and prosocial fantasies. There were few significant effects of the baseline demographic characteristics on the slope or quadratic parameters.

Teacher reports. For teacher-reported outcomes, intervention effects were estimated for two parameters: the intercept, centered at Wave 1 and representing intervention–control differences at the preintervention baseline (Fall 2004), the slope representing intervention–control differences in linear change in the outcome of interest across the four repeated assessments. As with the child-report outcomes, as shown in Table 3, there were no intervention–control differences at the preintervention baseline for any outcome examined (i.e., no significant effects for the intercept parameter). There were intervention–control differences in the slope parameter for three of the four teacher-report outcomes: teacher-reported ADHD Symptoms at $p < .1$ (ES = .12) which was accounted for in follow-up analyses by one of its subscales, Hyperactivity, (−.093 (.05), $p < .10$, ES = .13); teacher-reported Social Competence at $p < .05$ (ES = .14) which was accounted for in follow-up analyses by one of its subscales, Prosocial Behavior, (.135 (.05), $p < .05$, ES = .18); and teacher-reported Aggression at $p < .05$ (ES = .05). The effect for teacher-reported Aggressive Behavior is supported by significant intervention effects on Spring 2006 (Wave 4) levels of Aggression also at $p < .05$ (ES = .21), controlling for baseline levels. There were no intervention effects for teacher report of Academic Skills.

As shown in Table 3, there were a number of effects of the baseline covariates for the intercept parameter. Specifically, there were significant effects of gender on the intercept parameter such

Table 2
Unstandardized Estimates and (SE) of 4R's Impact on 4-Wave Trajectories of Child Self-Reports of their Social-Cognitive Processes and Social-Emotional Symptomatology

	Social-cognitive processes						Social-emotional symptomatology					
	HAB (CR)		AINS (CR)		NOBAGS (CR)		AGGFANT (CR)		PROFANT (CR)		DEP (CR)	
Intercept												
Intervention status (8 *df*)	−0.03	(0.02)	0.03	(0.03)	−0.08	(0.05)	−0.08	(0.05)	0.00	(0.15)	0.01	(0.03)
Child and teacher covariates												
Child gender, 1 = Girl (1176 *df*)	0.01	(0.02)	−0.08**	(0.02)	−0.17***	(0.03)	−0.18***	(0.03)	0.05	(0.03)	−0.01	(0.02)
Child Hispanic versus White (1176 *df*)	−0.01	(0.06)	0.01	(0.08)	−0.09	(0.09)	−0.16	(0.09)	0.14	(0.09)	−0.07	(0.05)
Child Black versus White (1176 *df*)	0.01	(0.06)	0.01	(0.08)	−0.03	(0.09)	−0.03	(0.09)	0.11	(0.09)	−0.04	(0.05)
Child other versus White (1176 *df*)	−0.07	(0.06)	−0.06	(0.09)	−0.08	(0.10)	−0.39***	(0.10)	0.06	(0.10)	−0.18**	(0.05)
SES risk (1176 *df*)	0.00	(0.01)	0.01	(0.01)	0.00	(0.02)	−0.02	(0.02)	0.02	(0.02)	−0.01	(0.01)
Community risk (1176 *df*)	0.00	(0.02)	0.00	(0.03)	0.02	(0.03)	0.04	(0.03)	0.01	(0.03)	0.00	(0.01)
Behavioral risk (1176 *df*)	0.06**	(0.02)	0.10*	(0.04)	0.03	(0.04)	0.10**	(0.04)	0.10**	(0.04)	0.02	(0.02)
Slope												
Intervention status (8 *df*)	−0.05†	(0.03)	0.04	(0.04)	0.08	(0.06)	0.04	(0.05)	0.05	(0.05)	−0.06*	(0.02)
Child and teacher covariates												
Child gender, 1 = Girl (1176 *df*)	−0.02	(0.02)	−0.07	(0.04)	−0.13*	(0.05)	−0.05	(0.04)	0.04	(0.04)	0.04	(0.02)
Child Hispanic versus White (1176 *df*)	−0.01	(0.07)	−0.02	(0.09)	0.03	(0.14)	−0.00	(0.12)	−0.05	(0.11)	0.12	(0.06)
Child Black versus White (1176 *df*)	0.06	(0.07)	0.05	(0.09)	0.04	(0.14)	0.12	(0.12)	−0.05	(0.11)	0.14*	(0.07)
Child other versus White (1176 *df*)	0.07	(0.08)	−0.03	(0.10)	−0.09	(0.16)	0.11	(0.13)	−0.02	(0.12)	0.16*	(0.07)
SES risk (1176 *df*)	0.01	(0.01)	−0.01	(0.02)	−0.01	(0.03)	0.00	(0.02)	0.00	(0.02)	0.00	(0.01)
Community risk (1176 *df*)	−0.01	(0.02)	0.04	(0.03)	−0.00	(0.03)	−0.02	(0.03)	−0.02	(0.03)	0.01	(0.02)
Behavioral risk (1176 *df*)	−0.03	(0.03)	−0.06	(0.04)	−0.08	(0.06)	−0.06	(0.05)	−0.06	(0.04)	0.02	(0.02)
Quadratic												
Intervention status (8 *df*)	0.01	(0.01)	−0.03†	(0.01)	−0.02	(0.02)	−0.01	(0.01)	−0.02	(0.01)	0.02	(0.00)
Child and teacher covariates												
Child gender, 1 = Girl (1176 *df*)	0.00	(0.01)	0.01	(0.01)	0.03	(0.02)	0.03*	(0.01)	−0.01	(0.01)	−0.01	(0.01)
Child Hispanic versus White (1176 *df*)	0.01	(0.02)	0.02	(0.03)	0.01	(0.05)	0.01	(0.04)	0.01	(0.03)	−0.02	(0.02)
Child Black versus White (1176 *df*)	−0.01	(0.02)	0.00	(0.03)	0.02	(0.05)	−0.02	(0.04)	0.01	(0.04)	−0.03	(0.02)
Child other versus White (1176 *df*)	−0.02	(0.02)	0.01	(0.03)	0.06	(0.05)	−0.02	(0.04)	0.01	(0.04)	−0.02	(0.02)
SES risk (1176 *df*)	−0.00	(0.00)	0.00	(0.01)	0.00	(0.01)	−0.00	(0.01)	−0.00	(0.01)	−0.00	(0.00)
Community risk (1176 *df*)	0.00	(0.01)	−0.00	(0.01)	0.00	(0.01)	0.01	(0.01)	0.01	(0.01)	−0.00	(0.01)
Behavioral risk (1176 *df*)	0.01	(0.01)	0.02	(0.01)	0.04*	(0.02)	0.01	(0.01)	0.01	(0.01)	−0.00	(0.01)

Note. Eight dummy variables representing eight of the nine school-level matched pairs are included in all models at the school level (Level 3) along with estimates of intervention on intercept, slope, and quadratic, with pair 9 serving as the referent group. HAB = hostile attributional biases; AINS = aggressive interpersonal negotiation strategies; NOBAGS = normative beliefs about aggression; AGGFANT = aggressive fantasies; PROFANT = prosocial fantasies; DEP = depression; CR = child report.
†$p < .10$. *$p < .05$. **$p < .01$. ***$p < .001$.

that teachers reported girls as having significantly lower levels of Aggression, ADHD Symptoms, and higher levels Social Competence and teacher-reported Academic Skills at baseline than boys. Teachers also reported Black children as having significantly higher levels of Aggression and lower levels of Social Competence compared to White children at baseline. Children rated as high on baseline behavioral risk were rated as significantly higher in intercept levels of Aggression, ADHD Symptoms, and lower in levels of Social Competence and Academic Skills. With regard to slope, children rated as high on baseline behavioral risk showed steeper declines in Aggression, ADHD Symptoms, and steeper increases in Social Competence. These somewhat unexpected slope effects are likely because of the high levels of these constructs at baseline for children rated as high on baseline behavioral risk.

School records. As noted earlier, the three school record outcomes were examined as two-level models with Year 2 scores as the dependent variable, the same set of baseline covariates presented before included at Level 1, and intervention and the eight school pair matches included at Level 2. (For attendance rate, these analyses were also

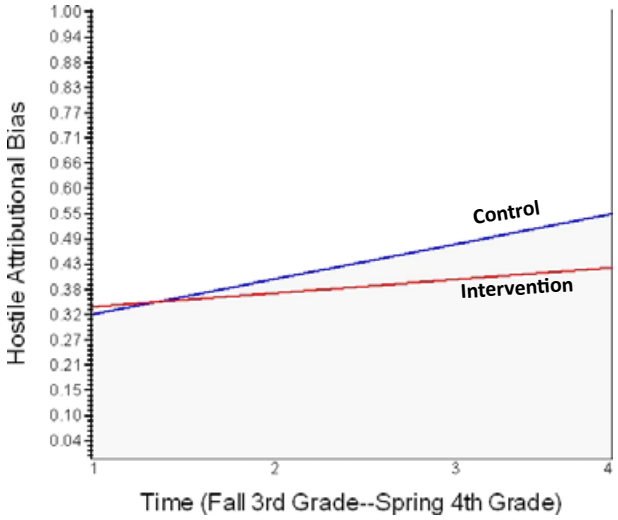

Figure 1. Impact of intervention on hostile attributional bias slope.

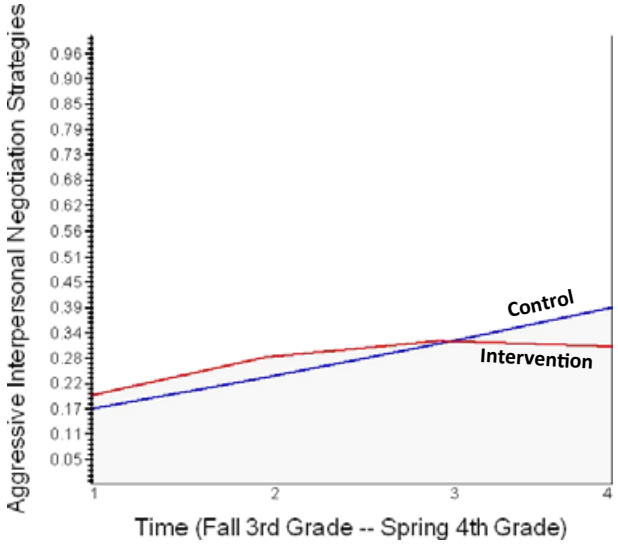

Figure 2. Impact of intervention on aggressive interpersonal negotiation strategies quadratic parameter.

conducted including attendance from the year prior to the onset of the study [i.e., when children were in second grade], and the intervention findings did not differ from what is reported here.) A baseline measure of the dependent variable could not be included in these models as universal testing did not begin until children were in third grade.

As shown in Table 4, there were no main effects of intervention for any of the three school record variables. There were main effects for baseline behavioral risk on the academic outcomes, including math achievement, reading achievement, and attendance such that children rated as high on baseline behavioral risk scored lower on standardized tests of math and reading achievement, and had lower rates of school attendance at the end of Year 2. In addition, for all three school record outcomes, there was a significant effect of socioeconomic risk at baseline, such that children with higher levels of socioeconomic risk did worse on the three indicators of school achievement at the end of Year 2.

Interaction Effects of Intervention by Baseline Child-Level Covariates

As noted earlier, cross-level interactions between intervention and the child-level baseline covariates (child gender, race/ethnic background, baseline behavioral, community, and SES risk) were estimated in a second set of models for the intercept, slope, and quadratic parameters for child-report outcomes, for the intercept and slope parameters for the teacher-report outcomes, and for the two-level models for the school record outcomes. There were no statistically significant intervention by baseline covariate interactions with three exceptions: as shown in Table 4, there were interactions of intervention by baseline behavioral risk for children's math achievement at $p < .1$ and children's reading achievement at $p < .05$. As shown in Figure 3 for math achievement, children with the highest level of baseline behavioral risk (based on teacher reports) show the greatest positive difference in math achievement between the intervention and control groups with effect sizes of .56 for a score of 2 on behavioral risk, of .18 for a score of 1, and .14 for a score of 0. For reading achievement a similar pattern holds, with effect sizes of .60 for a score of 2 on behavioral risk, of .39 for a score of 1, and .06 for a score of 0. In addition, there was a significant intervention by baseline behavioral risk interaction for teacher report of academic skills in the model that estimated intervention impacts on Wave 4 levels of academic skills controlling for baseline levels and covariates (0.35 (.18), $p < .1$, see Figure 4). Again, the pattern of this interaction is consistent with the school records outcomes, such that children with the highest level of baseline behavioral risk (based on teacher reports) show the greatest positive difference in teacher-reported Academic Skills between the intervention and control groups with effect sizes of .56 for a score of 2 on behavioral risk, of .25 for a score of 1, and .13 for a score of 0. Note the form of this analytic model is parallel to those conducted for the school

Table 3
Unstandardized Estimates and (SE) of 4R's Impact on 4-Wave Trajectories of Teacher Reports of Children's ADHD Symptoms, Aggressive and Prosocial Behavior, and Their Academic Skills

	ADHD (TR)		AGG (TR)		S-COMP (TR)		AC SKILL (TR)	
Intercept								
Intervention status (8 *df*)	−0.05	(0.04)	0.00	(0.03)	0.07	(0.04)	0.02	(0.11)
Child and teacher covariates								
Child gender, 1 = Girl (1173 *df*)	−0.33***	(0.04)	−0.19***	(0.03)	0.38***	(0.04)	0.24**	(0.07)
Child Hispanic versus White (1173 *df*)	−0.12	(0.10)	0.07	(0.07)	−0.17	(0.11)	−0.37*	(0.18)
Child Black versus White (1173 *df*)	−0.07	(0.10)	0.18**	(0.07)	−0.28**	(0.11)	−0.23	(0.18)
Child other versus White (1173 *df*)	−0.29**	(0.11)	−0.05	(0.07)	0.12	(0.12)	−0.02	(0.19)
SES risk (1173 *df*)	0.00	(0.02)	0.01	(0.01)	0.00	(0.02)	−0.08**	(0.03)
Community risk (1173 *df*)	0.01	(0.02)	0.01	(0.02)	−0.07*	(0.03)	−0.01	(0.05)
Behavioral risk (1173 *df*)	0.66***	(0.04)	0.83***	(0.02)	−0.56***	(0.04)	−0.30***	(0.07)
Y1–Y2 classroom size (1173 *df*)	−0.03***	(0.01)	−0.02***	(0.00)	0.01**	(0.01)	0.05***	(0.01)
Y1–Y2 teacher burnout (1173 *df*)	0.04	(0.02)	0.09***	(0.02)	−0.07**	(0.02)	−0.05	(0.01)
Y1–Y2 teacher experience (1173 *df*)	−0.00	(0.00)	−0.01**	(0.00)	0.02***	(0.00)	−0.00	(0.00)
Slope								
Intervention status (8 *df*)	−0.08†	(0.04)	−0.03*	(0.01)	0.11*	(0.04)	0.06	(0.08)
Child and teacher covariates								
Child gender, 1 = Girl (1173 *df*)	−0.03	(0.04)	−0.01	(0.01)	−0.06	(0.05)	0.05	(0.07)
Child Hispanic versus White (1173 *df*)	0.09	(0.11)	−0.01	(0.03)	0.03	(0.12)	0.10	(0.17)
Child Black versus White (1173 *df*)	0.11	(0.11)	0.01	(0.03)	−0.01	(0.12)	−0.10	(0.17)
Child other versus White (1173 *df*)	0.06	(0.12)	−0.00	(0.04)	0.01	(0.13)	0.10	(0.19)
SES risk (1173 *df*)	0.02	(0.02)	−0.00	(0.01)	−0.03	(0.02)	0.03	(0.03)
Community risk (1173 *df*)	0.01	(0.03)	0.00	(0.01)	0.06	(0.03)	0.02	(0.05)
Behavioral risk (1173 *df*)	−0.17***	(0.04)	−0.14***	(0.01)	0.09*	(0.05)	0.10	(0.07)
Y1–Y2 classroom size (1173 *df*)	0.00*	(0.00)	0.00	(0.00)	0.00	(0.00)	0.00	(0.00)
Y1–Y2 teacher burnout (1173 *df*)	−0.03**	(0.01)	−0.04***	(0.01)	0.03**	(0.01)	0.04**	(0.01)
Y1–Y2 teacher experience (1173 *df*)	0.00	(0.00)	0.01***	(0.00)	−0.01***	(0.00)	−0.00	(0.00)

Note. Eight dummy variables representing eight of the nine school-level matched pairs are included in all models at the school level (Level 3) along with estimates of intervention on intercept and slope, with pair 9 serving as the referent group. ADHD = attention/hyperactivity symptoms; AGG = aggression; S-COMP = social competence, AC SKILL = academic skills; TR = teacher report.
†$p < .10$. *$p < .05$. **$p < .01$. ***$p < .001$.

record outcomes, however, in this case, baseline levels of the dependent variable were controlled. Interactions were also tested in these models (i.e., Wave 4 controlling baseline levels) for the social-emotional outcomes (both child- and teacher-report) and no additional interactions were detected.

Discussion

This article presents 2-year, longitudinal experimental impacts of a universal, integrated intervention in elementary schools that embeds instruction in social-emotional learning into a balanced literacy curriculum. The article was designed to address a number of limitations in current research on whole school reform initiatives designed to prevent behavioral and academic maladaptation and promote children's social and emotional development. To begin, consistent with the 4Rs program theory and with developmental cascades theory, it focuses on outcomes in multiple domains, addressing key social-cognitive processes and behaviors, as well as academic outcomes. In addition, the data presented are analyzed in a manner consistent with the randomized, experimental design of the evaluation and with the developmental nature of the longitudinal data, estimating intervention effects at the school level on children's developmental growth parameters across four repeated time points. As such, and as noted in the beginning of the Results section, in interpreting the results we adhere to the traditional benchmarks for considering an alpha level of $p < .05$ as statistically significant; but given the nature of the design (resulting in only 8 *df* to estimate the intervention effect), we report effects up to the .10 level, particularly in the case of interactions (McClelland & Judd, 1993).

Table 4
Unstandardized Estimates and (SE) of 4R's Impact on Children's Academic Functioning in Spring of Fourth Grade

	Academic functioning		
	MATH (SR)	READ (SR)	ATTEND (SR)
Intervention status (8 *df*)	−4.35 (10.25)	−0.99 (10.73)	0.72 (0.87)
Baseline covariates (1176 *df*)			
Child gender, 1 = Girl	−4.09 (2.45)	8.73*** (2.47)	0.63 (0.72)
Child Hispanic versus White	−2.07 (6.56)	−5.97 (6.52)	0.90 (1.85)
Child Black versus White	−2.62 (6.70)	−5.45 (6.62)	2.92 (1.89)
Child other versus White	12.16 (7.45)	6.80 (7.42)	1.94 (2.10)
SES risk	−2.32* (1.18)	−2.22* (1.17)	−0.74* (0.34)
Community risk	−1.89 (1.82)	−1.94 (1.82)	−1.06* (0.53)
Behavioral risk	−7.99** (2.54)	−9.47*** (2.54)	−1.60* (0.74)
Interactions			
Treatment × Gender (16 *df*)	1.66 (5.60)	3.07 (5.32)	−0.21 (1.53)
Treatment × Hispanic (16 *df*)	−5.82 (13.60)	−11.08 (17.54)	−16.60 (14.35)
Treatment × Black (16 *df*)	−7.28 (13.56)	−17.83 (16.26)	−17.81 (15.29)
Treatment × Other (16 *df*)	11.04 (16.15)	−4.70 (17.74)	−21.85 (16.27)
Treatment × SES Risk (16 *df*)	1.93 (2.94)	−0.89 (3.25)	−0.04 (0.71)
Treatment × Community Risk (16 *df*)	−5.05 (5.13)	−5.71 (4.70)	−0.11 (1.09)
Treatment × Behavioral Risk (16 *df*)	13.28† (6.85)	16.12* (7.16)	1.49 (1.83)

Note. Eight dummy variables representing eight of the nine school-level matched pairs are included in all models at the school level (Level 2), with pair 9 serving as the referent group. MATH = math achievement, READ = reading achievement, ATTEND = attendance rate; SR = school records.
†$p < .10$. *$p < .05$. **$p < .01$. ***$p < .001$.

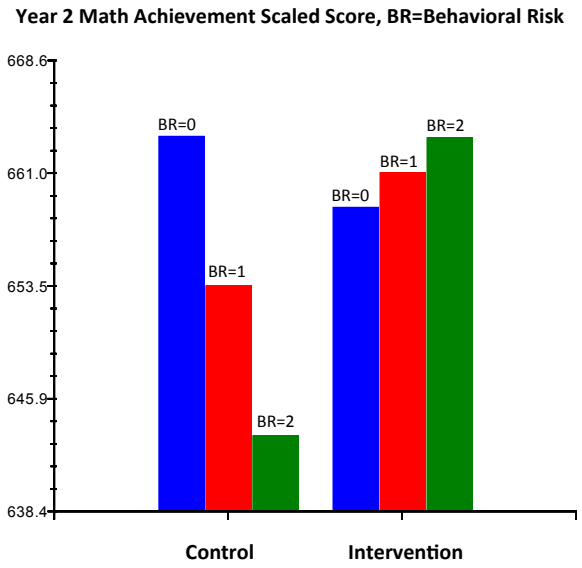

Figure 3. Interaction of intervention and baseline behavioral risk on Year 2 math achievement.

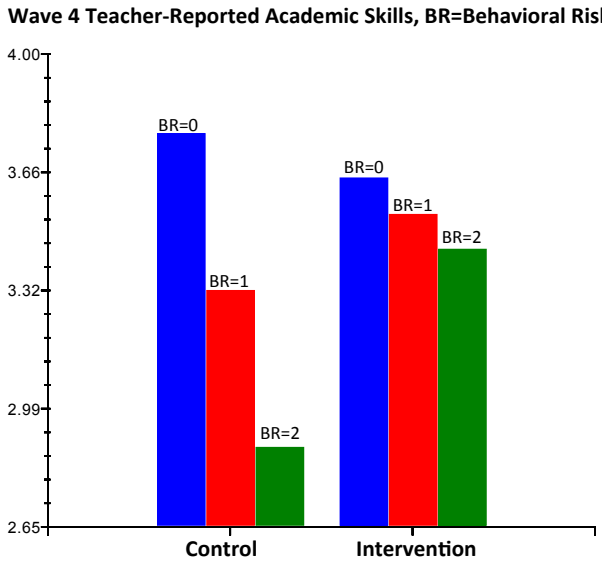

Figure 4. Interaction of intervention and baseline behavioral risk on Year 2 teacher report of academic skills.

Consistent with the dual focus of the 4Rs Program and with the emphasis of translational research on the promotion of social-emotional well-being as well as the prevention of problems, the primary findings presented in this article are summarized next by broad outcome domain. The findings are then considered in the context of intervention impacts after only 1 year of intervention (Jones et al., 2010).

Social-Cognitive Processes and Social-Emotional Symptomatology

The evidence presented in this article suggests that over two consecutive years, children in 4Rs schools reported slower rates of increase in HAB, a slowed rate of growth in AINS that appears to begin toward the outset of the 2nd year of exposure to intervention, and a steeper rate of decline in depressive and ADHD symptoms compared to children in the control schools.

Aggressive and Socially Competent Behavior

Teachers in 4Rs schools reported slower growth in children's aggressive behavior (compared to increases in control schools), and increases in social competence (compared to declines in control schools) over two school years.

Academic Functioning

While there were no main effects of intervention on teacher reports of children's academic skills or on the three school records outcomes, there were intervention by baseline behavioral risk interactions for standardized math and reading achievement and for teacher-reported academic skills. In short, children identified by teachers at greatest behavioral risk at baseline showed greater improvements as a result of exposure to 4Rs in their math and reading achievement and in teacher reports of their academic skills. Importantly, this set of intervention by baseline behavioral risk interactions were not evident for the social-emotional outcomes, regardless of the type of model examined (i.e., as a growth model estimating intervention effects on growth parameters, or a basic point-in-time model estimating intervention effects on Wave 4, controlling for baseline levels). This suggests these intervention by risk interactions are not an artifact of the form of model applied, but instead are tied to the developmental domain examined: children's academic functioning and not, in this case, their social-emotional skills.

This set of findings builds upon and expands those identified after only 1 year of intervention (Jones et al., 2010). After 1 year, we saw main effects for children's self-reports of HAB and depressive symptomatology; but there were no intervention main effects for child reports of AINS, and teacher reports of ADHD, social competence, and aggression. In addition, consistent with the moderating role of baseline behavioral risk after 2 years reported here, there were intervention by baseline behavioral risk interactions for teacher reports of children's academic skills and for children's school attendance after 1 year, with children at greatest risk showing the most substantial gains over time as a result of the 4Rs Program in the academic domain. Overall, the intervention main effects reported after both 1 and 2 years, are small in size (ranging in size from .05 for the teacher-reported aggression slope to .22 for the depressive symptoms slope). In contrast, the interactions of intervention with baseline behavioral risk are represented by effects for the highest risk group of moderate size (ranging from .56 for math achievement and academic skills to .60 for reading achievement).

Together, these findings indicate two primary types of effects of this universal intervention through the 2nd year. First, consistent with social information processing theory (e.g., Dodge, 1986), developmental-structural theory (e.g., Selman, 2003), and with our own prior work (Aber et al., 2003), we see effects for the general population of children in child reports of their social cognitions (i.e., HAB, AINS) and their social-emotional symptomatology (ADHD and Depressive Symptoms). Interestingly, intervention effects appear for children's AINS only after 2 years and after we see consistent positive changes over 2 years in their hostile attributions and depressive symptoms. Second, consistent with our theory that changes in children's social cognitions about aggression would lead to changes in their behavior, we see universal effects of the 4Rs Program after 2 years on teacher reports of aggressive behavior and social competence. While, this overall pattern is suggested by the findings, we have not yet conducted a direct test of the degree to which 4Rs-induced changes in children's social cognitions and social-emotional symptomatology *lead* to subsequent changes in their behavior.

Considering these findings from another perspective, there is growing evidence in the field of prevention science for ''sleeper effects,'' especially with respect to the prevention of externalizing behaviors such as conduct problems (e.g., Tremblay et al., 1995) and antisocial behavior (e.g., Smolkowski et al., 2005). In our case, we expected that the early changes we see for 4Rs children in HAB would translate into later changes in aggressive behavior. Consistent with the notion of sleeper effects, and with the application of cascades theory and translational research to developmental trajectories, we expect the changes we are currently seeing in children's social cognitions and their aggressive

behavior to offset problematic delinquent behaviors in later elementary and middle school when these more serious behaviors begin to increase. We also expect some of the early effects we are observing in depressive and ADHD symptoms to translate into universal benefits in other developmental domains later in elementary school and in middle school (e.g., reduced academic disengagement/failure and delayed onset of substance use; e.g., Eddy, Reid, Stoolmiller, & Fetrow, 2003; Lochman & Wells, 2004). The long-term effects of the 4Rs on such related outcomes are the focus of a follow-up study in middle school.

Finally, consistent with developmental contextualism (Cicchetti & Aber, 1998), we see 2-year effects for a targeted group of children at highest behavioral risk on their academic functioning; specifically children's scores on standardized tests of math and reading skills and teacher reports of children's academic skills. These findings are consistent with other experimental evaluations of school-based interventions that target primarily children's externalizing behavior problems. For example, the most recent report from the CPPRG indicates that the Fast Track Program showed significant effects for diagnoses and behavior in the externalizing domain (e.g., diagnoses of conduct disorder and antisocial behavior scores) for children after grades 3, 6 and 9, but only for children identified at Kindergarten at highest behavioral risk (using parent and teacher reports; CPPRG, 2007). In addition, Van Lier, Muthen, van der Sar, and Crijnen (2004) report positive impacts of the Good Behavior Game, a classroom-based behavior management program, on trajectories of conduct problems from grades 1 to 3 for children with the highest levels of disruptive behavior at baseline (Fall first grade), with similarly sized effects to those reported here (ES = .55).

Each of these studies report the impact of interventions that target one domain of functioning (i.e., externalizing behavior problems) on outcomes in that domain for children at highest behavioral risk at baseline. This study is, in contrast, the second in this project, and the first study we know of that reports significant experimental impacts of a universal, integrated, school-based social-emotional learning and literacy intervention for a subgroup of children identified based on functioning in one developmental domain, on outcomes in another developmental domain. In other words, this study reports effects of this integrated intervention on children identified at baseline as having serious behavioral difficulties on outcomes in the academic domain.

Future work with our sample will be focused on identifying the specific processes by which this universal intervention has impacts for this subgroup of children. For example, do universal, school-level population changes in the degree to which children generally attribute hostile intent to ambiguous provocations, in their aggressive and inattentive behaviors and their socially competent behavior, create the conditions in which children with particular problems (i.e., high level of behavioral difficulties) have greater learning opportunities (e.g., greater access to the teacher's focused attention)? These findings are consistent with the notion of developmental cascades (Masten & Cicchetti, 2010) and provide strong evidence for the power of translational research to influence developmental studies of trajectories of aggressive and competent behavior, particularly studies which address multiple domains of development and which consider the nuanced intersections between simultaneous changes in whole populations of children and those at particular need.

While this study has several strengths, our conclusions are tempered by limitations. A primary limitation is that our analyses were limited to outcomes at the end of the 2nd year of this 3-year intervention. With continued high quality and quantity implementation, we expect these impacts to both be sustained through the end of the 3rd year in the areas for which we already see impacts, and to expand to additional outcome domains. If, on the other hand, implementation quality and/or quantity diminishes in the 3rd year, we may see the effects reported to date attenuated.

A second factor to consider when interpreting the findings reported here is program implementation. As described in the Method section (a) there was both variability in implementation of the 4Rs Program in the intervention schools and the quantity of program implementation was below benchmarks, and (b) there was a great deal of program activity in the control schools in the broad domain of social and character development. That control schools were implementing programs and doing activities in the social and character domain to the degree they were is not surprising given the dramatic growth in interest in social-emotional learning in the last decade (Payton et al., 2008). Nonetheless, the intervention schools were implementing significantly more social and character activities, particularly and not surprisingly, those addressing violence prevention and social and emotional development. These implementation data underscore the need to examine intervention effects

in a manner that both accounts for dosage, and for implementation of similar forms of intervention activities in the control schools using sophisticated methodological tools (e.g., propensity scores, Lochman, Boxmeyer, Powell, Roth, & Windle, 2006). It is possible that such analyses will reveal stronger effects of the 4Rs, suggesting that our current findings may underestimate the impact of the 4Rs on children. These implementation data also provide a view into the possible sustainability of 4Rs in our research schools, or likely take-up of 4Rs in new schools. It is clear from the data from school districts, principals, and other decision makers in the control group that school officials are interested in social-emotional learning programs and have been purchasing and implementing a wide variety of programs in their schools. Our initial, rigorous empirical support for the 4Rs, combined with its relatively low cost (see Method), indicate that this program shows promise for its sustainability and scalability beyond the intensive support provided by this research program.

A third limitation is that the embedded nature of the intervention itself, and the design of the evaluation, precludes any ability to disentangle the impact of specific program components. As such, we cannot determine whether the activities in social-emotional learning are primarily responsible for outcomes in this domain, or whether the literacy activities are primarily responsible for outcomes in the academic domain.

Despite these limitations, this report of 2-year impacts of an integrated, social-emotional and literacy program provides evidence that this universal intervention has both universal impacts on social-cognitive processes and behaviors in the social-emotional domain, and subgroup impacts in the academic domain. This study provides good evidence that universal school-based interventions, delivered to whole populations of children, can result in positive changes in children's developmental health and well-being.

Implications for Policy and Practice

The translational value of this study also extends to both practice and policy arenas. From the standpoint of practice, our current evidence suggests that integrating pedagogical attention to building social-emotional skills through simultaneously enriching literacy practices, as instantiated by the 4Rs Program, can promote positive development in both social-emotional and academic domains. These findings challenge schools and school-based program practitioners to continue to conceptualize and operationalize a variety of practice models in which social-emotional and academic development can be fostered both in the classroom and in the school as a whole. From the standpoint of policy, findings from this study highlight the short-sightedness of educational policies that privilege and reward school and teacher attention to narrowly defined domains of development, such as academic performance at the exclusion of attention to children's social-emotional development. A growing theoretical and empirical literature supports the inextricably connected links between these domains and so we must now face the challenge of developing and adopting educational policies that both acknowledge and support this reality.

References

Aber, J. L., Brown, J., Chaudry, N., Jones, S. M., & Samples, F. (1996). The evaluation of the Resolving Conflict Creatively Program: An overview. *American Journal of Preventive Medicine*, *12*(5), 82–90.

Aber, J. L., Brown, J. L., & Jones, S. M. (2003). Developmental trajectories toward violence in middle childhood: Course, demographic differences, and response to school-based intervention. *Developmental Psychology*, *39*(2), 324–348.

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

Arnold, D. H. (1997). Co-occurrence of externalizing behavior problems and emergent academic difficulties in young high-risk boys: A preliminary evaluation of patterns and mechanisms. *Journal of Applied Developmental Psychology*, *18*, 317–330.

Bierman, K. L., Domitrovich, R. L., Nix, S. D., Gest, J. A. W., Greenberg, M. T., Blair, C., et al. (2008). Promoting academic and social-emotional school readiness: The Head Start Program. *Child Development*, *79*, 1802–1817.

Bierman, K. L., & Wargo, J. B. (1995). Predicting the longitudinal course associated with aggressive-rejected, aggressive (nonrejected), and rejected (nonaggressive) status. *Development and Psychopathology*, *7*(4), 669–682.

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

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

Brown, J. L., Roderick, T., Lantieri, L., & Aber, J. L. (2004). The Resolving Conflict Creatively Program: A school-based social and emotional learning program. In J. E. Zins, R. P. Weissberg, M. Wang, & H. J. Walberg (Eds), *Building academic success on social and emotional learning: What does the research say* (pp. 151–169). New York: Teachers College Press.

Chen, X., Rubin, K. H., & Li, D. (1997). Relation between academic achievement and social adjustment: Evidence from Chinese children. *Developmental Psychology*, *33*(3), 518–525.

Cicchetti, D., & Toth, S. L. (2006). Building bridges and crossing them: Translational research in developmental psychopathology. *Development and Psychopathology*, *18*, 619–622.

Clayton, C. J., Ballif-Spanvill, B., & Hunsaker, M. D. (2001). Preventing violence and teaching peace: A review of promising and effective antiviolence, conflict resolution, and peace programs for elementary school children. *Applied and Preventive Psychology*, *10*, 1–35.

Coie, J. D., & Dodge, K. A. (1998). Aggression and antisocial behavior. In N. Eisenberg (Ed.), *Handbook of child psychology (5th ed.), Vol. 3: Social, emotional, and personality development* (pp. 779–862). New York: John Wiley & Sons.

Conduct Problems Prevention Research Group (CPPRG). 1999. Initial impact of the Fast Track Prevention Trial for conduct problems: II. Classroom effects. *Journal of Consulting and Clinical Psychology*, *67*, 648–657.

CPPRG. (2007). Fast Track randomized controlled trial to prevent externalizing psychiatric disorders: Findings from grades 3 to 9. *Journal of the American Academy of Child and Adolescent Psychiatry*, *46*(10), 1250–1262.

Dalhberg, L. L., Toal, S. B., & Behrens, C. B. (1998). Measuring violence-related attitudes, beliefs, and behaviors among youths: A compendium of assessment tools. Centers for Disease Control and Prevention, National Center for Injury Prevention and Control, Atlanta, GA.

DeBaryshe, B. D., Patterson, G. R., & Capaldi, D. M. (1993). A performance model for academic achievement in early adolescent boys. *Developmental Psychology*, *29*, 795–804.

Dodge, K. A. (1986). A social information processing model of social competence in children. In M. Perlmutter (Ed.), *Minnesota symposium on child psychology* (Vol. 18, pp. 77–125). Hillsdale, NJ: Erlbaum.

Dodge, K. A. (2001). The science of youth violence prevention: Progressing from developmental epidemiology to efficacy to effectiveness to public policy. *American Journal of Preventive Medicine*, *20*(1) (Suppl.), 63–70.

Dodge, K. A. (2006). Translational science in action: Hostile attributional style and the development of aggressive behavior problems. *Development and Psychopathology*, *18*, 791–814.

Dodge, K. A., Bates, J. E., & Pettit, G. S. (1990). Mechanisms in the cycle of violence. *Science*, *250*, 1678–1683.

Dodge, K. A., Laird, R., Lochman, J. E., Zelli, A., & Conduct Problems Prevention Research Group (2002). Multidimensional latent-construct analysis of children's social-information processing patterns: Correlations with aggressive behavior problems. *Psychological Assessment*, *14*(1), 60–73.

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

Durlak, J. A. (1995). *School-based prevention programs for children and adolescents*. Thousand Oaks, CA: Sage.

Durlak, J. A., Weissberg, R. P., Dymnicki, A. B., Taylor, R. D., & Schellinger, K. (in press). The effects of social and emotional learning on the behavior and academic performance of school children. *Child Development*.

Eddy, J. M., Reid, J. B., Stoolmiller, M., & Fetrow, R. 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*(4), 525–552.

Embry, D. D. (2002). The Good Behavior Game: A best practice candidate as universal behavioral vaccine. *Clinical Child and Family Psychology Review*, *5*(4), 273–297.

Flay, B. R., & Allred, C. G. (2003). Long-term effects of the Positive Action program. *American Journal of Health Behavior*, *27*(Suppl 1), S6–S21.

Flay, B. R., Allred, C. G., & Ordway, N. (2001). Effects of the Positive Action program on achievement and discipline: Two matched-control comparisons. *Prevention Science*, *2*(2), 71–89.

Forehand, R., Brody, G. H., Armistead, L., Dorsey, S., Morse, E., Morse, P. S., et al. (2000). The role of community risks and resources in the psychosocial adjustment of at-risk children: An examination across two community contexts and two informants. *Behavior Therapy*, *13*, 395–414.

Gershoff, E. T., & Aber, J. L. (2006). Neighborhoods and schools: Contexts and consequences for the mental health and risk behaviors of children and youth. In L. Balter & C. Tamis-LeMonda (Eds.), *Child psychology: A handbook of contemporary issues* (2nd ed., pp. 611–645). New York: Psychology Press/Taylor & Francis.

Guerra, N. C., & Slaby, R. G. (1990). Cognitive mediators of aggression in adolescent offenders: II. Intervention. *Developmental Psychology*, *26*(2), 269–277.

Hinshaw, S. P. (1992). Externalizing behavior problems and academic underachievement in childhood and adolescence: Causal relationships and underlying mechanisms. *Psychological Bulletin*, *111*(1), 127–155.

Hoagwood, K., & Olin, S. S. (2002). The NIMH blueprint for change report: Research priorities in child and adolescent mental health. *Journal of the American Academy of Child and Adolescent Psychiatry*, *41*(7), 760–767.

Hudley, C. A., & Graham, S. (1993). An attributional intervention to reduce peer-directed aggression among African-American boys. *Child Development*, *64*, 124–138.

Huesmann, L. R., & Guerra, N. C. (1997). Children's normative beliefs about aggression and aggressive behavior. *Journal of Personality and Social Psychology*, *72*(2), 408–418.

Hundert, J., Boyle, M. H., Cunningham, C. E., Duku, E., Heale, J., McDonald, J., et al. (1999). Helping children adjust–a Tri-Ministry Study: II. Program effects. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *40*(7), 1061–1073.

Ialongo, N. S., Werthammer, L., Kellam, S., Brown, C. H., Wang, S., & Lin, Y. (1999). Proximal impact of two-first-grade preventive interventions on the early risk behaviors for later substance abuse, depression and antisocial behavior. *American Journal of Community Psychology*, *27*(5), 599–641.

Institute of Medicine. (1994). *Reducing risks for mental disorders: Frontiers for preventive intervention research*. Washington, DC: National Academy Press.

Jones, S. M., Brown, J. L., & Aber, J. L. (2008). Classroom settings as targets of intervention and research. In H. Yoshikawa & B. Shinn (Eds.), *Toward positive youth development: Transforming schools and community programs* (pp. 58–77). New York: Oxford University Press.

Jones, S. M., Brown, J. L., Hoglund, W., & Aber, J. L. (2010). Impacts of an integrated school-based social-emotional learning and literacy intervention on 3rd grade social-emotional and academic outcomes. *Journal of Consulting and Clinical Psychology*, *78*(6), 829–842.

Kellam, S. G., Brown, C. H., Poduska, J. M., Ialongo, N. S., Wang, W., Toyinbo, P., et al. (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*, *95S*, S5–S28.

Kellam, S. G., Brown, C. H., Rubin, B. R., & Ensminger, M. E. (1983). Paths leading to teenage psychiatric symptoms and substance use: Developmental epidemiological studies in Woodlawn. In S. B. Guze, F. J. Earls, & J. E. Barrett (Eds.), *Childhood psychopathology and development* (pp. 17–51). New York: Raven.

Kellam, S. G., & Langevin, D. J. (2003). A framework for understanding ''evidence'' in prevention research programs. *Prevention Science*, *4*(3), 137–153.

Kellam, S. G., Ling, X., Merisca, R., Brown, C. H., & Ialongo, N. (1998). The effect of the level of aggression in the first grade classroom on the course and malleability of aggressive behavior into middle school. *Development and Psychopathology*, *10*(2), 165–186.

Kellam, S. G., Rebok, G. W., Mayer, L. S., Ialongo, N., & Kalodner, C. R. (1994). Depressive symptoms over first grade and their response to a developmental epidemiologically based preventive trial aimed at improving achievement. *Development and Psychopathology*, *6*, 463–481.

Little, T. D. (1997). Mean and covariance structures (MACS) analyses of cross-cultural data: Practical and theoretical issues. *Multivariate Behavioral Research*, *32*, 53–76.

Lochman, J. E., Boxmeyer, C., Powell, N., Roth, D. L., & Windle, M. (2006). Masked intervention effects: Analytic methods for addressing low dosage of intervention. *New Directions for Evaluation*, *110*, 19–32.

Lochman, J. E., & Wells, K. C. (2004). The coping power program for preadolescent aggressive boys and their parents: Outcome effects at the 1-year follow-up. *Journal of Consulting and Clinical Psychology*, *72*(4), 571–578.

Lucas, C. P., Zhang, H., Fisher, P. W., Shaffer, D., Regier, D. A., Narrow, W. E., et al. (2001). The DISC Predictive Scales (DPS): Efficiently screening for diagnoses. *Journal of the American Academy of Child and Adolescent Psychiatry*, *40*, 443–449.

Maggs, J. L., & Schulenberg, J. (2001). Editors' introduction: Prevention as altering the course of development and the complimentary purposes of developmental and prevention sciences. *Applied Developmental Science*, *5*(4), 196–200.

Maslach, C., Jackson, S. E., & Schwab, R. L. (1996). Maslach Burnout Inventory-Educators Survey (MBI-ES). In C. Maslach, S. E. Jackson, & M. P. Leiter (Eds.), *MBI manual* (3rd ed.). Palto Alto, CA: Consulting Psychologists Press.

Masten, A. S., & Cicchetti, D. (2010). Developmental cascades. *Development and Psychopathology*, *22*, 491–495.

Masten, A., Coatsworth, J. D., Neeman, J., Gest, S. D., Tellegen, A., & Garmezy, N. (1995). The structure and coherence of competence from childhood through adolescence. *Child Development*, *66*, 1635–1659.

Masten, A. S., Long, J. D., Kuo, S. I.-C., McCormick, C. M., & Desjardins, C. D. (2009). Developmental models of strategic interventions. *European Journal of Developmental Science*, *3*, 282–291.

Maughan, B., Pickles, A., Hagell, A., Rutter, M., & Yule, W. (1996). Reading problems and antisocial behavior: Developmental trends in comorbidity. *Journal of Child Psychology and Psychiatry*, *37*(4), 405–418.

McCartney, K., & Rosenthal, R. (2000). Effect size, practical importance, and social policy for children. *Child Development*, *71*(1), 173–180.

McClelland, G. H., & Judd, C. M. (1993). Statistical difficulties of detecting interactions and moderator effects. *Psychological Bulletin*, *114*, 376–390.

Milch, R., Loney, J., & Landau, S. (1982). Independent dimensions of hyperactivity and aggression: A validation with playground observation data. *Journal of Abnormal Child Psychology*, *91*, 183–198.

Miles, S. B., & Stipek, D. (2006). Contemporaneous and longitudinal associations between social behavior and literacy achievement in a sample of low-income elementary school children. *Child Development*, *77*(1), 103–117.

National Research Council. (1993). *Understanding and preventing violence*. Washington, DC: National Academy Press.

Ollendick, T. H., Weist, M. D., Borden, M. G., & Greene, R. W. (1992). Sociometric status and academic, behavioral, and psychological adjustment: A five-year

longitudinal study. *Journal of Consulting and Clinical Psychology*, *60*, 80–87.

Olweus, D. (1983). Low school achievement and aggressive behavior in adolescent boys. In D. Magnusson & V. Allen (Eds.), *Human development: An interactive perspective* (pp. 353–365). New York: Academic Press.

Payton, J., Weissberg, R. P., Durlak, J. A., Dymnicki, A. B., Taylor, R. D., Schellinger, K. B., et al. (2008). *The positive impact of social and emotional learning for kindergarten to eighth-grade students: Findings from three scientific reviews*. Chicago: Collaborative for Academic, Social, and Emotional Learning.

Prothrow-Stith, D. (2007). A major step forward in violence prevention research. *American Journal of Preventive Medicine*, *33*(2S), S109–S111.

Raver, C. C., Garner, P. W., & Smith-Donald, R. (2007). The roles of emotion regulation and emotion knowledge for children's academic readiness: Are the links causal? In R. C. Pianta, M. J. Cox, & K. L. Snow (Eds.), *School readiness and the transition to kindergarten in the era of accountability* (pp. 121–147). Baltimore: Paul H Brookes.

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

Reynolds, C. R., & Kamphaus, R. W. (1998). *Behavioral assessment system for children*. Circle Pines, MN: American Guidance Service.

Rosenfeld, E., Huesmann, L. R., Eron, L. D., & Torney-Purta, J. V. (1982). Measuring patterns of fantasy behavior in children. *Journal of Personality and Social Psychology*, *42*, 347–366.

Selman, R. (2003). *The promotion of social awareness: Powerful lessons from the partnership of developmental theory and classroom practice*. New York: The Russell Sage Foundation.

Selman, R. L., Beardslee, W., Schultz, L. H., Krupa, M., & Podoresky, D. (1986). Assessing adolescent interpersonal negotiation strategies: Toward the integration of structural and functional models. *Developmental Psychology*, *22*, 450–459.

Shinn, M. R., Walker, H. M., & Stoner, G. (2002). *Interventions for academic and behavior problems II: Preventive and remedial approaches*. Washington, DC: NASP.

Shonkoff, J. P., & Phillips, D. A. (eds). (2000). *From neurons to neighbourhoods: The science of early childhood development*. Washington, DC: National Academy Press.

Smolkowski, K., Biglan, A., Barrera, M., Taylor, T., Black, C., & Blair, J. (2005). Schools and Homes in Partnership (SHIP): Long-term effects of a preventive intervention focused on social behavior and reading skill in early elementary school. *Prevention Science*, *6*(2), 113–125.

Snijders, T., & Bosker, R. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. Thousand Oaks, CA: Sage.

Social and Character Development (SACD) Research Consortium (2008). *Efficacy of school-wide programs to promote social and character development and reduce problem behavior in elementary school children*. National Center for Education Research, Institute of Education Sciences, U.S. Department of Education. Washington, DC: U.S. Government Printing Office.

Tolan, P., Gorman-Smith, D., & Henry, D. (2004). Supporting families in a high-risk setting: Proximal effects of the SAFE children preventive intervention. *Journal of Consulting and Clinical Psychology*, *72*(5), 855–869.

Tremblay, R. E., Pagani-Kurtz, L., Masse, L. C., Vitaro, F., & Pihl, R. O. (1995). A bimodal preventive intervention for disruptive kindergarten boys: Its impact through mid-adolescence. *Journal of Consulting and Clinical Psychology*, *63*, 560–568.

Van Lier, P., Muthen, B., van der Sar, R. M., & Crijnen, A. (2004). Preventing disruptive behavior in elementary school children: Impact of a universal classroom-based intervention. *Journal of Consulting and Clinical Psychology*, *72*(3), 467–478.

Vitaro, F., Brendgen, M., & Tremblay, R. E. (2001). Preventive intervention: Assessing its effects on the trajectories of delinquency and testing for mediational processes. *Applied Developmental Science*, *5*(4), 201–213.

Vuchinich, S., Flay, B. R., Aber, J. L., Bickman, L., Farmer, T., Gottfredson, G., et al. (under review). The effects of person mobility on the design and analysis of place-randomized trials. *Prevention Science*.

Walker, H. M., & Shinn, M. R. (2002). Structuring school-based interventions to achieve integrated primary, secondary, and tertiary prevention goals for safe and effective schools. In M. Shin, H. Walker & G. Stoner (Eds.), *Interventions for academic and behavior problems II: Preventive and remedial approaches* (pp. 1–25). Washington, DC: National Association of School Psychologists.

Webster-Stratton, C., Reid, M. J., & Stoolmiller, M. (2008). Preventing conduct problems and improving school readiness: Evaluation of the incredible year teacher and child training programs in high risk schools. *The Journal of Child Psychology and Psychiatry*, *49*(5), 471–488.

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

Wentzel, K. R. (1991). Relations between social competence and academic achievement in early adolescence. *Child Development*, *62*, 1066–1078.

Williams, S., & McGee, R. (1994). Reading attainment and juvenile delinquency. *Journal of Child Psychology and Psychiatry*, *35*, 441–461.

Wilson, S. J., Lipsey, M. W., & Derzon, J. H. (2003). The effects of school-based intervention programs on aggressive behavior: A meta-analysis. *Journal of Consulting and Clinical Psychology*, *71*(1), 136–149.

Zins, J. E., Weissberg, R. P., Wang, M., & Walberg, H. J. (eds.) (2004), *Building academic success on social and emotional learning: What does the research say?* New York: Teachers College Press.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Table S1. Psychometric and Descriptive Characteristics of Dependent Variables at Wave 1 (Baseline), Wave 2, Wave 3, and Wave 4 by Intervention and Control Schools