ORIGINAL PAPER

Promoting Children's Prosocial Behaviors in School: Impact of the "Roots of Empathy" Program on the Social and Emotional Competence of School-Aged Children

Kimberly A. Schonert-Reichl · Veronica Smith ·
Anat Zaidman-Zait · Clyde Hertzman

Abstract This study examines the effects of the *Roots of Empathy* (ROE) program on children's social-emotional competence. ROE is a theoretically derived universal preventive program that focuses on decreasing children's aggression and facilitating the development of their social-emotional understanding and prosocial behaviors. The program has as its cornerstone monthly visits by an infant and his/her parent(s) that serve as a springboard for lessons on emotion understanding, perspective taking, caring for others, and infant development. The study included a quasi-experimental control-group pretest–posttest, multi-informant design with 585 4th- to 7th-grade children from 28 classrooms. Outcome measures included self-reports of understanding of infant distress, empathy, and perspective taking, and peer and teacher reports of prosocial and aggressive behaviors. Measures assessing implementation were also included. Children in the ROE intervention classrooms showed significant improvement across several of the domains assessed: self-reports of causes for infant crying, peer nominations of prosocial behaviors, and teacher reports of proactive and relational aggression. Self-reported empathy and perspective taking showed no significant changes. According to ROE instructors' diaries assessing implementation, students in the experimental condition were exposed to all or most of the ROE curriculum. These findings support and extend recent research examining the positive impacts of classroom-based social and emotional learning (SEL) programs on children's social development and behavioral adjustment.

Keywords Primary prevention · Social and emotional competence · Prosocial and aggressive behaviors · School-aged children

K. A. Schonert-Reichl (✉)
Department of Educational and Counselling Psychology, and Special Education, University of British Columbia, 2125 Main Mall, Vancouver, BC V6T 1Z4, Canada
e-mail: kimberly.schonert-reichl@ubc.ca

V. Smith
Department of Educational Psychology, University of Alberta, Edmonton, AB, Canada

A. Zaidman-Zait
Human Early Learning Partnership, College of Interdisciplinary Studies, University of British Columbia, Vancouver, BC, Canada

C. Hertzman
School of Population and Public Health, University of British Columbia, Vancouver, BC, Canada

Introduction and Background

Identifying the factors that predict children's social and emotional competence has garnered considerable empirical attention in recent decades, in part, because such knowledge affords an understanding of the mechanisms and processes that may propel children away from or toward maladaptation: strategic prerequisites for preventive efforts to guide children onto successful developmental paths (Izard, 2002; Staub, 2003). Most of the studies conducted in this domain have focused on the identification of factors that predict negative outcomes, such as aggression and mental illness, with relatively little attention given to delineating the ways in which children's prosocial behaviors and characteristics can be advanced. However, recent years have witnessed a shift from a preoccupation with repairing weaknesses to the enhancement of positive qualities and in preventing or heading off problems *before* they arise (Damon, 2004; Damon & Gregory, 2003;

Huebner & Gilman, 2003; Seligman & Csikszentmihalyi, 2000). Implicit in this trend is the assumption that educational interventions can be designed not to only deter children's problem behaviors but to foster children's strengths, positive behaviors, and resiliency (e.g., Masten & Motti-Stefanidi, 2009).

School-based prevention efforts have been heralded as an effective and cost-effective way to promote children's positive development and stave off an upward trajectory of mental illness and aggressive behaviors (Greenberg, 2010; Weissberg & Greenberg, 1998)—problems that have been shown to increase over the course of the elementary school years (e.g., Aber, Brown, & Jones, 2003; Farmer & Xie, 2007; Murray-Close, Ostrov, & Crick, 2007; Neal, 2010). Bolstered by evidence that empirically based curricula can prevent the onset of problem behaviors, such as anxiety and aggression, and decrease antisocial behavior (e.g., Ang & Hughes, 2001; Cook, Murphy, & Hunt 2000; Durlak, Weissberg, Dymnicki, Taylor, & Schellinger, 2011; Fonagy et al., 2009; Gresham, Cook, Crews, & Kern, 2004; Jones, Brown, & Aber, 2011; Lösel & Beelmann, 2003), many school districts throughout North America have strengthened their efforts to include programs that promote students' social and emotional competence. Because virtually all children go to school during key developmental years, this context is important when fostering children's positive social and emotional development. Due to many competing demands and time constraints, however, schools need to prioritize their efforts and implement curricular approaches that are evidence-based and both time- and cost-effective.

Although there has been increased theoretical and research attention aimed at delimiting the social and emotional factors that significantly impact children's psychological well-being and behavioral adjustment (Denham, 1998; Hymel, Schonert-Reichl, & Miller, 2007; Saarni, 1999), to date, few studies exist that examine the effectiveness of interventions specifically targeted to promoting children's prosocial behaviors—those behaviors and characteristics such as cooperativeness, helpfulness, trustworthiness, and kindness (for exceptions, see Battistich, Solomon, Watson, Solomon, & Schaps, 1989; Conduct Problems Research Group, 1999; Denham & Burton, 1996; Greenberg, Kusché, Cook, & Quamma, 1995). As noted by Izard (2002), "Although arousal of any negative emotion may provide the basis for strong learning, the induction and utilization of the emotions relating to empathy, sympathy, and prosocial behavior have the greatest relevance for preventive interventions" (p. 801). This present study contributes to ongoing scholarship in the school-based promotion of children's social and emotional competence by reporting empirical impacts of a unique social and emotional learning (SEL) intervention—the "Roots of Empathy" (ROE)—on a sample of 4th- to 7th-grade children's social-emotional understanding and social behaviors after exposure to the intervention.

Description of the Intervention

ROE Curriculum

ROE is a program for children in the kindergarten through 8th grade (Gordon, 2005). It was developed in 1996 and has been implemented to children in urban, rural, and Aboriginal communities in Canada, internationally in the United States, New Zealand, the Isle of Man, Northern Ireland, Republic of Ireland, and Scotland with pilot implementations elsewhere (Kendall et al. 2006; Yau, 1999). Since the program began, it has reached over 325,000 children worldwide.

The primary goals of ROE are to (1) develop children's social and emotional understanding, (2) promote children's prosocial behaviors and decrease their aggressive behaviors, and (3) increase children's knowledge about infant development and effective parenting practices. ROE is a 9-month program that has as its cornerstone a monthly classroom visit by an infant and his/her parent(s) whom the class "adopts" at the beginning of the school year. It is during these monthly visits that children learn about the baby's growth and development via interactions and observations with the baby. Each month, an ROE program instructor[1] visits his/her participating ROE classroom three times, once for a pre-family visit, another time for the visit with the parent and infant, and finally, a postfamily visit. The lessons for the bimonthly visits from the instructor are designed to foster empathy, emotional understanding, and problem-solving skills through discussion and activities in which the parent–infant visit serves as a springboard for discussions about perspective taking, caring for others, infant development, and effective parenting practices. Each lesson is designed to capitalize on shared observations from the family visit. Lesson plans and accompanying activities are scripted to match the age of the baby and are calibrated to the students' level of development. (Four different curriculums are available: kindergarten, 1st to 3rd grade, 4th to 6th grade, and 7th and 8th grade). In addition to the three classroom visits per month by the instructor, each classroom teacher utilizes the lessons and ideas presented during the specific ROE lesson within the general education program.

Each ROE lesson provides opportunities to discuss and learn about different dimensions of empathy, in particular emotion identification and explanation, perspective taking,

---

[1] See the Method section for a full description of instructor training and selection.

and emotional sensitivity. Across various lessons, children are invited to identify the emotions of the baby and to provide explanations for those emotions. Children then become engaged in lessons through stories, art projects, and general classroom activities in which they reflect and discuss their own emotions and the emotions of others. For theme 3 (Caring and Planning for the baby), for instance, in the pre-family visit, the instructor reads the book *Sasha and the Wriggly Tooth* (by R. Tregebov, 1993) to the children. Afterward, the ROE instructor leads a discussion about the mixed feelings that can ensue when one looses a tooth (e.g., "happy to be getting a visit by the tooth fairy," "embarrassed because you may look or talk funny with a missing tooth"). In the subsequent parent and infant visit, children are provided opportunities to perspective-take via asking questions of the parents about their feelings about their infant's teething experience (e.g., "how does it feel to see your baby in pain," "what do you do to help your baby feel better?").

In keeping with other comprehensive social and emotional competence promotion programs, embedded within the ROE program are explicit components aimed at creating an ecology in the classroom environment in which belonging, caring, collaboration, and understanding others are emphasized (Cohen, 2001; Noddings, 1992). Throughout each of the lessons are opportunities for children to be engaged in activities that benefit their baby. For example, in one lesson, children create a class recording of nursery rhymes and songs for their baby, and in another lesson, at the end of the school year, the children create a "wishing tree" whereby each child in the classroom writes a wish that she or he has for the baby's life. Such activities, specifically those in which individuals work collectively on projects that benefit others, have been shown elsewhere to promote altruism and a prosocial value orientation (Battistich, Solomon, Watson, & Schaps, 1997; Noddings, 1992; Staub, 1988, 2003).

The ROE Program's Theory of Change

ROE is based on empirical understandings of the processes and mechanisms that elicit both aggression and prosocial behaviors, as well as theoretical models that guide related programs targeting a range of social behaviors (see Conduct Problems Research Group, 1999; Consortium on the School-Based Promotion of Social Competence, 1996). The ROE curriculum model draws on research and theory in developmental psychology, suggesting that *emotion processes* and social understanding play critical roles in children's interpersonal relationships and social behaviors (e.g., Izard, 2002; Shipman, Zeman, Penza, & Champion, 2000), such that "emotions form the motivational bases for empathy and prosocial behavior" (Izard, Fine, Mostow, Trentacosta, & Campbell, 2002, p. 761). In particular, ROE

draws on the *functionalist* approach to emotions (Campos, Mumme, Kermoian, & Campos, 1994), wherein emotion understanding and expressivity are seen as playing central roles in the establishment and maintenance of children's interpersonal relationships (Saarni, 1999). Empathy, in particular, is core to the ROE curriculum because it is central to the genesis and enactment of prosocial behavior. Indeed, there is a great deal of empirical evidence indicating that not only empathy leads individual to desist from aggression but also the ability to empathize is central to the development of positive social relationships and caring and kind behaviors (Eisenberg & Miller, 1987; Eisenberg, Fabes, & Spinrad, 2006). Feshbach (1979) has put forth a three-component model in which empathy is conceptualized as comprising two *cognitive* components and one *affective* component: the ability to discriminate and label affective states in others; the ability to assume the perspective and role of another person; and emotional responsiveness (i.e., the affective ability to experience emotions). This framework provides a conceptual basis for the content of the ROE lessons and an organizing scheme for the measures utilized in this evaluation.

The Present Study

The current research was designed to extend understanding of the school-based promotion of children's social-emotional competence in several ways. First, we evaluated the effectiveness of ROE on the development of multiple dimensions of children's emotional understanding. More specifically, we included assessments of children's self-reported empathy and perspective taking along with assessments of their knowledge of emotion causes and knowledge of emotion management/regulation strategies using the crying infant paradigm—a paradigm that has previously demonstrated significant associations with empathy in school-aged children (Catherine & Schonert-Reichl, 2011). According to the functionalist theory of emotion (Barrett & Campos, 1987; Campos et al., 1994), the development of emotion regulation skills enables children to adapt successfully within their social environment because such skills assist children in their ability to cope with their own emotions as well as the emotional displays of others (e.g., Rubin, Coplan, Fox, & Calkins, 1995). The development of emotional regulation/management is posited to arise through socialization practices that give children opportunities to learn how to label and interpret emotions, when emotion expression is warranted, and how to manage emotion arousal (Lewis & Michalson, 1983)—skills that are specifically targeted in the ROE program curriculum. To our knowledge, no previous studies have examined the effects of an intervention on children's understanding of infant emotions.

Second, cognizant of recent research indicating that children's prosocial behaviors are important correlates of empathy (e.g., Eisenberg, Fabes, & Spinrad, 2006; Schonert-Reichl, 1993; see Schonert-Reichl & Oberle, 2011 for a review) and intellectual competencies such as academic performance (e.g., Caprara, Barbanelli, Pastorelli, Bandura, & Zimbardo, 2000; Wentzel, 1993), we also included outcome measures that assessed a range of prosocial behaviors. These behaviors such as sharing, helping, cooperation, and kindness are particularly important to examine in the context of an evaluation of a school-based social-emotional competence promotion program because these behaviors are considered to be hallmarks of social competence in childhood and adolescence (Wentzel, Filisetti, & Looney, 2007). In addition, it is well established that these behaviors have been related theoretically and empirically to mental health and well-being (Hay, 1994) and other forms of social competence such as peer acceptance and positive social relationships (Bukowski & Sippola, 1996; Oberle, Schonert-Reichl, & Thomson, 2010; Schonert-Reichl, 1999).

Third, in addition to assessing the effects of the ROE program on children's prosocial behaviors, we also examined a broader corpus of aggressive behaviors than typically assessed in evaluations of interventions, going beyond the usual exclusive focus on physical aggression. Historically, researchers have often focused on overt forms of aggression, such as hitting, pushing, and verbal threats. However, recent theoretical and methodological advances in the study of aggression have revealed the existence of a wider range of aggressive behaviors that can be distinguished with regard to their nature (e.g., overt, covert), function (e.g., social dominance), and underlying social-cognitive processes (e.g., Arsenio & Lemerise, 2001; Crick & Dodge, 1999; Sutton, Smith, & Swettenham, 1999). Dodge and Coie (1987) and others (e.g., Vitaro, Brendgen, & Tremblay, 2002), for instance, have suggested two types of aggressive behavior that can be distinguished by their underlying function and motivation: proactive and reactive aggression. Proactive aggressive behavior is instrumental and involves attempts to influence or coerce others through aversive means in an unprovoked situation, requiring neither provocation nor anger. Proactive aggression has frequently been described as "cold-blooded" because proactively aggressive individuals typically "… view aggression as an effective and viable means for obtaining social goals" (Crick & Dodge, 1999, p. 998). In contrast, reactive aggressive behavior has been described as affective, impulsive, defensive, and "hotheaded," involving angry outbursts in response to actual or perceived threats or provocations. Distinguishing between proactive and reactive aggression is important on both theoretical and empirical grounds and may also be important for the design of prevention and intervention programs.

In the present study, we also assessed intervention effects on another form of aggression that has risen to prominence in the theoretical and empirical literature in recent years—a form of aggression that includes behaviors that are more covert and indirect in nature and are aimed at damaging the target's social relationships as opposed to inflicting bodily harm. Distinct from physical aggression, different terms have been used to describe this more subtle form of aggression, including indirect aggression (Lagerspetz, Bjorkvist, & Peltonen, 1988), social aggression (Cairns, Cairns, Neckerman, Gest, & Gariépy, 1988), and relational aggression (Crick & Grotpeter, 1995). In the present investigation, we use Crick's (1996) definition in which relational aggression refers "… to harming others through purposeful manipulation or damage to their peer relationships (e.g., using social exclusion as a form of retaliation)." (p. 2317). Several recent studies have found that this form of aggression becomes increasingly common over the course of the elementary school years (for a review, see Neal, 2010), with one study showing a linear increase in relational aggression among 4th-grade girls over one calendar year (Murray-Close et al., 2007; see Neal, 2010 for a review). A recent meta-analysis of 148 studies on child and adolescent direct and relational aggression found gender effects for direct aggression (favoring boys) and negligible gender effects for relational aggression (Card, Stucky, Sawalani, & Little, 2008). This finding of trivial gender differences in relational aggression challenges current portrayals of this form of aggression being more enacted by girls. Unfortunately, few studies to date have examined how a classroom-based intervention can reduce relational aggression. To our knowledge, this is one of the first evaluations of a primary preventive intervention aimed at promoting social and emotional competence in which the dimensions of physical, proactive, reactive, and relational aggression were examined concomitantly in one study.

Fourth, we assessed children's prosocial and aggressive behaviors via multi-informants. Specifically, teachers and peers were selected as informants as they are typically used as the main informants in research on children's social behaviors in school (e.g., Ladd & Troop-Gordon, 2003). Moreover, Cappella and Weinstein (2006), in their evaluation of a social aggression prevention program (SAPP) for 5th-grade girls, noted disparate intervention effects depending on whether the informant was the teacher or a peer, rendering multiple informants important when investigating the effectiveness of an SEL intervention. For our peer assessments, we utilized a peer behavioral assessment method whereby children nominate their classmates who fit particular behavioral characteristics. Following the methodology outlined by Parkhurst and Asher (1992), we used single, precisely focused peer

assessment items to measure specific prosocial and aggressive behaviors. Using single item for each behavioral dimension is an approach that has been frequently used in research on children's peer relationships (Coie & Dodge, 1988). Peer behavioral assessments provide data from multiple raters of a child's behavior and may also provide information regarding a wider corpus of observations in settings beyond the classroom. Peer reports increase the power of assessment because many peers are involved in the evaluation rather than just a single teacher. Moreover, because peers can act as "participant observers," they are in a particularly significant position to provide assessment information to the researcher about the social behaviors of their classmates (Younger, Schneider, Wadeson, Guirguis, & Bergeron, 2000). For the present investigation, children provided ratings of their classmates on dimensions of prosocial and antisocial/aggressive behaviors.

Finally, recognizing that there are no "true control groups" in school-based experimental or quasi-experimental research (Oakley et al., 2006), and following recommendations by Greenberg (2004) and Durlak and DuPre (2008), we obtained information in order to document the experiences and interventions received by children in both the intervention and control classrooms, other than the ROE program.

Aims and Hypotheses

This evaluation study is a quasi-experimental control-group pretest–posttest, multi-informant design. We matched intervention with control classes based on average age, gender, and English as a second language (ESL). We hypothesized that when compared to control children, ROE program children would show significant and positive changes from pretest to posttest on our measures of understanding of infant crying, empathy, perspective taking, and teacher- and peer-rated social behavior. Our dual focus on increasing social and emotional competence and reducing aggression was warranted due to empirical evidence indicating that prognosis for children is poorest when they demonstrate a combination of low social-emotional competence alongside aggressive behavior (CPPRG, 1999; Tolan, Guerra, & Kendall, 1995). Moreover, cognizant of recent research indicating increases in children's aggression during the late elementary school years (e.g., Aber et al., 2003; Murray-Close et al., 2007), we examined both the magnitude and direction of changes in aggressive and prosocial behaviors across the school year (Wilson, Lipsey, & Derzon, 2003).

A second purpose of our study was to examine the fidelity of the ROE program when implemented in a "real-world" setting delivered in regular elementary school classrooms with minimal support from the program developers (Domitrovich & Greenberg, 2000; Durlak & DuPre, 2008). Central to our question was the degree to which the ROE program's objectives and procedures were put into everyday practice in the classroom. Hence, program implementation was monitored with lesson diaries and a self-report survey of lesson implementation to determine the degree to which the program lessons were delivered with fidelity.

Method

Participants

Children

Participating children were drawn from regular education classrooms in the 4th through 7th grades of 28 public elementary schools across two Canadian cities (18 schools in Vancouver, BC; 10 schools in Toronto, Ontario). Schools were chosen based on their representativeness of the range of socioeconomic and racial/ethnic diversity of students in the school district. Fourteen of the schools had a classroom in which the ROE program was implemented. To avoid contamination, 14 control classrooms were selected from 14 schools where no ROE program was currently being implemented. The program classrooms were matched as closely as possible with respect to gender of teacher, class size, grade level, and child characteristics (e.g., race/ethnic and gender composition of children, percent of children whose home language was different than English). Active consent for the assessment phase of the intervention was obtained using parental permission forms that were translated into seven languages (i.e., Chinese, Vietnamese, Punjabi, Spanish, Polish, Russian, and Korean). All children recruited were from regular classrooms and hence were deemed to be competent to read and write English. Prior to providing children with the parental permission slips, either a trained research assistant or the principal investigator of the research project provided a 15-min presentation to each participating class, describing the study in age-appropriate language and answering children's questions. Permission rates were high, with an average consent rate of 92%.

The recruited sample included 638 children attending school in the fall of 4th to 7th grades. Fifty-three children (8%) were lost to the study because they moved away prior to posttesting. Attrition was comparable across the ROE program and control groups [$ns = 34$ and 19, respectively; $X^2 (1, N = 53) = 2.74$, ns]. Further analyses revealed that there were no significant differences in any of the pretest scores between those children who attrited and those who remained, $Fs < 1$.

The final sample was comprised of 585 children (280 girls, 305 boys) in the 4th to 7th grades who remained in their respective intervention ($n = 306$) or control ($n = 279$) classrooms throughout the one school-year intervention and were available for pre- and posttesting. Ages ranged from 8 years, 8 months to 12 years, 8 months at pretest, with a mean of 10 years, 7 months (SD = .87). With regard to first language learned, 40% of the participants reported English as their first language, 39% reported Chinese, 8% reported Vietnamese, and the remaining 13% reported other languages (e.g., Punjabi, Spanish, Polish, Russian, Tamil, and Filipino). These percentages reflect the ethnic diversity of the cities in which the research took place. Because the design of this study was quasi-experimental, it was important to demonstrate the similarity of the groups across background and potentially contaminating variables (Cook & Campbell, 1979). To examine for potential differences between our two groups, T tests were utilized for continuous variables and chi-square analyses for categorical variables. Analyses revealed no differences between intervention and control classrooms with regard to age, gender, or first language learned ($ps > .25$), suggesting that our matching process was successful.

Teachers

Teachers in both the ROE and control group completed a questionnaire that assessed the following dimensions: sex, race/ethnicity, and teaching background (years of teaching experience). Of the 28 teachers who participated in the research, 27 completed background questionnaires (14 program and 13 control). Analyses of background characteristics of the intervention and control teachers revealed no significant differences. The majority of the teachers were women (71% program and, 69% control), most identified themselves as European Canadian (86% program and 84% control), and the majority of teachers in both groups had 10 or more years of teaching experience (66% program and 70% control).

ROE Program Instructors

ROE instructors also completed a background questionnaire that allowed for comparisons between instructors on the following dimensions: sex, ethnicity, and experience teaching the ROE program. The majority of ROE instructors were women (93%) and included one counselor, three youth and family workers, one community school coordinator, four community health nurses, one parent, and four non-enrolling teachers. Years of experience in education or a similar position ranged from 5 to 33 with a mean of 17.5 (SD = 9.7). Exactly half of the instructors had previous experience of delivering the ROE program.

The Intervention

The ROE program began in late fall, and all 26 lessons across the nine themes (i.e., Meeting the Baby, Crying, Caring and Planning for the Baby, Emotions, Safety, Sleep, Communication, Who Am I?, Goodbye, and Good Wishes) were completed by the end of the school year across all of the participating classrooms. The time for each classroom lesson ranged from 30 to 45 min. Each instructor utilized the ROE intermediate-grade curriculum (Gordon, 2001) to implement the program and followed the guidelines explicated in each lesson. Each theme involved three lessons: (1) a pre-family visit that introduced the theme; (2) a class visit by the "classroom baby" and his/her parent(s)/ caregiver(s), where the instructor directed interactions and observations with baby and parent(s) to the students to enhance learning about infant development; and (3) a postfamily visit where the students were guided to discuss the baby visit and conclude the theme. In consonance with the goals of the ROE stated earlier, each lesson fostered these goals through discussion and activities in which the parent–infant visit served as a springboard for discussions about understanding feelings, infant development, and effective parenting practices.

Because "pre-implementation training and on-site coaching enable service providers to implement programs with greater integrity and to have more positive effects on the children they serve" (Weissberg & Greenberg 1998, p. 925), ROE instructors underwent an intensive three-day training session, prior to program implementation, wherein they were provided with program goals, as well as its theoretical and empirical underpinnings. Instructors were provided with a training manual (Gordon, 2000) delineating the theory and research guiding each ROE lesson, and a curriculum manual in which plans for each lesson were thoroughly described (Gordon, 2001). The training session included discussions on infant and child social and emotional competence as well as lecture, video, and readings, and role-plays of curriculum instructional techniques. At the end of the three-day training session, all instructors were tested on the program and its components, and those who passed became certified ROE program instructors. ROE instructors were mentored throughout the course of the program implementation. Mentors, individuals with extensive experience implementing the ROE program, were trained and certified as "Roots of Empathy" leaders. They visited the classrooms during program implementation to provide support and feedback to the instructor, as well as helping trouble-shoot special situations and facilitating instructors' learning from their programs and each other. All instructors received a "booster session" mid-way through the school year. This session allowed an opportunity for instructors to share their experiences with one

another, ask questions, and obtain assistance for any issues that they confronted with regard to program implementation. Finally, all school principals and teachers were provided with explicit guidelines to follow in order to maintain the integrity of ROE program delivery.

As can be surmised, intervention fidelity was ensured by manualization of all components, mid-way booster session, and ongoing supervision of ROE instructors via the ROE mentor. Outside interventions were neither encouraged nor discouraged and were assumed to occur at the same rate for intervention and control groups. The control condition was a "treatment as usual" comparison that included regular school prevention programs to the extent that schools chose to use them.

Procedure

Data were collected two to four weeks prior to exposure to the intervention in October (time 1) and two to four weeks following the end of the intervention in June (time 2). At both time periods, self-report, peer-report, and teacher-report instruments were used to assess participants' emotion understanding, empathy and perspective taking, and prosocial and aggressive/antisocial behaviors. Child self-report and peer behavioral assessment measures were group-administered to children in their classrooms by research assistants, while a second research assistant circulated to monitor children's survey completion and answer any of their questions. Children who were non-consented worked on an alternative activity related to the regular school curriculum. Teachers completed ratings of children's social behaviors at pretest and posttest. They were given a $200.00 honorarium for their participation in the study.

Measures

Assessment of Implementation

To assess program adherence and participant responsiveness, ROE instructors were asked to complete a monthly "diary" in which they indicated the total number of lessons completed (exposure), the percentage of curricular components completed within each lesson (adherence), and a rating of the quality of student engagement (participant responsiveness) during each lesson on a 5-point scale ranging from 1 ("not engaged") to 5 ("very engaged").

In order to garner information about program differentiation, we also collected data assessing the degree to which both ROE and control teachers implemented SEL programs and practices outside of those associated with the ROE program. Specifically, we asked teachers to rate the

frequency with which they implemented SEL activities (other than those associated with the ROE program) in their classroom across the school year on a 5-point scale (1 = *not at all* to 5 = *a lot*). Additionally, we asked teachers in both ROE and control classroom to report on their training (e.g., workshops) on SEL as well as their beliefs about the importance of teaching activities to promote students' social and emotional learning (SEL). With regard to this latter dimension, we asked teachers to rate on a 5-point scale (1 = *not very important* to 5 = *very important*) "What importance do you give to activities or teaching strategies that influence students' social and emotional learning in your classroom?"

Direct Child Assessments: Demographic Information

At pretest, children provided information about their sex, ages, grade levels, first language learned, and family composition.

Direct Child Assessments: Understanding of Infant Crying

Children's understanding of infants' crying was assessed via a task adapted from the infant facial expression of emotion (IFEEL) pictures (Emde, Osofsky, & Butterfield, 1993; modified by Catherine & Schonert-Reichl, 2011), designed to assess an individual's understanding of the causes of infants' cries and knowledge of behavioral strategies for helping a crying infant. Data were collected using a booklet of self-report questionnaires. On one page, children were shown a black-and-white photograph of an infant crying. The picture was cropped so that only the face of the infant was shown and was gender-neutral, with no apparent clothing or identifiers depicting either a male or female infant. On a separate page, children were presented with two questions designed to address two conceptual categories, tapping their thoughts about (a) *causes* for an infant's distress (i.e., "Why does this baby cry?") and (b) emotion regulation *strategies or action responses* to the display of emotions (i.e., "What are some ways to help a crying baby?"). Our assessment focused on two dimensions of emotion understanding— causality of emotions and emotion regulation strategies— that, based on the ROE curriculum, we believed would be of particular importance in determining the ROE program's effectiveness in advancing children's understanding of infant emotions. Children were not limited with regard to the number of causes and strategies that they could offer. In some cases, several responses produced by the children were infrequently used, did not fit the definitions of the other categories, or did not address the question properly (e.g., responding "baby is wrestling/

fighting" as a cause). These responses were excluded from further analyses.

Responses to the question assessing children's understanding for *causes* for infants' cries were coded following a modified version of emotion casual explanations' coding developed by Fabes, Eisenberg, Nyman, and Michealieu (1991). Each response was first assigned to two superordinate categories depending on whether the child's responses indicated references to *internal* states (e.g., "the baby is sad because he wants his mom") or references to *external* behaviors or actions (e.g., "the baby is crying because somebody took his toy"). References to internal states were then further subdivided into those references that referred to (a) an emotional state (e.g., feelings—"the baby is upset," "the baby is scared") or (b) a physical state (e.g., being sick or hurt, physical needs—"the baby is thirsty," "the baby is tired," "the baby has an ear infection").

Coding of responses to the question assessing *emotion regulation strategies* (i.e., "What are some ways to help a crying baby?") was based on past research regarding the categorization of emotion management processes for regulating emotional distress (Shipman, Zeman, Penza, & Champion, 2000). Specifically, children's responses were coded into one of the following three categories: 1) provide *physical support* (e.g., "give it a bottle," "change the baby's diaper"), 2) provide *comfort object/environment* (e.g., "give the baby a pacifier/soother," "put the baby in the swing"), and 3) provide *social/relational support* (e.g., "put the baby in his/her mother's lap," "make it feel loved"). The children's answers were scored from the questionnaires. One coder with no additional information about the child (i.e., the child's responses to questions about other emotions or the child's gender or group assignment) coded all of the children's responses. A second coder, also with no information about the child or other responses, independently coded a random selection of 15% of children's responses. Inter-rater reliability was fairly high for both the emotion causes category (K = .82) and the emotion regulation strategies category (K = .92). All scores in data analyses were based on adjudication of two coders' decisions where there had been disagreements.

Direct Child Assessments: Empathic Concern and Perspective Taking

Participants' empathy-related responding was assessed via the Interpersonal Reactivity Index (IRI; Davis, 1983), a self-report measure comprising four 7-item subscales (empathic concern, perspective taking, fantasy, personal distress), each of which taps a separate dimension of empathy. For the purposes of the present study, only the empathic concern and perspective-taking subscales were used. The empathic concern scale assesses the tendency to feel compassion and concern for other individuals (e.g., "I often feel sorry for people who don't have the things I have"), whereas the perspective-taking subscale measures the tendency to consider things from others' viewpoints (e.g., "Sometimes I try to understand my friends better by imagining how they think about things"). Participants rated each item on a 5-point rating scale (ranging from $1 = never$ to $5 = very often$). Scores were computed by averaging item scores within subscales so that higher scores signified greater empathic concern and perspective taking, respectively. Supportive evidence for the construct validity of the empathic concern and perspective taking subscales of the IRI has been obtained in previous research (Davis, 1983), including significant correlations with related constructs in expected directions (Schonert-Reichl, 1993). For the present study, both scales demonstrated adequate internal consistency both at pretest and at posttest. Specifically, Cronbach alphas were .80 at pretest and .86 at posttest for the empathic concern subscale and .74 at pretest and .80 at posttest for perspective-taking subscale.

Peer Report of Child: Peer-Rated Prosocial and Aggressive/Antisocial Behaviors

We used a peer nomination and behavioral assessment method to assess children's prosocial and antisocial/aggressive behaviors. Specifically, unlimited and cross-gender peer nominations were used to obtain independent assessments of children's social behavior. Six types of *prosocial* behaviors ("shares," "cooperates," "helps other kids when they have a problem," "kind," "understands other kids' point of view," and "fair"), and two types of *aggressive/antisocial* behaviors ("starts fights" and "breaks rules") were assessed. For each behavior, children were provided with a list of all of their classmates participating in the research and were asked to circle the names of their classmates on each list who fit the behavior description; children could circle as many or as few names as they wanted. For each question, the percentage of nominations each child received was computed by dividing the number of nominations received by the total number of participating children in the classroom. For each of the prosocial and aggressive/antisocial characteristics and behaviors, children's nominations were standardized within each classroom, and a proportional nominations score was calculated per child for each of the behaviors/characteristics.

Teacher Report of Child: Children's Prosocial and Aggressive Behaviors

Teachers rated children's aggressive and prosocial behaviors on corresponding subscales of the Child Behavior

Scale (CBS; Ladd & Profilet, 1996), an instrument designed to assess children's behaviors with peers at school. Both of these subscales have evidence supporting their reliability and validity with school-age children (see Ladd & Profilet, 1996). The aggressive and prosocial behavior subscales from the CBS demonstrated adequate internal reliability in the present study at pretest and posttest, as reported below. The CBS Aggressive subscale contained seven items that referred to physical (e.g., "Kicks, bites, or hits other children") and verbal (e.g., "Threatens other children") aggression, and the CBS Prosocial Behavior subscale was composed of seven items tapping e.g., cooperative (e.g., "Cooperative with Peers"), and empathic and prosocial (e.g., "Shows a recognition of the feelings of others; is empathic," "Offers help or comfort when other children are upset") behavior. An additional CBS subscale included six "filler" items that are not scored. These items refer to other types of classroom behaviors (e.g., "Listens to classmates") and were included as a way to discourage teachers from forming specific hypotheses about the specific constructs being assessed. Teachers rated children's behavior with scale points labeled and defined as follows: 1 = *doesn't apply* ("child seldom displays the behavior"), 2 = *applies sometimes* ("child occasionally displays the behavior"); 3 = *certainly applies* ("child often displays the behavior").

To assess proactive, reactive, and relational aggression, items were taken from existing measures of proactive aggression (9 items; e.g., "Gets other children to gang up on a peer that he/she does not like"), reactive aggression (11 items; e.g., "When a peer accidentally hurts this child, such as by bumping into him/her, this child assumes that the peer meant to do it, and them overreacts with anger or fighting") (Dodge & Coie, 1987), and relational aggression (4 items; e.g., "When mad at someone, tries to get others to dislike that person") (Crick, 1996) and were added to the CBS to complement the range of behaviors that teachers rated. As with the CBS aggressive and prosocial subscales, teachers rated children's proactive, reactive, and relational aggression on a 3-point scale ranging from 1 (*doesn't apply*) to 3 (*certainly applies*). Items for each subscale were averaged with higher scores signifying higher levels of each construct. Supportive evidence for the reliability and validity of the scales assessing proactive, reactive, and relational aggression has been obtained (e.g., Crick, 1996; Dodge & Coie, 1987). For the present investigation, reliability, assessed via Cronbach's alpha, was found to be adequate at both pretest and posttest: Aggressive Behavior, $\alpha = .91$ and $.86$, respectively; Prosocial Behavior, $\alpha = .90$ and $.91$, respectively; Proactive Aggression, $\alpha = .77$ and $.76$, respectively; Reactive Aggression, $\alpha = .92$ and $.93$, respectively; and Relational Aggression, $\alpha = .90$ and $.87$, respectively.

Results

The results will be described in two parts. First, preliminary analyses are presented. In this section, we describe results of analyses examining baseline differences between intervention and control groups along with analyses examining implementation fidelity and dosage. Following, we present results from a series of multivariate analyses of covariance (MANOVAs) and analyses of covariance (ANCOVAs) employed to address the study's primary questions.

Preliminary Analyses

Baseline Comparisons

A series of ANOVAs was performed to check for mean pretest differences between ROE program and control children on all outcome measures. As expected, owing to the matching procedures we employed, there were no intervention-control differences at the preintervention baseline for any outcomes examined ($Fs < 1$).

Intervention Implementation

Implementation adherence, as assessed via instructors' diaries, revealed that the average number of ROE lessons taught was 25.21 (SD = 1.25) out of the total 26 lessons (range: 23 to 26 lessons), indicating an implementation level of 96%. Of the 14 ROE instructors, 11 (80%) of them reported that they implemented all 26 lessons, and the remaining three instructors reported implementing 23 of the 26 lessons. In addition to reporting the total number of lessons completed and to obtaining more in-depth information about program adherence, ROE instructors also reported the percentage of content/material that they implemented within each of the 26 lessons. Overall, ROE instructors reported implementing, on average, 95% (range from 89% to 100%) of the lesson plan content across the 26 lessons. Given the structure of the ROE program in which each of the nine themes is broken down into three components (pre-family visit, baby visit, postfamily visit), we also examined the percentage of lesson content presented for each of the three lesson types. Our findings revealed that, on average, ROE instructors implemented 93% of lesson content across the nine pre-family visit lessons (range from 85% to 100%), 95% of the content for the nine baby-visit lessons (range from 77% to 100%), and 91% of content for the 8 postfamily visit lessons (range from 79% to 100%). Thus, it appears that the implementation fidelity, assessed via ROE instructors' self-report on number of lessons completed and in the percentage of lesson content presented, was high across instructors and classrooms.

Because of the high level of implementation adherence, no interaction effects were tested with these as potential moderators.

ROE instructors also rated students' level of engagement for each of the ROE lessons on a 5-point scale ranging from 1 ("*not engaged*") to 5 ("*very engaged*"). Because we were interested in determining students' level of engagement for lessons when the baby was present and when the baby was absent, we calculated separate means for each of these lesson types. Average level of student engagement for lessons in which the baby was present ranged from 4.72 to 5.00 ($M = 4.93$, SD $= .10$), and the average level of engagement for lessons in which the baby was absent ranged from 3.72 to 4.95 ($M = 4.42$, SD $= .41$).

We next tested for differences between ROE and control teachers on their training in SEL, beliefs about SE, and frequency with which they implemented SEL activities in their classrooms throughout the school year (other than the ROE program). With regard to training in SEL and SEL beliefs, teachers in both groups were equivalent in terms of having received coursework or workshops in SEL (71% program and 70% control), and all (100%) teachers rated SEL as "important" or "very important." An independent-samples t test revealed that, on a scale ranging from 1 (*not at all*) to 5 (*a lot*), control teachers implemented significantly more SEL activities in their classrooms ($M = 3.17$, SD $= .94$) than did ROE teachers ($M = 2.21$, SD $= 1.05$), $t(24) = -2.42$, $p = .023$.

Main Analyses

Analytic Strategy

The effect of condition on child outcomes was tested using either analysis of covariance (ANOVA) or multiple analysis of variance (MANCOVA), where appropriate. A 2 (group: ROE program, control condition) $\times$ 2 (gender) model was utilized and included changes from pretest to posttest (i.e., difference scores) as the dependent variable (Zumbo, 1999; for a recent discussion regarding the utility of difference scores, see Thomas & Zumbo, 2011). Statistically comparable to performing a repeated-measures analysis, difference or "change" scores (posttest–pretest) provide an unbiased estimate of true change regardless of baseline value. Difference scores are seen as an alternative to ANCOVA (in which pretest scores are statistically controlled) when the researcher is interested in examining the *direction of change* and determining whether improvement in scores from pretest to posttest is greater in the treatment group than it is for the control group, as was the case in the present study (Tabachnick & Fidell, 2001). In all analyses, first language learned (ESL status) and age were included as covariates in

order to control for potential confounds. In order to provide information about the magnitude of program effects beyond statistical significance, we calculated effect sizes (Cohen's *d*) for each of our analyses. According to the criteria proposed by Cohen (1988), an effect size of .2 is considered small, .5 is considered to be a medium effect, and .8 is considered a large effect.[2] To control for family wise error, we set the alpha level at .01 for all tests of significance.

Understanding of Infant Crying

Table 1 provides a summary of the outcome analyses on children's understanding of the *causes* for infant crying. The effect of condition on difference scores for internal (i.e., emotion state, physical state) and external causes was tested using a series of 2 (group) $\times$ 2 (gender) ANCOVAs, with age and ESL status as covariates. As can be seen, a significant group effect was found for one of the three categories: ROE children showed significant improvements in their ability to spontaneously generate causes for infants' crying that reflected attention to the *emotions* of the baby (i.e., making reference to the infant's feelings such as "the baby is crying because he is upset" or "misses his mommy") in contrast to children in control classrooms. The effect for this finding was $d = .26$, or a small effect. Neither interaction effects nor other main effects emerged for the other *causes* dimensions for infant crying.

We next conducted a series of 2 (group) $\times$ 2 (gender) ANCOVAs, with age and ESL status as covariates to examine intervention effects on the three dimensions of children's strategy knowledge regarding infant crying. Analyses revealed a significant group $\times$ gender interaction for providing social/relational support dimension (e.g., "put the baby in his/her mother's arms or lap"). Girls in ROE classrooms, in contrast to girls in control classrooms, increased significantly in the number of times they reported that providing social/relational support is a viable strategy to utilize to help a crying infant (*Mean* difference scores .16 and $-.14$, respectively), $F(3, 276) = 5.44$, $p < .02$. The effect size for this interaction is $d = .28$, or a small effect. There was no significant difference between boys in

---

[2] The use of multilevel modeling (MLM)—analyzing data at the level of the classroom—was inappropriate due to the small classroom sample size in this study. As recommended by Raudenbusch (1997), a sample size of at least 40 clusters (e.g., classrooms, schools) needs to be given in order to achieve satisfactory statistical power in classroom-based program evaluation using MLM. Hence, analyses were conducted at the level of the individual (i.e., child), rather than at the classroom level. Although methodological research has indicated that significance levels resulting from individual level analyses where a program was implemented at the level of the classroom may be overstated (e.g., Donner & Klar, 2000), it has also supported the notion that effect sizes remain unbiased (Raudenbusch & Bryk, 2002).

Table 1 Casual explanations for infant crying: adjusted means at pretest and posttest for ROE program and control groups and adjusted difference scores, with age and ESL status as covariates

Casual explanation	Intervention ($n = 306$)			Control ($n = 279$)			Group difference in change	
	Pretest	Posttest	Difference score	Pretest	Posttest	Difference score	F value	Effect size Cohen's d
Internal causes								
Emotion state	.52 (.81)	.82 (1.13)	.30 (1.22)	.45 (.72)	.47 (.72)	.03 (.85)	9.82**	.26
Physical state	2.44 (1.53)	2.84 (1.53)	.40 (1.56)	2.12 (1.50)	2.43 (1.58)	.32 (1.58)	.36$_{ns}$	.05
External causes	.15 (.46)	.12 (.35)	−.03 (.59)	.22 (.51)	.20 (.50)	−.02 (.59)	.25$_{ns}$	.01

Numbers in parentheses are standard deviations

** $p < .01$

Table 2 Strategy knowledge for infant crying: adjusted means at pretest and posttest for ROE program and comparison groups and adjusted difference scores, with age and ESL status as covariates

Strategy knowledge	Intervention ($n = 306$)			Control ($n = 279$)			Group difference in change	
	Pretest	Posttest	Difference score	Pretest	Posttest	Difference score	F value	Effect size Cohen's d
Physical strategies	1.34 (1.10)	1.54 (1.09)	.20 (1.22)	1.33 (1.10)	1.49 (1.14)	.16 (1.23)	.00$_{ns}$	.03
External comfort object	.17 (.43)	.07 (.28)	−.10 (.49)	.11 (.32)	.08 (.27)	−.03 (.34)	4.37**	−.16
Social/relational support	.72 (.81)	.79 (.82)	.07 (1.02)	.64 (.89)	.59 (.75)	−.05 (.95)	2.15$_{ns}$	.11

Numbers in parentheses are standard deviations

** $p < .01$

Table 3 Empathy and perspective taking: adjusted means at pretest and posttest for ROE program and comparison groups and adjusted difference scores, with age and ESL status as covariates

Variable	Intervention ($n = 306$)			Control ($n = 279$)			Group difference in change	
	Pretest	Posttest	Difference score	Pretest	Posttest	Difference score	F value	Effect size Cohen's d
Empathy	3.69 (.79)	3.62 (.81)	−.08 (.75)	3.54 (.72)	3.52 (.82)	−.02 (.65)	1.07$_{ns}$	.08
Perspective taking	3.25 (.72)	3.25 (.78)	.00 (.70)	3.08 (.75)	3.14 (.76)	.06 (.66)	1.18$_{ns}$	.08

Numbers in parentheses are standard deviations

ROE classrooms and boys in control classrooms on this dimension (*Mean* differences scores −.02 and .03, respectively). No other interactions emerged for the other two strategy knowledge dimensions. However, as can be seen in Table 2, the ANCOVA analyses revealed a significant main effect for group assignment for one of the three strategies for helping a crying infant: in contrast to children in control classrooms, ROE program children significantly decreased from pretest to posttest in the frequency with which they mentioned providing external comfort object/environment strategies. This indicates that ROE children were less likely than control children to suggest that providing an external comfort object/environment (e.g., "give the baby a pacifier," "put the baby in a swing") as a strategy to help a baby who is crying. The effect for this finding was small.

Empathy and Perspective Taking

To test for intervention effects on empathy and perspective taking, we conducted a series of 2 (group) × 2 (gender) ANCOVAs for empathy and for perspective taking, respectively, with age and ESL status as covariates. As can be seen in Table 3, there were no significant intervention effects for either empathy or perspective taking. For both of these outcome variables, the analyses of potential interactions were non-significant.

Peer-Nominated Prosocial and Aggressive/Antisocial Behaviors

We next examined changes from pretest to posttest on dimensions of peer-nominated prosocial and aggressive/

antisocial behaviors. A MANCOVA analysis revealed a significant multivariate main effect for intervention status, $F(8, 570) = 12.29$, $p < .001$. As can be seen in Table 4, follow-up univariate analyses revealed that ROE children, relative to control children, increased significantly from pretest to posttest on all of the six dimensions of peer-nominated prosocial behaviors. In other words, children in ROE classrooms were more likely than children in control classrooms to increase in the frequency with which they were nominated by their peers as being prosocial with respect to "shares," "cooperates," "helps other kids when they have a problem," "kind," "understands other kids' point of view," and "fair." The average effect size across all six dimensions was .48, or a moderate effect. Effect sizes ranged from small (i.e., .28 for the dimension "cooperates) to large (i.e., .79 for the dimension of "fair"). No significant gender-by-intervention interactions emerged for prosocial behavior, and there were no significant main effects or interaction effects for peer-nominated aggressive/antisocial behaviors.

Teacher-Rated Prosocial and Aggressive Behaviors

For teacher-reported outcomes, a 2 (group) by 2 (gender) MANCOVA was conducted to determine intervention effects on teachers' ratings of children's prosocial and aggressive behaviors (i.e., physical, reactive, proactive, and relational aggression). A significant main effect for intervention status was found, $F(5, 566) = 12.53$, $p < .001$. As shown in Table 5, follow-up univariate analyses revealed that ROE program children, relative to control children, decreased significantly in teacher-rated proactive and relational aggression. The effect size for proactive

aggression was −.53 (a medium effect), and the effect size for relational aggression was −.36 (a small–medium effect). None of the remaining effects were significant.

Because we noted that the difference scores for proactive and relational aggression were in opposite directions for the ROE and control groups, reflecting a *decrease* in these aggressive behaviors among ROE children and an *increase* in these aggressive behaviors among control children, we conducted paired *t* tests to determine whether the within-group change was significant from pretest to posttest for proactive and relational aggression for the ROE program children and control children, separately. Analyses revealed that, whereas children in the ROE program group evidenced significant *decreases* in their proactive and relational aggression from pretest to posttest, $(t(296) = 5.24$ and $t(296) = 2.88$, respectively, $ps < .01$, $d = .30)$, control children evidenced significant *increases* in their proactive and relational aggression, $(t(274) = −3.83$ and $t(274) = −2.47$, respectively, $ps < .01$, $d = .19)$.

Discussion

This study represents one of the first systematic attempts to examine the effectiveness of the ROE program—an SEL program in which a caregiver and an infant serve as a springboard for lessons about emotions, caring for others, and infant development. Using a rigorous quasi-experimental design, we found that children in ROE program classrooms, in contrast to children in control classrooms, improved in their understanding of infant crying; significantly increased in prosocial behaviors as obtained via peer reports; and significantly decreased in proactive and

Table 4 Peer behavioral assessment: adjusted means at pretest and posttest for ROE program and control groups and adjusted difference scores, with age, gender, and ESL status as covariates

	Intervention ($n = 306$)			Control ($n = 279$)			Group difference in change	
	Pretest	Posttest	Difference score	Pretest	Posttest	Difference score	F value	Effect size Cohen's d
Prosocial behavior								
Shares	.43 (.19)	.54 (.18)	.11 (.15)	.42 (.19)	.47 (.20)	.05 (.14)	20.87***	.41
Cooperates	.52 (.22)	.59 (.24)	.07 (.15)	.49 (.23)	.52 (.25)	.03 (.13)	7.22**	.28
Helps	.38 (.19)	.44 (.21)	.06 (.15)	.37 (.21)	.38 (.20)	.01 (.14)	15.35***	.34
Kind	.53 (.20)	.59 (.21)	.05 (.14)	.51 (.22)	.48 (.23)	−.03 (.13)	40.23***	.59
Takes others' view	.35 (.18)	.40 (.18)	.04 (.13)	.34 (.19)	.32 (.19)	−.02 (.13)	29.62***	.46
Fair	.48 (.19)	.53 (.19)	.05 (.14)	.49 (.22)	.43 (.20)	−.06 (.14)	68.75***	.79
Antisocial behavior								
Starts fights	.13 (.19)	.16 (.22)	.03 (.13)	.15 (.19)	.19 (.21)	.03 (.11)	.32$_{ns}$	.00
Breaks rules	.14 (.20)	.19 (.24)	.05 (.13)	.19 (.20)	.23 (.26)	.04 (.11)	5.03$_{ns}$	.08

Numbers in parentheses are standard deviations

** $p < .01$; *** $p < .001$

Table 5 Teacher reports of child behaviors: adjusted means at pretest and posttest for ROE program and control groups and adjusted difference scores, with gender, age, and ESL status as covariates

	Intervention ($n = 297$)			Control ($n = 275$)			Group difference in change	
	Pretest	Posttest	Difference score	Pretest	Posttest	Difference score	F value	Effect size Cohen's d
Aggressive with peers	1.22 (.34)	1.25 (.37)	.03 (.32)	1.21 (.36)	1.24 (.38)	.03 (.30)	$.06_{ns}$	.00
Prosocial with peers	2.45 (.49)	2.46 (.46)	.01 (.49)	2.35 (.58)	2.40 (.60)	.05 (.43)	1.85_{ns}	.00
Reactive aggression	1.20 (.37)	1.22 (.37)	.02 (.32)	1.20 (.34)	1.23 (.38)	.03 (.25)	$.09_{ns}$	.00
Proactive aggression	1.25 (.29)	1.17 (.25)	−.08 (.26)	1.15 (.24)	1.20 (.28)	.05 (.23)	39.93**	−.53
Relational aggression	1.17 (.35)	1.11 (.31)	−.06 (.34)	1.10 (.27)	1.16 (.35)	.06 (.32)	14.77**	−.36

Numbers in parentheses are standard deviations

** $p < .001$

relational aggression as obtained through teacher reports. Additionally, a review of ROE instructors' diaries indicated that the program was implemented as intended (i.e., adherence), with instructors reporting that students were exposed to all or most of the components of the ROE curriculum.

Our findings both replicate and extend what is known about the effectiveness of school-based prevention programs in several specific ways. First, one of the strengths of the present study is that it provides support for the effectiveness of an SEL program implemented under "real-world" conditions in classrooms that reflect the diversity of many large urban school districts. Our study, although not a randomized controlled trial, is high in ecological validity because it is relevant for contexts where teachers are expected to address the needs of an increasingly diverse group of typical children in regular classrooms. Research has indicated that within one school district, school-age populations vary considerably in terms of their family experiences, cultural experiences, and socioeconomic status. Determining the effectiveness of a social-emotional competence promotion program within such a diverse routine practice condition provides high external validity to the ROE program in these conditions. Wilson et al. (2003), in their meta-analysis of the effects of school-based interventions on aggressive behavior, cogently argue for a differentiation between evaluations of *demonstration programs* and *routine practice programs*. According to them, demonstration programs are those set up by researchers whose primary mandate is to determine program efficacy under controlled conditions. Such programs are generally implemented by the researchers and are often delivered by research staff and/or teachers who are closely supervised by the researchers. In contrast, routine practice programs, such as the one presented herein, are those programs that are already being implemented in schools on a routine, ongoing basis. Although evaluations of demonstration programs have high internal validity, such efficacy

trials are low in external validity and tell us little about the program's effectiveness under typical conditions of community ownership. In their review, Wilson et al. emphasized the relative dearth of evaluations of routine practice programs and concluded that there was insufficient evidence about the effectiveness of routine practice programs to offer any firm conclusions about their efficacy in reducing aggression. Hence, the findings of the present study make an important addition to a small, but growing corpus of evidence for routine practice programs.

Second, because in evaluating our results, we not only utilized the traditional benchmark tests of statistical significance to examine the effects of the ROE program on a range of child outcomes but also calculated effect sizes (Wilkinson and the Task Force on Statistical Inference, 1999), the findings add to what is known about the practical effects of a classroom-based SEL program. Following Cohen's (1988) criteria, our intervention effects on behaviors revealed an average effect size of .48 (range: .28 to .79) for peer-nominated prosocial behaviors and an average effect size of .45 for teacher-rated proactive and relational aggression. Applying Cohen's (1988) criteria for small, medium, and large effects, our average effect sizes would be considered to be in the medium or moderate range. These effect sizes are in accord those found in other meta-analyses of school-based violence prevention programs (e.g., Wilson, Gottfredson, & Najaka, 2001; Wilson et al., 2003) and are slightly higher than those found in a recent meta-analysis conducted by Durlak et al. (2011) of 213 school-based universal social and emotional learning programs involving 270,034 students. Our moderate effect sizes may be due, in part, to the high fidelity of program implementation that we found (as discussed below). Indeed, as noted by Greenberg (2010) "quality of implementation has a significant impact on outcome, with higher quality implementation related to a greater effect size" (p. 31).

Third, evidence from this investigation increases both researchers' and educators' understanding of program

implementation by showing that a relatively easy-to-use and cost-effective SEL program that incorporates an infant and his/her caregiver can be implemented as intended, or in other words, with high implementation integrity. Specifically, our analyses of ROE instructors' diaries indicated that 96% of ROE lessons were implemented as intended, with an average of 95% of lesson plan content being delivered. This rate of implementation fidelity is higher than that typically reported in other evaluations of classroom-based SEL program. As noted by Durlak and DuPre (2008) in their meta-analysis examining the influence of implementation on program outcomes, "Expecting perfect or near-perfect implementation is unrealistic. Positive results have often been obtained with levels around 60%; few studies have attained levels greater than 80%. No study has documented 100% implementation for all providers." (p. 331).

What can account for the high level of implementation integrity as reported by the ROE instructors? Perhaps the successful implementation of the ROE program was due, in part, to the implementation supports that were provided to the ROE instructors. Dane and Schneider (1998), in their review of studies of school-based interventions, posit that one important dimension of implementation integrity is program "promotion." For them, program promotion includes uses of a program manual, formal training opportunities, and ongoing consultation and support for program providers. According to these criteria, the ROE program would be considered to be high on this dimension of implementation. Specifically, the ROE curriculum is manualized. Moreover, each of the ROE instructors received an intensive 3-day training along with a 1-day booster session mid-way through the school year. Moreover, each ROE instructor was provided with an expert ROE mentor or "coach" to provide support and feedback to the instructor, as well as helping trouble-shoot special situations and facilitating instructors' learning. Joyce and Showers (2002) have emphasized the importance of coupling training with coaching. They found that when training was combined with coaching, 95% of teachers acquired knowledge and developed skills for applying that knowledge in the classroom. In cases in which coaching was absent, a small percentage (i.e., 5%) of teachers applied the skills in the classroom.

Or perhaps high program fidelity was maintained because of the inclusion of an infant and his/her caregiver into the curriculum. That is, because a commitment was made by the ROE parent to participate in the program and attend the monthly ROE lessons, all of the ROE lessons were completed and thus the program was implemented as intended. High fidelity may also have been maintained because the ROE instructor regularly contacted the ROE parent and scheduled the parent/infant classroom visits to accommodate the parents' schedule, hence reducing the likelihood for the parent and his/her infant to miss a lesson. This finding is provocative because it raises questions about the availability of a parent and infant to participate in the ROE program. As described earlier, a parent and his/her infant play a central role in the ROE program lesson content and delivery. Indeed, the entire delivery of the ROE program is dependent on this dimension of the program. The finding in the present study was that program implementation was high for the parent/infant visit lesson content, with an average of 95% of content being covered for each of the parent/infant visits. Perhaps this high degree of implementation fidelity for the parent/infant visits is attributable, in part, to the larger cultural and political context in which the program was implemented. Specifically, perhaps high implementation of the parent/infant portion of the ROE curriculum was due, in part, to Canada's parental leave policy. In 2000, Canada greatly expanded its parental leave from 10 to 35 weeks that can be divided as desired between two parents. This time is in addition to 15-week paid maternity leave. Moreover, in most situations, maternity leave can be combined with parental leave to yield a maximum of 50 weeks (note also that parental leave and maternity leave benefits are paid by the Canadian Employment Insurance system; http://www.servicecanada.gc.ca/eng/ei/types/special.shtml). Hence, it may be that because many parents are at home with their infant for almost an entire year, many of them are willing and available to participate in the ROE program. Future research examining the implementation of the ROE program in countries other than Canada should examine further this relation between implementation fidelity and parental leave policies.

Fourth, because proactive, reactive, and relational aggression were examined prospectively across a school year in both ROE and control classrooms, the findings from the current study add to what is known about stability and change in these forms of aggression during the elementary school years and how a classroom-based SEL program can interrupt the "normal" trajectory. Previous research has demonstrated a linear increase in aggression over the course of the elementary school years (Aber, Brown, & Jones, 2003; Fonagy et al., 2009). What our data have to say about aggression in schools and the role of preventive interventions in interrupting a trajectory characterized by increased aggression is somewhat unique to what has been learned previously in several respects.

One of these advances was attained by examining *change* over time rather than simply conducting analyses comparing posttest differences between intervention and control children. Examining change from pretest to posttest, our findings revealed *increases* in proactive aggression among control-group children and *decreases* in these forms

of aggression among ROE program children. The effect size for this difference was in the moderate range, $d = .53$. This finding is noteworthy given the insidious and disturbing nature of these forms of aggression. Indeed, relative to children in the ROE program group, it may not be too surprising that proactive aggression increased during the school year among children in the control since it is an instrumental form of aggression that is more likely to be acquired and reinforced than other forms of aggression (Bandura, 1983).

Our results indicated no intervention effects for reactive aggression. As noted by Poulin and Boivin (2000), "Distinguishing reactive aggressive behaviors from proactive aggressive behaviors at the morphological level is an important first step in understanding the nature and functional value of both behaviors" (p. 115). Indeed, while a number of previous empirical investigations exist lending support to the distinction of these two dimensions of aggression in both laboratory and school settings (Boivin, Dodge, & Coie, 1995; Dodge & Coie, 1987), the present investigation is one of the first to provide evidence of the importance of distinguishing between reactive and proactive aggression in an evaluation of primary preventive efforts.

A similar, but less robust pattern ($d = .36$) was found for relational aggression. Specifically, we found increases in relational aggression among control children and decreases in this form of aggression among ROE program children. This increase in relational aggression among children in the control group mirrors those findings of recent investigations examining the developmental trajectories of relational aggression during the elementary school years. Murray-Close et al. (2007), for example, examined trajectories of relational aggression and found a rise in relational aggression over the course of one calendar among a large, diverse sample of fourth-grade girls. Similarly, Tremblay (2000) and others (e.g., Cote, Vaillancourt, Farhat, LeBlanc, Nagin, & Tremblay, 2006) have noted that although physical aggression decreases from the time children begin school until the end of high school, the middle childhood years are characterized as a time in the life span in which significant linear increases in relational and other forms of indirect aggression occur.

Our findings indicating a rise in relational aggression in control classrooms may be due, in part, to the nature of classroom dynamics and group formation in the elementary school years. As noted by Farmer and Xie (2007), "When same-age children or youth are aggregated together they tend to selectively and systematically synchronize their behavior with each other…. This leads to the formation of (1) differential preferences for peer interactions, (2) distinct social groups, and (3) hierarchical social structures" (p. 462). Although social hierarchies can function as a way

to reduce aggression, they can also exacerbate aggression among peer group members when children feel that their position in the social system is vulnerable and their roles and responsibilities in the group are threatened (Farmer & Xie, 2007). More specifically, when social hierarchies are seen as uncertain, as may be the case among children whose self-esteem and identity are in flux due to the biological, cognitive, social, and emotional changes that are occurring as they make transition from childhood to adolescence, the expression of aggression may become more endemic to the peer culture and occur more frequently (Adler & Adler, 1998; Pellegrini & Long, 2002). It is likely that as children vie for top positions in the peer hierarchy, even "ordinary" children may rely on relational aggression to either maintain or increase their status in the peer group. Hence, the mechanism that may underlie the increased aggression in the control group classrooms may be the putative behavioral and psychological stressors that ensue as children seek to maintain or improve their status in the peer group.

Our findings suggest that the ROE program can serve as a catalyst to redirect the downward spiral toward increased aggression that evolves during a typical elementary school year. The findings from our research examining the effectiveness of the ROE program are not unique however. Grossman et al. (1997), in their evaluation of the school-based social and emotional promotion program *Second Step®*, found a pattern similar to the one in the current study. That is, in their study of 790 2nd- and 3rd-graders, they found that those children who received the *Second Step®* program decreased in physical aggression over the course of the school year, whereas the children who did not receive the curriculum showed increases in both verbal and physically violent behaviors during the same time period. Our research findings extend those of Grossman et al. by showing that a classroom-based social and emotional competence promotion program can also influence other forms of aggression.

Fifth, the results of this investigation extend knowledge about trajectories of prosocial behavior in school-aged children and demonstrate that a classroom-based SEL program can lead to improvements in children's sharing, helping, and cooperative behaviors. Similar to what we found for teacher-rated relational and proactive aggression, our data indicated that peer-nominated prosocial behaviors worsened (i.e., decreased) among children in control classrooms over the course of the school year and improved among children who received the ROE intervention.

These findings challenge a commonly held belief in developmental psychology—that prosocial behaviors, including sharing, helping, comforting, and cooperation *increase* in frequency as children grow older. Nantel-Vivier et al. (2009), for instance, in their longitudinal,

multi-informant study examining trajectories of children's prosocial behaviors in Canada and Italy, found that, contrary to commonly held beliefs, prosocial behavior remained stable or declined among children between the ages of 10 and 15.

The increases in prosocial behavior among children in the ROE program relative to controls may reflect the ROE program emphasis. That is, because ROE lessons focus on caring for the baby and there are multiple opportunities provided throughout the curriculum for children to engage in activities that benefit the baby (e.g., creating a CD of lullabies for the baby, creating a book for the baby that documents his/her month by month growth and developmental milestones), the ROE program creates a classroom environment in which prosocial behaviors become the norm. As noted by Staub (1988): "Acting to benefit others can result in personal changes that lead to more significant helping. It can promote a more positive evaluation of the welfare of people who have benefited from one's actions and of people in general and a perception of oneself as a caring person willing to extend effort and make sacrifices for others' sake" (p. 95). Moreover, perhaps teaching emotional understanding skills in conjunction with group activities that benefit others, such as helping the baby in the ROE program, can lead to increases in children's prosocial behaviors with one another.

As with other studies with multiple reporters assessing comparable constructs, we found differential outcomes by reporter for similar behaviors (Achenbach, 2006). As Achenbach has noted, when diverse perspectives are needed to illustrate a construct, as in the case of children's aggressive and prosocial behaviors, it is important to take into consideration which set of windows are most likely to maximize one's view of the behavior. When teachers, not peers, reported on children's aggressive behaviors, a positive intervention impact was found for proactive and relational aggression. In contrast, when peer nominations were used to assess children's social behaviors, we found positive intervention effects for prosocial behaviors, including sharing, helpfulness, cooperation, kindness, perspective taking, and being seen as fair. The discrepancy by informer may reflect the different contexts in which the behavior is being observed as well as the differential salience of the behavior for the informant. In the case of teachers, their observations of children's social behaviors may be restricted primarily to the classroom, and teachers, in contrast to peers, may perceive aggressive behaviors as more salient than other types of social behaviors. In contrast, children may display different prosocial behaviors in their peer groups, and these behaviors may take place across a range of contexts, including the playground and the lunchroom. Taken together, our findings provide strong evidence for the utility of collecting multi-informant data

in evaluations of SEL programs and suggest that a range of informants, including peers and teachers, are in a position to supply common and unique information about the behavioral outcomes of children and, together, produce a better understanding of program impacts.

Finally, our data provide important information about the development of children's perceptions and comforting strategies to infant crying. We found that children who received the ROE curriculum, in contrast to control children, significantly increased in their understanding of infant crying in two important ways. First, ROE children showed significant improvements in their ability to spontaneously generate causes for infants' crying that reflected attention to the *emotions* of the baby (i.e., making reference to the infant's feelings such as "the baby is crying because he is upset" or "misses his mommy"). Second, with regard to strategies for helping a crying baby, we found a significant group-by-gender interaction for the dimension that focused on providing social/relational support dimension (e.g., "put the baby in his/her mother's arms or lap"). Girls in ROE classrooms, in contrast to boys, increased significantly in the number of times they reported that providing social/relational support was a viable strategy for helping a crying infant. Third, ROE program children significantly decreased from pretest to posttest in the frequency with which they mentioned providing external comfort object/environment strategies. Collectively, these findings extend previous work on children's understanding of infant distress and prosocial responding.

For decades now, researchers have studied infant crying and its many functions. Responding sensitively to infant crying is central to optimal infant development and secure attachment (e.g., Ainsworth, Blehar, Waters, & Wall, 1978; Fish, Stifter, & Belsky, 1991). Until now, however, very little research has examined the manner in which understanding of and responsiveness to infant crying emerge during the childhood years and how a classroom-based SEL intervention can advance children's understanding of infant distress and their responsiveness to it. Murray (1979) proposed an ethological framework where cries were viewed as elicitors of emotions that promote prosocial behaviors such as helping, comforting, and protecting the infant (Hoffman, 1978). Researchers examining children's understanding and responses to infant crying have shown that caregiving responses to infant crying increase with age (e.g., Zahn-Waxler, Friedman, & Cummings, 1983) and that such increases are due, in part, to increased sophistication in understanding others' needs. That is, one mechanism underlying developmental changes in the nature and type of responses generated may reflect advanced social-cognitive abilities. A possible explanation for the greater frequency with which ROE program children identified emotions as a cause of infant crying may be a result of the

program focus on the emotional life of infants and that these discussions occurred in situations in which a real infant was present (and sometimes crying). Such experiences may have been central for advancing children's social-cognitive understanding of infant crying. It may be that such changes would not be evident had it not been for the presence of the ROE infant. Further research is clearly needed to shed further light on this supposition.

The findings on empathy are puzzling. Contrary to expectations, there were no significant differences in self-reported empathy and perspective taking between ROE and control children. Because the bases for our null findings are indeterminant, we have no way of knowing whether the failure to find significant improvements in empathy-related responding among children who received the ROE program was due to the investigation's methodology, an invalid theoretical premise, or to other unknown factors that we did not consider. However, the following observations are offered to assist investigators who may wish to examine the development of empathy in future investigations into classroom-based SEL program. First, it may be that because the measure we utilized in our study to assess empathy and perspective taking (i.e., Davis' Interpersonal Reactivity Index) is a measure that has been used primarily to measure empathy in adolescents and adults (for a review, see Zhou, Valiente, & Eisenberg, 2003), it may not have been sensitive enough in detecting developmental changes in children's empathy and perspective taking across middle childhood. Beyond this consideration, it is also plausible that because there was such an explicit teaching of "empathy" in the ROE curriculum, the children in the ROE program became sensitized to the meaning of empathy and hence, at posttest, approached the items on the measure differently. Future research is clearly needed to compare the measurement of empathy using different approaches.

Limitations and Directions for Future Research

The data reported here are encouraging of continued investigation into the effectiveness of the ROE program. Nonetheless, while evaluating the results, several limitations should be noted. Firstly, one limitation derives from the utilization of a quasi-experimental design rather than the "gold-standard" (see Biglan, Mrazek, Carnine, & Flay, 2003 for a discussion about the utility of other designs in adding to the evidence base) and not the recommended "gold-standard" randomized clinical trial (RCT). That is, classrooms were not randomly assigned to receive the ROE program. Instead, teachers volunteered to have the program implemented in their classroom. This makes causal inferences between the ROE program and children's social and emotional competence difficult to make in the absence of further evidence. The observed effects may be due to specific

dimensions of the ROE curriculum, the characteristics of the ROE instructor/classroom teacher, or some combination of the two. Nonetheless, there are several reasons for confidence in our findings. First, children were not assigned to teachers on the basis of the ROE program, and thus, there was no reason to expect that children who received the program were different from children who did not receive it. We tested baseline differences between children in the ROE program and control groups on demographic and outcome measures at baseline and found no significant differences between groups, suggesting that our matching process was successful. Second, the pattern of change across several of our dimensions of social behaviors are patterns that Shadish, Cook, and Campbell (2002) have identified as tenable for asserting causal claims in quasi-experimental designs. For instance, our findings with regard to proactive aggression indicated a different pattern of change from pretest to posttest for ROE program and comparison groups, respectively. More specifically, whereas children in the ROE program *decreased* in proactive aggression, control children *increased* in proactive aggression. This pattern is consistent with a pattern in which other plausible explanations of threats to internal validity are low. Finally, it is significant that the teachers who volunteered for ROE in their classrooms did not rate social and emotional factors as more important than control teachers, suggesting that the teachers differed only according to the opportunity to have the program and not according to key attitudinal factors. Future research on ROE could be extended by using an RCT design to determine the extent to which our positive findings can be replicated.

Secondly, because we relied solely on ROE instructors' self-reports of implementation fidelity/adherence, we cannot be certain that the ROE program was truly implemented as intended. That is, because we have no objective data on implementation integrity, we have no way of determining whether the ROE was implemented at the high level conveyed via the ROE instructors' monthly diaries. As noted by Schulte, Easton, and Parker (2009), "Research has consistently shown that implementers overestimate their own level of treatment integrity (Dusenbury, Brannigan, Hansen, Walsh, & Falco, 2005; Lillehoj, Griffin, & Spoth, 2004; Resnicow et al., 1998)" p. 469. Hence, some caution should be taken when interpreting our implementation findings. In future research on the ROE program, it is important to include multiple measures for monitoring implementation, including observer-report and direct observations, in addition to self-reports (see Leff, Hoffman, & Lakin Gullan, 2009). For instance, one way to objectively monitor implementation may be through videotaping lessons that can then be coded later by independent observers (Schulte et al., 2009).

Thirdly, neither teachers nor peers were blind to experimental condition when providing their ratings of

children's behaviors. We readily acknowledge that teachers can be a critical source of information due to the fact that they have many opportunities to observe children for several hours each day. Nonetheless, although our teacher-report measure of child behaviors included filler items to lessen response bias in our teachers' reports, teachers were aware of intervention status. With regard to our peer behavioral assessments, peers were also not blind to treatment condition, and they would have easily been able to discern which condition their classroom had been assigned. Although peers as participant observers can provide important sources of information about their classmates' behaviors both inside and outside of the classroom, our peer behavioral assessment measure of prosocial and aggressive/antisocial behaviors may have been influenced by peers' knowledge about experimental condition. We speculate here that peers' ratings of classmates' behaviors would be less likely than teachers to be influenced by knowledge of the intervention status given that it is unlikely that children would be able to generate specific hypothesis of the study. However, we have no data to support such a claim, and future investigations into the ROE program would benefit from collecting data from observers blind to intervention status in order to allow for a more objective measure of children's behaviors.

Thirdly, our analyses were conducted at the individual child's level even though selection was by classroom. Unfortunately, the small number of classrooms did not provide sufficient statistical power to use multilevel modeling. The clustering of children within classrooms results in the non-independence of subjects, which could bias the statistical tests used to identify intervention effects. This is a major challenge to evaluations of universal school-based interventions when insufficient resources exist to recruit large numbers of classrooms or schools (Stoolmiller, Eddy, & Reid, 2000). Nonetheless, as noted by Slavin (2008), although analyzing data at the child level when randomization was done at the classroom level is discouraged by methodologists (Donnor & Klar, 2000) because the findings can exaggerate statistical significance, "… their effect sizes are unbiased (Raudenbush & Bryk, 2002) and therefore are of value…" (p. 9).

A final limitation is that our analysis did not address questions regarding differential program effects for individual children. Although we did control statistically for important child demographic characteristics, namely age and ESL status, such analyses does not allow an exploration of differential program effects for children of different ages and/or differing ethnic and cultural backgrounds. Given the wide-ranging ethnic and cultural diversity of the children in our sample, including a larger number of participants might have allowed for a more specific examination of the manner in which children's background characteristics interact with classroom contexts to determine outcomes. Finally, because the ROE curriculum is multifaceted with many lessons and foci, it is difficult to determine which aspects of the program are responsible for the effects in social-emotional understanding and behavioral change.

Despite these limitations, the results of this study provide empirical evidence that a relatively easy-to-implement universal classroom-based intervention that does not focus exclusively on disturbed or atypical children substantially improves children's cooperative and kind behavior to one another and decreases their aggression. Such an evaluation not only provides support for the continued implementation and evaluation of preventive efforts that are grounded in the scientific study of children's development but also contributes to the ongoing scholarship at the intersection of translational research, developmental and prevention sciences, and educational reform.

References

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*, 324–348.

Achenbach, T. M. (2006). As others see us: Clinical and research implications of cross-informant correlations for psychopathology. *Psychological Science, 15*, 94–98.

Adler, P. A., & Adler, P. (1998). *Peer power: Preadolescent culture and identity*. New Brunswick, NJ: Rutgers University Press.

Ainsworth, M. D. S., Blehar, M. C., Waters, E., & Wall, S. (1978). *Patterns of attachment: A psychological study of the Strange Situation*. Hillsdale, NJ: Erlbaum.

Ang, R., & Hughes, J. (2001). Differential benefits of skills training with antisocial youth based on group composition: A meta-analytic investigation. *School Psychology Review, 31*, 164–185.

Arsenio, W. F., & Lemerise, E. A. (2001). Varieties of childhood bullying: Values, emotion processes, and social competence. *Social Development, 10*, 59–73.

Bandura, A. (1983). Psychological mechanisms of aggression. In R. G. Green & E. I. Donnerstein (Eds.), *Aggression: Theoretical and empirical views* (Vol. 1, pp. 1–40). New York: Academic.

Barrett, K., & Campos, J. (1987). Perspectives on emotional development II: A functionalist approach to emotions. In J. Osofsky (Ed.), *Handbook of infant development* (2nd ed., pp. 555–578). New York: Wiley.

Battistich, V., Solomon, D., Watson, M., & Schaps, E. (1997). Caring school communities. *Educational Psychologist, 32*, 137–151.

Battistich, V., Solomon, D., Watson, M., Solomon, J., & Schaps, E. (1989). Effects of an elementary school program to enhance prosocial behavior on children's cognitive social

problem-solving skills and strategies. *Journal of Applied Developmental Psychology, 10*, 147–169.

Biglan, A., Mrazek, P. J., Carnine, D., & Flay, B. R. (2003). The integration of research and practice in the prevention of youth problem behaviors. *American Psychologist, 58*, 433–440.

Boivin, M., Dodge, K. A., & Coie, J. D. (1995). Individual-group behavioral similarity and peer status in experimental play groups of boys: The social misfit revisited. *Journal of Personality and Social Psychology, 69*, 269–279.

Bukowski, W. M., & Sippola, L. K. (1996). Friendship and morality. In W. M. Bukowski & A. F. Newcomb (Eds.), *The company they keep: Friendship in childhood and adolescence* (pp. 238–261). Cambridge: Cambridge University Press.

Cairns, R. B., Cairns, B. D., Neckerman, H. J., Gest, S. D., & Gariépy, J. L. (1988). Social networks and aggressive behavior: Peer support or peer rejection? *Developmental Psychology, 24*, 815–823.

Campos, J., Mumme, D., Kermoian, R., & Campos, R. (1994). A functionalist perspective on the nature of emotion. *Monographs of the Society for Research in Child Development, 59*(2–3), 284–303.

Cappella, E., & Weinstein, R. (2006). The prevention of social aggression among girls. *Social Development, 15*, 434–462.

Caprara, G. V., Barbanelli, C., Pastorelli, C., Bandura, A., & Zimbardo, P. G. (2000). Prosocial foundations of children's academic achievement. *Psychological Science, 11*, 302–306.

Card, N. A., Stucky, B. D., Sawalani, G. M., & Little, T. D. (2008). Direct and indirect aggression during childhood and adolescence: A meta-analytic review of gender differences, intercorrelations, and relations to maladjustment. *Child Development, 79*, 1185–1229.

Catherine, N., & Schonert-Reichl, K. A. (2011). Children's perceptions and comforting strategies to infant crying: Relations to age, sex, and empathy-related responding. *British Journal of Developmental Psychology, 29*, 524–551. doi:10.1348/026151010X521475.

Cohen, J. (1988). *Statistical power analysis* (2nd ed.). Hillsdale, NJ: Erlbaum.

Cohen, J. (Ed.). (2001). *Caring classrooms/intelligent schools: The social emotional education of young children*. New York: Teachers College Press.

Coie, J. D., & Dodge, K. A. (1988). Multiple sources of data on social behavior and social status in the school: A cross-age comparison. *Child Development, 59*, 815–829.

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

Consortium on the School-Based Promotion of Social Competence. (1996). The school-based promotion of social competence: Theory, research, practice, and policy. In R. J. Haggerty, L. R. Sherrod, N. Garmezy, & M. Rutter (Eds.), *Stress, risk, and resilience in children and adolescents: Processes, mechanisms, and interventions* (pp. 268–316). Cambridge, MA: Cambridge University Press.

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues in field settings*. Chicago: Rand McNally.

Cook, T. D., Murphy, R. F., & Hunt, H. D. (2000). Comer's School Development Program in Chicago: A theory-based evaluation. *American Educational Research Journal, 37*, 535–597.

Cote, S., Vaillancourt, T., Farhat, A., LeBlanc, J. C., Nagin, D., & Tremblay, R. (2006). The development of physical aggression from toddlerhood to pre-adolescence: A nation wide longitudinal study of Canadian children. *Journal of Abnormal Child Psychology, 34*, 71–85.

Crick, N. R. (1996). The role of overt aggression, relational aggression, and prosocial behavior in the prediction of children's future social adjustment. *Child Development, 67*, 2317–2327.

Crick, N. R., & Dodge, K. A. (1999). 'Superiority' is in the eye of the beholder: A comment on Sutton, Smith, and Swettenham. *Social Development, 8*, 128–131.

Crick, N. R., & Grotpeter, J. K. (1995). Social aggression, gender, and social psychological adjustment. *Child Development, 66*, 710–722.

Damon, W. (2004). What is positive youth development? *Annals of the American Academy of Political and Social Science, 591*, 13–24.

Damon, W., & Gregory, A. (2003). Bringing in a new era in the field of youth development. In R. M. Lerner & P. L. Benson (Eds.), *Developmental assets and asset-building communities: Implications for research, policy, and practice* (pp. 47–64). New York: Kluwer/Plenum.

Dane, A. V., & Schneider, B. H. (1998). Program integrity in primary and early secondary prevention: Are implementation effects out of control? *Clinical Psychology Review, 18*, 23–45.

Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology, 44*, 113–126.

Denham, S. A. (1998). *Emotional development in young children*. New York: Guilford.

Denham, S. A., & Burton, R. (1996). A social-emotional intervention program for at-risk four-year-olds. *Journal of School Psychology, 34*, 224–245.

Dodge, K. A., & Coie, J. D. (1987). Social-information processing factors in reactive and proactive aggression in children's peer groups. *Journal of Personality and Social Psychology, 53*, 1146–1158.

Domitrovich, C. E., & Greenberg, M. T. (2000). The study of implementation: Current findings from effective programs that prevent mental disorders in school-age children. *Journal of Educational and Psychological Consultation, 11*, 193–221.

Donner, A., & Klar, N. (2000). *Design and analysis of cluster randomized trials in health research*. New York: Oxford University Press.

Durlak, J. A., & DuPre, E. P. (2008). Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology, 41*, 327–350.

Durlak, J. A., Weissberg, R. P., Dymnicki, A. B., Taylor, R. D., & Schellinger, K. B. (2011). Enhancing students' social and emotional development promotes success in school: Results of a meta-analysis. *Child Development, 82*, 474–501.

Dusenbury, L., Brannigan, R., Hansen, W. B., Walsh, J., & Falco, M. (2005). Quality of implementation: Developing measures crucial to understanding the diffusion of preventive interventions. *Health Education Research, 20*, 308–313.

Eisenberg, N., Fabes, R. A., & Spinrad, T. (2006). Prosocial development. In N. Eisenberg (Ed.), *Handbook of child psychology: Social, emotional, and personality development* (pp. 646–718). Hoboken, NJ: Wiley.

Eisenberg, N., & Miller, P. A. (1987). The relation of empathy to prosocial and related behaviors. *Psychological Bulletin, 101*, 91–119.

Emde, R. N., Osofsky, J. D., & Butterfield, P. M. (1993). *The IFEEL pictures: A new instrument for interpreting emotions*. Madison, WI: International Universities Press.

Fabes, R., Eisenberg, N., Nyman, M., & Michelieu, Q. (1991). Children's appraisals of others' spontaneous emotional reactions. *Developmental Psychology, 27*, 858–866.

Farmer, T. W., & Xie, H. (2007). Aggression and school social dynamics: The good, the bad, and the ordinary. *Journal of School Psychology, 45*, 461–478.

Feshbach, N. D. (1979). Empathy training: A field study in affective education. In S. Feshbach & A. Fraczek (Eds.), *Aggression and behaviors change: Biological and social processes* (pp. 234–249). New York: Praeger.

Fish, M., Stifter, C. A., & Belsky, J. (1991). Conditions of continuity and discontinuity in infant negative emotionality: Newborn to five months. *Child Development, 62*, 1525–1537.

Fonagy, P., Twemlow, S. W., Vernberg, E. M., Nelson, J. M., Dill, E. J., Little, T. D., et al. (2009). A cluster randomized controlled trial of child-focused psychiatric consultation and a school systems-focused intervention to reduce aggression. *Journal of Child Psychology and Psychiatry, 50*, 607–616.

Gordon, M. (2000). *Roots of Empathy: Training manual*. Available from the Roots of Empathy, www.Rootsofempathy.org

Gordon, M. (2001). *Roots of Empathy: Training manual*. Available from the Roots of Empathy, www.Rootsofempathy.org

Gordon, M. (2005). *Roots of Empathy: Changing the world child by child*. Toronto, ON: Thomas Allen.

Greenberg, M. T. (2004). Current and future challenges in school-based prevention: The researcher perspective. *Prevention Science, 5*, 5–13.

Greenberg, M. T. (2010). School-based prevention: Current status and future challenges. *Effective Education, 2*, 27–52.

Greenberg, M. T., Kusché, C. A., Cook, E. T., & Quamma, J. P. (1995). Promoting emotional competence in school-aged children: The effects of the PATHS curriculum. *Development and Psychopathology, 7*, 117–136.

Gresham, F. M., Cook, C. R., Crews, S. D., & Kern, L. (2004). Social skills training for children and youth with emotional and behavioral disorders: Validity considerations and future directions. *Behavioral Disorders, 30*, 32–46.

Grossman, D. C., Neckerman, H. J., Koepsell, T. D., Liu, P. Y., Asher, K. N., Beland, K., et al. (1997). The effectiveness of a violence prevention curriculum among children in elementary school. *Journal of the American Medical Association, 277*, 1605–1611.

Hay, D. F. (1994). Prosocial development. *Journal of Child Psychology and Psychiatry, 35*, 29–71.

Hoffman, M. L. (1978). Empathy, its development and prosocial implications. In C. B. Keasy (Ed.), *Nebraska symposium on motivation*, Vol. 25, pp. 169–218.

Huebner, E., & Gilman, R. (2003). Toward a focus on positive psychology in school psychology. *School Psychology Quarterly, 18*, 99–102.

Hymel, S., Schonert-Reichl, K. A., & Miller, L. D. (2007). Reading, 'Riting, 'Rithmetic and relationships: Considering the social side of education. *Exceptionality Education Canada, 16*, 149–191.

Izard, C. (2002). Translating emotion theory and research into preventive interventions. *Psychological Bulletin, 128*, 796–824.

Izard, C. E., Fine, S., Mostow, A., Trentacosta, C., & Campbell, J. (2002). Emotional processes in normal and abnormal development and preventive intervention. *Developmental and Psychopathology, 14*, 761–787.

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

Joyce, B., & Showers, B. (2002). *Student achievement through staff development* (3rd ed.). Alexandria, VA: Association for Supervision and Curriculum Development.

Kendall, G., Schonert-Reichl, K., Smith, V., Jacoby, P., Austin, R., Stanley, F., et al. (2006). *The evaluation of the "Roots of Empathy" in Western Australia 2005*. A report written for the Department of Education and Training, Western Australia, Perth: Telethon Health Research Institute.

Ladd, G. W., & Profilet, S. M. (1996). The Child Behavior Scale: A teacher-report measure of young children's aggressive, withdrawn, and prosocial behaviors. *Developmental Psychology, 32*, 1008–1024.

Ladd, G. W., & Troop-Gordon, W. (2003). The role of chronic peer difficulties in the development of children's psychological adjustment problems. *Child Development, 74*, 1344–1367.

Lagerspetz, K. M. J., Bjorkqvist, K., & Peltonen, T. (1988). Is indirect aggression typical of females? Gender differences in aggressiveness in 11 to 12 year old children. *Aggressive Behavior, 14*, 403–414.

Leff, S. S., Hoffman, J. A., & Lakin Gullan, R. (2009). Intervention integrity: New paradigms and applications. *School Mental Health, 1*, 103–106.

Lewis, M., & Michalson, L. (1983). *Children's emotions and moods*. New York: Plenum Press.

Lillehoj, C. J., Griffin, K. W., & Spoth, R. (2004). Program provider and observer ratings of school-based preventive intervention implementation: Agreement and relation to youth outcomes. *Health Education & Behavior, 31*, 242–257.

Lösel, F., & Beelmann, A. (2003). Effects of child skills training in preventing antisocial behavior: A systematic review of randomized evaluations. *Annals of the American Academy of Political and Social Sciences, 857*, 177–186.

Masten, A. S., & Motti-Stefanidi, F. (2009). Understanding and promoting resilience in children: Promotive and protective processes in schools. In B. Gutkin & C. R. Reynolds (Eds.), *The handbook of school psychology*. New York: Wiley.

Murray, A. D. (1979). Infant crying as an elicitor of parental behavior: An examination of two models. *Psychological Bulletin, 86*, 191–215.

Murray-Close, D., Ostrov, J. M., & Crick, N. R. (2007). A short-term longitudinal study of growth of relational aggression during middle childhood: Associations with gender, friendship intimacy, and internalizing problems. *Development and Psychopathology, 19*, 187–203. doi:10.1017/S0954579407070101.

Nantel-Vivier, A., Kokko, K., Caprara, G. V., Pastorelli, C., Gerbino, M. G., Paciello, M., et al. (2009). Prosocial development from childhood to adolescence: A multi-informant perspective with Canadian and Italian longitudinal studies. *Journal of Child Psychology and Psychiatry, 50*, 590–598. doi:10.1111/j.1469-7610.2008.02039.

Neal, J. W. (2010). Social aggression and social position in middle childhood and early adolescence: Burning bridges or building them? *Journal of Early Adolescence, 30*, 122–137.

Noddings, N. (1992). *The challenge to care in schools: An alternative approach to education*. New York: Teachers College Press.

Oakley, A., Strange, V., Bonnell, C., Allen, E., & Stephenson, J. (2006). Health services research: Process evaluations in randomised control trials of complex interventions. *British Medical Journal, 332*, 413–416.

Oberle, E., Schonert-Reichl, K. A., & Thomson, K. (2010). Understanding the link between social and emotional well-being and peer relations in early adolescence: Gender-specific predictors of peer acceptance. *Journal of Youth and Adolescence, 39*, 1330–1342.

Parkhurst, J., & Asher, S. R. (1992). Peer rejection in middle school: Subgroup differences in behavior, loneliness, and interpersonal concerns. *Developmental Psychology, 28*, 231–241.

Pelligrini, A. D., & Long, J. D. (2002). A longitudinal study of bullying, dominance, and victimization during the transition from primary school through secondary school. *British Journal of Developmental Psychology, 20*, 259–280.

Poulin, F., & Boivin, M. (2000). Reactive and proactive aggression: Evidence of a two-factor model. *Psychological Assessment, 12*, 115–122. doi:10.1037/1040-3590.12.2.115.

Raudenbush, S. W. (1997). Statistical analysis and optimal design for cluster randomized trials. *Psychological Methods, 2*, 173–185.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods*. Thousand Oaks, CA: Sage.

Resnicow, K., Davis, M., Smith, M., Lazarus-Yaroch, A., Baranow-ski, T., Baranowskii, J., et al. (1998). How best to measure implementation of school health curricula: A comparison of three measures. *Health Education Research, 13*, 239–250.

Rubin, K. H., Coplan, R. J., Fox, N. A., & Calkins, S. D. (1995). Emotionality, emotion regulation, and preschoolers' social adaptation. *Development and Psychopathology, 7*, 49–62.

Saarni, C. (1999). *The development of emotional competence*. New York: Guilford.

Schonert-Reichl, K. A. (1993). Empathy and social relationships in adolescents with behavioral disorders. *Behavioral Disorders, 18*, 189–204.

Schonert-Reichl, K. A. (1999). Moral reasoning during early adoles-cence: Links with peer acceptance, friendship, and social behaviors. *Journal of Early Adolescence, 19*, 249–279.

Schonert-Reichl, K. A., & Oberle, E. (2011). Teaching empathy to children: Theoretical and empirical considerations and implica-tions for practice. In B. Weber & E. Marsal (Eds.), *"The politics of empathy" New interdisciplinary perspectives on an ancient phenomenon*. Berlin, Germany: Lit Verlag.

Schulte, A. C., Easton, J. E., & Parker, J. (2009). Advances in treatment integrity research: Multidisciplinary perspectives on the conceptualization, measurement, and enhancement of treat-ment integrity. *School Psychology Review, 38*, 460–475.

Seligman, M. E. P., & Csikszentmihalyi, M. (2000). Positive psychology. *American Psychologist, 55*, 5–14.

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

Shipman, K., Zeman, J., Penza, S., & Champion, K. (2000). Emotion management skills in sexually maltreated and nonmaltreated girls: A developmental psychopathology perspective. *Develop-ment and Psychopathology, 12*, 47–62.

Slavin, R. E. (2008). Perspectives on evidence-based research in education. *Educational Researcher, 37*, 5–14.

Staub, E. (1988). The evolution of caring and nonaggressive persons and societies. *Journal of Social Issues, 44*, 81–100.

Staub, E. (2003). Notes on cultures of violence, cultures of caring and peace, and the fulfillment of basic human needs. *Political Psychology, 24*, 1–21.

Stoolmiller, M., Eddy, M. J., & Reid, J. B. (2000). Detecting and describing preventive intervention effects in a universal school-based randomized trial targeting delinquent and violent youth. *Journal of Consulting and Clinical Psychology, 68*, 296–306.

Sutton, J., Smith, P. K., & Swettenham, J. (1999). Social cognition and bullying: Social inadequacy or skilled manipulation? *British Journal of Developmental Psychology, 17*, 435–450.

Tabachnick, B. G., & Fidell, L. S. (2001). *Computer-assisted research design and analysis*. Boston, MA: Allyn & Bacon.

Thomas, D. R., & Zumbo, B. D. (2011). Difference scores from the point of view of reliability and repeated-measures ANOVA: In defence of difference scores for data analysis. *Educational and Psychological Measurement*. doi:10.1177/0013164411409929.

Tolan, P. H., Guerra, N. G., & Kendall, P. C. (1995). A developmental-ecological perspective on antisocial behavior in children and adolescents: Toward a unified risk and intervention framework. Special Section: Prediction and prevention of child and adolescent antisocial behavior. *Journal of Consulting and Clinical Psychology, 63*, 579–584.

Tregebov, R. (1993). *Sasha and the wiggly tooth*. Toronto, ON: Second Story Press.

Tremblay, R. E. (2000). The development of aggressive behavior during childhood: What have we learned in the past century? *International Journal of Behavioral Development, 24*, 129–141.

Vitaro, F., Brendgen, M., & Tremblay, R. E. (2002). Reactively and proactively aggressive children: Antecedent and subsequent characteristics. *Journal of Child Psychology and Psychiatry, 43*, 495–505.

Weissberg, R. P., & Greenberg, M. T. (1998). School and community competence-enhancement and prevention programs. In I. E. Sie-gel & K. A. Renninger (Eds.), *Handbook of child psychology: Vol. 5. Child psychology in practice* (5th ed., pp. 877–954). New York: Wiley.

Wentzel, K. R. (1993). Does being good make the grade? Social behavior and academic competence in middle school. *Journal of Educational Psychology, 85*, 357–364.

Wentzel, K. R., Filisetti, L., & Looney, L. (2007). Adolescent prosocial behavior: The role of self-processes and contextual cues. *Child Development, 78*, 895–910.

Wilkinson, L., & Task Force on Statistical Inference. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist, 54*, 594–604.

Wilson, D., Gottfredson, D., & Najaka, S. (2001). School based prevention of problem behaviors: A meta analysis. *Journal of Quantitative Criminology, 17*, 247–272.

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*, 136–149.

Yau, M. (1999). *An evaluation of the Roots of Empathy program: Some preliminary findings, 1997–99*. Toronto District School Board: Academic Accountability Department.

Younger, A. J., Schneider, B. H., Wadeson, R., Guirguis, M., & Bergeron, N. (2000). A behavior-based peer-nomination measure of social withdrawal in children. *Social Development, 9*, 544–564.

Zahn-Waxler, C., Friedman, S. L., & Cummings, E. M. (1983). Children's emotions and behaviors in response to infants' cries. *Child Development, 54*, 1522–1528. doi:10.2307/1129815.

Zhou, Q., Valiente, C., & Eisenberg, N. (2003). Empathy and its measurement. In S. J. Lopez & C. R. Snyder (Eds.), *Positive psychological assessment: A handbook of models and measures* (pp. 269–284). Washington, DC: American Psychological Association.

Zumbo, B. D. (1999). The simple difference score as an inherently poor measure of change: Some reality, much mythology. *Advances in Social Science Methodology, 5*, 269–304.