Journal of Consulting and Clinical Psychology
2009, Vol. 77, No. 2, 302–316

0022-006X/09/$12.00 DOI: 10.1037/a0015302

Targeting Children's Behavior Problems in Preschool Classrooms: A Cluster-Randomized Controlled Trial

C. Cybele Raver
New York University

Stephanie M. Jones
Harvard University

Christine Li-Grining
Loyola University Chicago

Fuhua Zhai
New York University

Molly W. Metzger
Northwestern University

Bonnie Solomon
University of Illinois at Chicago

The present study evaluated the efficacy of a multicomponent, classroom-based intervention in reducing preschoolers' behavior problems. The Chicago School Readiness Project model was implemented in 35 Head Start classrooms using a clustered-randomized controlled trial design. Results indicate significant treatment effects (*d*s = 0.53–0.89) for teacher-reported and independent observations of children's internalizing and externalizing behavior problems. Moreover, there was some evidence for the moderating role of child gender, race/ethnic group membership, and exposure to poverty-related risk, with stronger effects of intervention for some groups of children than for others. Findings contribute to a growing area of research on poverty and preventive intervention in early childhood.

Keywords: randomized, trial, classroom-based consultation, teacher training

Recently, researchers and policy makers have expressed the concern that preschoolers' behavior problems may significantly compromise preschoolers' chances for later success in school (Gilliam, 2005; Raver, 2002). Young children who are persistently sad, withdrawn, or disruptive have been found to receive less instruction, to have fewer opportunities for learning from peers, and to be less engaged and less positive about their role as learners (Arnold et al., 2006). Young children facing economic disadvantage may be at particularly high risk. Exposed to a wide range of psychosocial stressors, children in poor neighborhoods are at greater risk for developing emotional and behavioral difficulties and have minimal access to mental health services (Fantuzzo et al., 1999). In light of the growing evidence of onset of behavior problems as early as toddlerhood (Carter, Briggs-Gowan, Jones, & Little, 2003; Shaw, Dishion, Supplee, Gardner, & Arnds, 2006), early childhood represents a particularly important time to target children's risk of behavior problems.

Given the consequences of behavioral difficulty for children's school readiness, how can children's behavior problems be reduced in early childhood? Given that 67% of young children in the United States are enrolled in center-based or non-relative care prior to enrollment in kindergarten (Innes, Denton, & West, 2001), preschool classrooms are an increasingly important service setting outside the home (Spoth, Kavanagh, & Dishion, 2002). Pathbreaking studies in the last decade suggest that targeting classroom processes can be an effective way to reduce children's behavioral problems (e.g., August, Realmuto, Hektner, & Bloomquist, 2001; Conduct Problems Prevention Research Group, 1999; Ialongo et al., 1999; Lochman & Wells, 2003; Webster-Stratton, Reid, & Hammond, 2004). However, most of those classroom-based studies targeted low-income children in early elementary grades (see Berryhill & Prinz, 2003, and Jones, Brown, & Aber, 2008, for reviews). It is unclear from these important school-based efficacy trials whether the same classroom processes hold for urban settings

C. Cybele Raver, Department of Applied Psychology and Institute of Human Development and Social Change, New York University; Stephanie M. Jones, Risk and Prevention Program, Harvard Graduate School of Education, Harvard University; Christine Li-Grining, Department of Psychology, Loyola University Chicago; Fuhua Zhai, Institute of Human Development and Social Change, New York University; Molly W. Metzger, Department of Human Development and Social Policy, Institute for Policy Research, Northwestern University; Bonnie Solomon, Department of Human Development and Social Policy, Institute of Policy Research, Northwestern University.

We gratefully acknowledge support (NICHD Award R01 HD46160-01) from the federal Interagency School Readiness Consortium, which includes the National Institute of Child Health and Human Development (NICHD), the Administration for Children and Families, the Assistant Secretary for Planning and Evaluation in the Department of Health and Human Services, and the Office of Special Education and Rehabilitation Services of the U.S. Department of Education. This research was completed with support from the William T. Grant Foundation and the McCormick Tribune Foundation. We express our heartfelt thanks to senior policy leaders and program staff at Chicago's Department of Children and Youth Services, including Vanessa Rich, Anthony Raden, and Mary Ellen Caron, as well as our thanks to the dedicated center directors, teachers, families, and Chicago School Readiness Project intervention staff, who made this work possible. Thanks also to Emily Pressler and Kathleen Zadzora for their research assistance.

Correspondence concerning this article should be directed to C. Cybele Raver, Department of Applied Psychology, New York University, Kimball Hall, 246 Greene Street, Room 518E, New York, NY 10003. E-mail: cybele.raver@nyu.edu

in which younger children are served and where preschool teachers have substantially lower levels of training, salary, and support, on average, than do teachers in elementary schools (Granger & Marx, 1992).

In short, this study addresses a gap in our understanding of the types of classroom-based interventions that might reduce behavior problems among low-income children in preschool settings. Results of a recent nationally representative sample suggest that low-income children experienced short-term academic benefit from having attended preschool but that their emotional and behavioral adjustment was placed at substantially greater risk in the long run (Magnuson, Ruhm, & Waldfogel, 2007). This "trade-off" is alarming and signals the need for targeting preschool classroom processes that might support rather than compromise young children's emotional and behavioral development. In sum, it is imperative to learn whether interventions that target social–emotional development in preschool can avert the risk of higher behavior problems among low-income children and also support their emotional, behavioral, and academic adjustment.

Moderating Role of Economic Risk, Race/Ethnicity, and Gender

Recent research in prevention science has highlighted the ways in which low-income children of color face both significantly higher risks of behavioral difficulty and large disparities in their access to mental health and behavioral health services (Fantuzzo et al., 1999; Yoshikawa & Knitzer, 1997). As a key step in closing the gap in these behavioral health disparities, we need to learn whether interventions demonstrate similar or different levels of efficacy for boys and girls and for children in different sociocultural and socioeconomic contexts (Knight & Hill, 1998). Similarly, recent findings from a number of efficacy trials suggest that interventions demonstrate significantly stronger impacts for families facing a greater versus smaller number of poverty-related risks (Aber, Jones, Brown, Chaudry, & Samples, 1998; Tolan, Gorman-Smith, & Henry, 2004). Based on these bodies of research, child gender and family cumulative exposure to poverty-related risks were hypothesized to play key moderating roles in the efficacy of our intervention. Finally, Hispanic children represent the fastest growing group of children in poverty in cities such as Chicago, and programs increasingly identify that their services must meet the needs of Hispanic children as well as those of African American children (Goerge, Dilts, Yang, Wasserman, & Clary, 2007). Accordingly, we examined the moderating role of children's racial/ethnic status when testing the efficacy of the intervention.

The Current Study

The Chicago School Readiness Project (CSRP) intervention is based on several theoretical models of preschool children's behavioral problems in early educational settings. First, much recent research suggests that children's behavioral difficulty may be part and parcel of relatively low quality of care in preschool classrooms (Ritchie & Howes, 2003). For example, teachers are expected to manage large numbers of preschoolers in their classrooms (NICHD Early Child Care Research Network, 1999), though they often have little training or support in effective methods of classroom management. In this model, children with more emotional and behavioral difficulty engage in escalating, emotionally deregulating "coercive processes" with teachers (Arnold, McWilliams, & Arnold, 1998; Kellam, Ling, Merisca, Brown, & Ialongo, 1998). On the basis of this theoretical framework, we provided intensive training in strategies that teachers could employ to provide their classrooms with more effective regulatory support and better classroom management as one mechanism to reduce children's behavior problems (Raver et al., 2008).

Classroom-based research suggests a second, complementary theoretical model, whereby teachers may experience "burnout" marked by emotional exhaustion and depersonalization as a result of trying to meet too many classroom demands with too little support (Brouwers & Tomic, 2000). Review of this literature suggests that teachers might be unlikely to take new, proactive steps to support children's behavioral self-regulation, if teachers themselves feel unsupported (Woolfolk, Rosoff, & Hoy, 1990). This framework led us to include two additional components to the model. As part of a model of classroom- and child-centered consultation, a weekly Mental Health Consultant (MHC) provided 20 weeks of consultation as a "coach" who supported teachers while they tried new techniques learned in the teacher training (Donohue, Falk, & Provet, 2000; Gorman-Smith, Beidel, Brown, Lochman, & Haaga, 2003). As an additional program component, MHCs spent a significant portion of the school year (in winter) conducting stress reduction workshops to help teachers reduce stress and limit burnout. One critique might be that MHCs bring "an extra pair of hands" to the classroom in addition to their clinical expertise. As a control for improvements in adult–child ratio introduced by the presence of MHCs in treatment classrooms, control group classrooms were assigned a lower cost Teacher Aide (TA) for the same amount of time per week.

A final theoretical model drawn from child clinical research suggests that low-income preschoolers face much higher likelihood of exposure to a range of poverty-related risks that pose serious threats to their mental health and school adjustment (Campbell, 1995; Shaw et al., 2006). Exposure to high levels of family and community violence may represent a particularly pernicious threat to the emotional and behavioral adjustment and school readiness of young, low-income children (Margolin & Gordis, 2000). In recent research, up to 30% of inner-city, low-income preschool and school-age children were reported to have been exposed to severe violence that included robbery, threats with a weapon, or shooting or stabbing (Randolph, Koblinsky, & Roberts, 1996). Given such findings, it is clear that many children in inner-city, economically disadvantaged communities are likely to come to preschool with considerably more emotional and behavioral concerns than preschool teachers can reasonably handle. Although Head Start is mandated to serve children with special needs, children with emotional and behavioral problems remain substantially underreferred and underserved for emotional disturbance (Fantuzzo et al., 1999). In short, this model suggests that MHCs could aid parents, teachers, and children by providing both classroom-based and individualized consultative services (Keenan et al., 2007; Yoshikawa & Knitzer, 1997). Thus, a fourth component of the CSRP model is the provision of child-focused mental health consultation for three to five children in each classroom in the late spring of the preschool school year.

Our principal aim in the CSRP intervention was to marshal these primary programmatic components to improve the school readiness of low-income preschool-age children by increasing their emotional and behavioral adjustment. Our immediate research aim, addressed in this article, was to test whether this multicomponent intervention yields short-term benefits by reducing children's behavioral problems in the spring of the children's preschool year. Following Flay et al.'s (2005) recommendation to use multiple methods, we assessed the impact of the CSRP intervention for the full sample of CSRP-enrolled children via survey methods more commonly used in school settings. In addition, we used a more labor-intensive observational method to assess children's problem behaviors (including aggressive and nonaggressive disruptive behavior as well as withdrawn and disconnected behaviors), in which a small subsample of focal children within each classroom is randomly selected and observed (Miller, Gouley, Seifer, Dickstein, & Shields, 2004). Our aim was to capitalize on the strengths of each methodological approach and to balance research limitations of feasibility and cost. We hypothesized that the CSRP intervention would significantly benefit children enrolled in treatment-assigned Head Start programs as compared with children enrolled in control group assigned programs, and we predicted that observational findings would be consistent with (if not confirmatory of) results yielded from teacher reports.

Hypotheses

First, we expected that the intervention would decrease children's internalizing and externalizing behavior problems by spring of the Head Start school year. We expected that the precision of our estimates of the impact of treatment would be greater when we took important baseline child-, teacher-, and classroom-level covariates into account. Second, we expected that estimates of the impact of intervention would be moderated by children's exposure to higher levels of poverty-related risk at baseline, with effects of the intervention larger for children facing higher levels of poverty-related risk. Third, we expected that the impact of intervention might be moderated by children's racial/ethnic status. Because research on racial/ethnic differences in treatment effects is limited, no specific hypotheses were made. Fourth, we examined whether the intervention had a significantly larger impact on girls than on boys.

The present study capitalizes on recent methodological advances in prevention science and educational research in which the impact of interventions on children's outcomes is considered to be "nested" within classroom and institutional contexts. Recent school-based prevention trials aimed at supporting low-income youths have demonstrated the importance of disaggregating potential confounds such as family income (e.g., August, Egan, Realmuto, & Hektner, 2003; Tolan et al., 2004). In addition, recent research suggests that interventions may work quite differently in settings marked by low versus high levels of institutional resources, teacher education, and motivation (Gottfredson, Jones, & Gore, 2002). Building on these recent innovations in educational research, we employed a cluster-randomized experimental design to evaluate the efficacy of the CSRP model. A "settings-level" approach (with the inclusion of a large number of child-, classroom-, and school-level baseline characteristics) has been argued to increase the precision of treatment impact estimates (Bloom, 2005; Cook, 2005). This study provides us with an opportunity to examine the merits of these analytic approaches for a classroom-based preventative intervention.

Method

Sample

Following recent school-based intervention models, this study used a cluster-randomized design. As such, random assignment occurred at the site level, with matched pairs of Head Start–funded programs assigned to treatment and control conditions.

School and participant selection. In an effort to balance generalizability and feasibility, we selected preschool sites on the basis of (a) receipt of Head Start funding, (b) having two or more classrooms that offered "full day" programming, and (c) location in one of seven high-poverty neighborhoods (see Raver et al., 2008, for a detailed discussion of exclusionary criteria). CSRP staff completed block-by-block surveys of all seven neighborhoods, in which all child-serving agencies were identified and screened to determine whether they met site selection criteria (including receipt of Head Start funding). Eligible sites were then invited to self-nominate for participation in the research project. Eighteen sites across seven neighborhoods completed the process and were included as CSRP sites, and two classrooms within each site were randomly selected for participation. Research staff successfully recruited 83% of the children enrolled in classrooms between Labor Day and the assigned enrollment cutoff date in mid-October of the school year. Teacher reports of child behavior problems were collected for the full sample across fall and spring of the school year (see Figure 1). In addition, observational assessments of children's externalizing/disruptive and internalizing/disconnected behavior were collected for a stratified, randomly selected subsample of the full sample.

Randomization. Each site was matched with another "sister" site that most closely resembled it on a range of demographic characteristics of families and site characteristics indicating program capacity. Methods employing sum of squared distances and sum of absolute distances were used to estimate best matches for pairs of sites across 14 site-level demographic characteristics (list available from C. Cybele Raver upon request). One member of each pair was randomly assigned to treatment, and the other member of the pair was assigned to control group. Within each of the nine treatment sites, 2 classrooms participated, for a total of 18 treatment classrooms. Across the nine control sites, there were 17 classrooms (2 classrooms in eight sites, and 1 classroom in the remaining site, which lost a Head Start–funded preschool classroom due to funding cuts). Treatment classrooms received the multiple components of the intervention package across the school year, and control classrooms were paired with TAs as described above. Additional information on program design and implementation is detailed below.

The CSRP intervention was implemented for two cohorts of children and teachers, with Cohort 1 participating in 2004–05 and Cohort 2 participating in 2005–06. As with other recent efficacy trials implemented with multiple cohorts, the sites enrolled in Cohorts 1 and 2 differed on several program-level and demographic characteristics, and therefore those characteristics were included in all analyses (e.g., Gross et al., 2003).

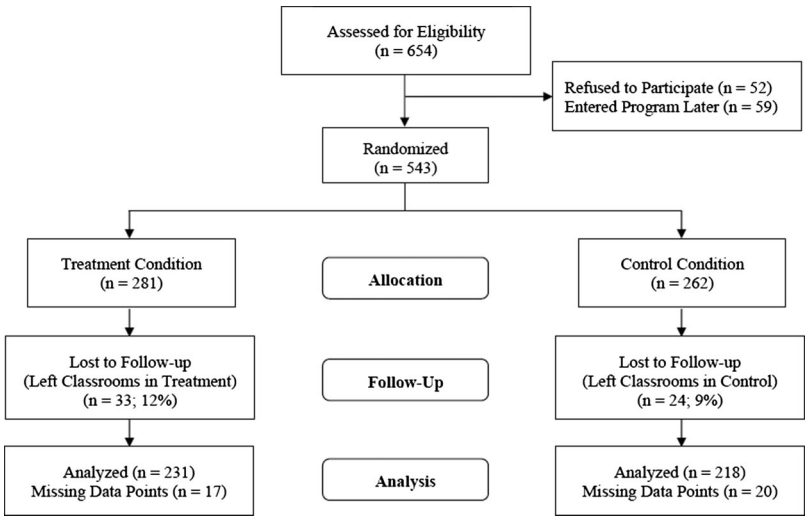

Figure 1. Participant flowchart for Chicago School Readiness Project.

Because we planned to model child outcomes as potentially responsive both to the intervention and to teacher- and classroom-level characteristics, teachers were included as participants. Teachers were enrolled in two cohorts that were also pooled into a single data set ($N = 90$). A total of 87 teachers participated at baseline. The number of teachers increased to 90 by the spring of the Head Start year. This net increase reflected the entry of 7 more teachers and the exit of 4 teachers who either moved or quit during the school year.

At baseline, a total of 543 children participated in CSRP. By the spring, the number of participating children was reduced to 509. Nearly all of the exits were due to children voluntarily leaving the Head Start program, though 1 child was requested to leave the Head Start program and one parent opted to withdraw her child from participating in CSRP.[1]

The sample for analyses of teacher-report data in this study included 449 children, who had complete data on children's behavior problems in the fall and spring, as well as child and family background characteristics in the fall. Representativeness analyses compared children across children's gender, race/ethnicity, risk of behavior problems in the fall, whether parents were single, whether families included four or more minors, whether parents spoke Spanish, and family risk factors. Overall, analyses revealed no significant differences between children who were in the analytic sample and children who were excluded due to missing data. However, the analytic sample included a lower percentage of parents who spoke Spanish, $\chi^2(1, N = 602) = 18.341$, $p < .001$. Similarly, there were no significant differences across analytic and omitted samples of teachers in terms of teachers' age, education level, and depressive symptoms. The one exception was that differences were found in levels of teacher job overload, with teachers in the analytic sample reporting lower job demands and control, $t(600) = 3.314$, $p < .01$, and $t(600) = -2.414$, $p < .05$, respectively. These variables were included as covariates in all analyses.

Because of the importance of multimethod, multimeasure approaches to estimates of intervention efficacy (Flay et al., 2005), independently collected observational assessments of children's problem behaviors were conducted for a stratified, randomly selected subsample of children within the CSRP sample. Roughly 6 children per classroom were selected; they represented low (z score < 0), average ($0 < z$ score < 1), or high ($1 < z$ score) levels of behavior problems based on fall (baseline) teacher reports on the Behavior Problems Index (described below). Each child's z score was centered by child gender and classroom membership. The sample size for analysis was 172 (54% female), and percentages of each racial/ethnic group in the observational subsample matched those in the total CSRP sample.

Design

As described earlier, the CSRP intervention model comprises four specific components: (a) teacher training in behavior management strategies, (b) MHCs' provision of "coaching" to teachers in implementing these strategies in the classroom, (c) MHCs' provision of stress reduction workshops, and (d) MHCs' provision of targeted, direct services for children with the highest emotional and behavioral problems. The intervention emphasized three central principles—MHC–teacher collaboration, MHCs' cultural competence, and sustainability—as well as a dual focus on supporting behavioral change at the classroom and child levels. The four components of teacher training, coaching, stress reduction, and direct, "one-on-one" child-focused services were delivered by consultants to classrooms randomized to treatment. The full model was manualized, and intervention staff underwent 2 days of training in implementation, as well as clinical supervision (once every

[1] Attrition was due to the exit and entry of two groups of children. The group of 543 children was reduced when 88 children exited the study, leaving 455 children who entered at baseline and remained in the study. In addition to the original group at baseline, 59 children entered the study later in the school year (5 of these children subsequently exited and 54 of them remained). Thus, there were 509 children participating in the study by the spring. Attrition analyses suggest that differences in exits and entry between intervention and control groups were minimal and unlikely to bias analyses of treatment impact (see Raver et al., 2008).

2 weeks) and administrative supervision (once every week) throughout the year.

Treatment classrooms: Teacher training. Treatment teachers were invited to participate in 30 hr of workshop-style training sessions (across 5 Saturdays in October through January). These trainings, adapted from the evidence-based Incredible Years Teacher Training Program (Webster-Stratton et al., 2004) and led by a licensed clinical social worker (and experienced trainer), apply behavioral principles to teachers' approaches to reducing children's challenging behaviors. MHCs attended the sessions as well, as one way of fostering their collaborative partnership with teachers. Teacher incentives included payment ($15 hr), catered lunches, and on-site child care. Concerns regarding the relevance and acceptability of the training program for ethnic minority teachers of young African American and Hispanic children were addressed by piloting the training component in two demonstration sites not included in the full study. All training materials were translated into Spanish, and piloting and follow-up focus groups were conducted in English and Spanish with 12 teachers in two pilot sites that served low-income African American and Hispanic families. Teachers indicated that the training and coaching formats were "helpful" to "very helpful."

Mental health consultation: Coaching and stress reduction. Teacher training was combined with weekly provision of mental health consultation to classrooms, anchored in clinically trained consultants' provision of "coaching" and stress reduction strategies. Training alone may not ensure that teachers transfer learned material to their classrooms, given the everyday challenges of running a classroom smoothly (Gorman-Smith et al., 2003; Wasik, Bond, & Hindman, 2006). Therefore, during the first 10 weeks (or the first third) of the intervention, MHCs followed specific coaching steps to help teachers strengthen their ability to promote children's positive emotional and behavioral development: establishing shared goals with teachers, observing teacher–child interactions, sharing and discussing feedback, engaging in collaborative problem solving, and supporting the use of specific techniques (for more details, see Madison-Boyd et al., 2006). During the second third of the intervention, MHCs held a 1-day stress reduction workshop for each Head Start site. MHCs also discussed teachers' strategies for relieving stress during class visits based on the premise that teachers might be better able to focus on meeting the needs of children if they themselves felt supported (Curbow, 1990).

Mental health consultation: Direct service to children. In the last 10 weeks of the intervention, MHCs provided targeted, direct intervention services (including individual and group therapies) to a small number of children. During the first 6 months of the intervention MHCs identified 3 to 4 children per class (approximately 70 in total) on the basis of (a) clinical judgment, (b) consultation with teachers, and (c) review of teacher-reported measures of children's behavioral problems in the fall. In addition, MHCs recorded the number of school visits spent working with individual children as well as the primary clinical concerns addressed during one-on-one sessions with individual children (see below).

Program satisfaction. Exit survey data (from Cohort 1 treatment-group teachers) suggested high acceptability of mental health consultation. Over 80% of teachers reported that MHCs were "somewhat" to "very" helpful in strengthening classroom rules and routines, helping teachers to build positive behaviors with children who were "difficult," and allowing teachers to spend more time teaching in groups. Also, 91% of teachers reported that having MHCs allowed them to devote time to individual children who needed extra help. When asked whether they would choose to have a MHC in their classroom again, 89% of teachers said "yes."

Control classrooms. Classrooms within sites randomly assigned to the control group were paired with associate's degree–level TAs during the intervention year. Control classroom-assigned TAs provided an "extra pair of hands and eyes" during everyday classroom activities.

Facilitators and providers: Program coordinator, clinical supervisor, and trainer. The program coordinator acted as a community liaison and helped ensure that the project ran smoothly across the 35 classrooms and 18 sites participating in CSRP. For example, the program coordinator organized all teacher trainings and encouraged teachers in the intervention group to register for the training sessions. The program coordinator also monitored TAs in the control condition on the number of hours they spent in control classrooms. In the treatment condition, the MHCs received biweekly clinical supervision from a licensed clinical social worker with a master's degree. A second licensed clinical social worker served as a leader of teacher training using the Incredible Years Teacher Training Program (Webster-Stratton et al., 2004) and had extensive prior experience leading parent and teacher groups in the Chicagoland area.

MHCs. Each MHC held a master's degree in social work. Given the focus of CSRP, MHCs were required to have had experience working in early childhood settings and with families facing multiple poverty-related risks. Additional emphasis was placed on MHCs' ability to deliver culturally competent services; languages spoken (e.g., Spanish); and cultural match across MHCs, teachers, and children when hiring decisions were made.

Treatment fidelity. The fidelity of implementation of the CSRP model was monitored in several ways. These included tracking (a) the number of trainings attended by all intervention-assigned teachers and TAs, (b) records of the number of weekly visits made by MHCs to their assigned classrooms, (c) MHCs' detailed weekly and monthly reports of content of all classroom- and child-focused consultation services, and (d) ratings by both the MHCs and intervention-assigned teachers on the extent to which intervention services had been successfully delivered and implemented.

Analysis of (a) teacher attendance at trainings suggests that teachers attended three of the five trainings, on average. Nearly all teachers attended the stress reduction workshop (held on site at CSRP-enrolled Head Start sites). MHCs followed up with a review of training topics with all treatment-assigned teachers (including those who might have missed an earlier training session) during classroom visits. Records suggested (b) that MHCs completed 29 classroom visits to their assigned classrooms, on average, across the school year (range = 21–40). This resulted in classrooms participating in an average of 128 hr of mental health consultation by the end of the school year. Service delivery was relatively evenly distributed across programs (Li-Grining et al., 2007).

In addition, (c1) MHCs' classroom- and child-focused consultation services were monitored via weekly reports completed at the end of each classroom visit. MHCs reported on the number of social services they provided (e.g., coaching on strategies covered

in teacher training, direct services to children), the number of strategies from teacher training sessions that teachers tried, and the quality of this implementation. Review of these weekly service provision reports suggests that MHCs coached teachers in the use of the five specific behavior management strategies taught during the trainings in every classroom, throughout the school year. Every month, (c2) MHCs submitted spreadsheets (and accompanying treatment plans) that identified the number of children to whom they had delivered direct services. These monthly reports identified dates of visits, child ID, and the nature of the behavioral concern. Review of these logs showed that 137 children received child-focused consultation services and that individual children received 5.49 MHC visits, on average (range = 1–23).

Finally, (d1) teachers rated the quality of trainings as very helpful and rated MHCs, on average, as somewhat to very helpful in providing classroom- and child-focused consultation. Ratings were also obtained (d2) from MHCs on the degree to which teachers were successful in trying the strategies in their classrooms. In their weekly reports, MHCs rated teachers as "somewhat successful," on average, in teachers' use of classroom management strategies that were covered in trainings.

Procedures

Data collection. In the fall, families with children ages 3–4 were recruited from each of the 35 classrooms to participate in the study. Approximately 17 children in each classroom were enrolled in CSRP. Consent forms for each child were signed by his or her parent or guardian, who also completed a demographic interview (e.g., parent's marital status and education level).

Children's behavior problems were rated by teachers in both fall and spring of the Head Start year. Teachers and TAs were given the Behavior Problems Index (BPI; Zill, 1990) in the fall and spring, as well as the Caregiver–Teacher Report Form (C–TRF; Achenbach & Rescorla, 2001) in the spring, for each CSRP-enrolled child in their classrooms. Teachers' reports were collected by CSRP research staff (blind to the treatment status of the classroom) within 6 weeks, and teachers were reimbursed a nominal participant payment ($20 per packet).

For the observational subsample of CSRP, we selected the observer rating version of the Penn Interactive Peer Play Scale (PIPPS; developed by Fantuzzo et al., 1995; adapted by Milfort & Greenfield, 2002) as a valid coding system with low-income, ethnic minority children. It can be reliably coded "in vivo" to yield data on low-frequency behaviors, such as children's aggressive and withdrawn behaviors (see Miller et al., 2004). Children's externalizing/disruptive and internalizing/disconnected behaviors were observed in 20-min blocks during the course of the school day by the coding team (see below).

Both trained observers and teachers provided classroom-level data in the fall to account for classroom-level differences in resources and support for children's social–emotional development. Trained observers, who were blind to randomization, assessed the quality of children's classrooms using the Classroom Assessment Scoring System (CLASS; La Paro, Pianta, & Stuhlman, 2004) and the Early Childhood Environment Rating Scale, Revised Edition (ECERS–R; Harms, Clifford, & Cryer, 2003). The team consisted of 12 individuals who each had at least a bachelor's (BA) degree. Of the 12 members, 6 were African American and 6 were Caucasian or Asian; thus, approximately half the time, the race of the observer matched that of most children. Using the ECERS–R and CLASS, observers rated overall classroom quality as well as dimensions of emotional climate (e.g., teacher sensitivity, behavior management, negative climate). While conducting observations, staff noted the number of children and adults in the classroom. Teachers and their assistants also completed self-reports of demographic information (e.g., level of education) and psychosocial well-being (e.g., depressive symptoms; Kessler et al., 2002).

In the fall, administrators at each Head Start site provided CSRP with access to site-level characteristics regarding staff (e.g., percentage of teachers with a BA, availability of a family support worker on site). Administrators also provided site-level data on children and families (e.g., total enrollment, percentage of families led by single parents).

Measures

Dependent measures of intervention impact. To assess the impact of the CSRP intervention on children's behavior problems, we included two teacher-reported measures of children's internalizing and externalizing behavior problems. The Achenbach System of Empirically Based Assessment profile entitled the Caregiver–Teacher Report Form (C-TRF; Achenbach & Rescorla, 2001) was completed by teachers in May. The measure consists of 100 items asking the respondent to rate the child on a scale from 0 to 2 (0 = *not true,* 1 = *somewhat or sometimes true,* 2 = *very true or often true*). Responses were summed into Internalizing (α = .90) and Externalizing (α = .97) subscales. For all hierarchical linear modeling (HLM) analyses, we used raw Internalizing and Externalizing scores to avoid truncating variance in teacher report scores (Achenbach, Edelbrock, & Howell, 1987). Standardized *t*-score values were calculated so that the percentage of CSRP-enrolled children with "elevated" levels of behavioral problems (with *t* scores $\geq$60) could be included in descriptive analyses.

In addition, in both the fall and spring, teachers completed the Behavior Problems Index (BPI), a 28-item rating scale originally designed for parent report of child behavior and adapted from multiple studies of children's behavior problems (Zill, 1990). CSRP modified the original version in several minor ways. For the purposes of this study, items were summed into Internalizing (α = .80) and Externalizing (α = .92) subscales, following the National Longitudinal Survey of Youth (NLSY79; Zill, 1990). For both the BPI and C–TRF, children's scores were averaged across the two reporters (i.e., the child's teacher and TA).

To provide independent observational assessments of children's behavior problems, CSRP received permission from the author, John Fantuzzo, to use an observer rating version of the measure that has been shown to be valid and reliable (Fantuzzo et al.,1995; Milfort & Greenfield, 2002). The original measure was altered slightly. The CSRP version included 30 of the 32 original items and used a dichotomous rating that indicated whether or not the coder observed the specific behavior (rather than a 4-point scale that ranged from *never* to *always*). Interrater reliability was high, ranging from an alpha of .79 to .92 for 30% of the cases (where a second coder observed 2 out of 6 of the children in each classroom). On the basis of results from principal-components analysis, we created three subscales, two of which met adequate levels of

interitem reliability (Aggression/Disruption, $\alpha = .71$; Withdrawal/Disconnection, $\alpha = .64$) and were subsequently used in analyses. Concurrent validity has been demonstrated for the Aggression/Disruption subscale, which was positively correlated with the Externalizing subscale of the C–TRF ($r = .16$, $p < .05$).

Child-level covariates. Child-level demographic characteristics were included in the following analyses. These included (a) child gender, (b) child membership in the race/ethnic category of African American versus Hispanic, (c) parent's self-identification as Spanish-speaking in the home, (d) large family size (with ≥4 children), (e) single-headed household, and (f) family's cumulative exposure to three poverty-related risks (i.e., mothers' educational attainment of less than high school degree, family income-to-needs ratio for the previous year being less than half the federal poverty threshold, mothers' engagement in 10 hr or fewer of employment per week; Raver, 2004). A final child-level covariate was the child's total behavioral problems score (as rated by teachers on the BPI) in the fall.

Classroom/teacher-level covariates. A number of teacher characteristics were included as proxy assessments of classroom quality and were assessed through teacher report. These included teacher reports regarding their age, level of education (teacher's attainment of BA as well as teacher assistant's attainment of BA), and teacher reports on several psychosocial characteristics that might affect teachers' perceptions of children's behavioral difficulty (see Anthony, Anthony, Morrel, & Acosta, 2005). To assess teachers' psychosocial characteristics, we briefly assessed teachers' depressive symptoms at baseline using the 6-item K6, a scale of psychological distress developed for the U.S. National Health Interview Survey (Kessler et al., 2002). The K6 items are coded as 0–4 and then summed ($\alpha = .65$). In addition, teachers reported job overload on the 6-item Job Demands and 5-item Job Control subscales of the Child Care and Early Education Job Inventory (Curbow, Spratt, Ungaretti, McDonnell, & Breckler, 2000). Rating was on a scale of 1–5. Subscales demonstrated adequate internal consistency ($\alpha = .67$ and $\alpha = .56$, respectively) and were summed. To calculate classroom-level covariates, we averaged scores on each variable across all teachers in each classroom (Gerard & Buehler, 2004).

To control for additional variation in classroom quality, we collected observational measures in the fall using the CLASS (La Paro et al., 2004) and the Early Childhood Environment Rating Scale—Revised (ECERS-R; Harms et al., 2003). CLASS indicators included 7-point Likert scores on negative climate, teacher sensitivity, and behavior management (see Raver et al., 2008). Three quarters of the observations were double coded "live" by two observers, and intraclass correlation values (α) indicated adequate to high levels of interobserver agreement (negative climate, $\alpha = .70$; teacher sensitivity, $\alpha = .77$; behavior management, $\alpha = .66$).

The 43-item ECERS–R (Harms et al., 2003) is a widely used research tool that measures early childhood classroom quality. Items are scored on a scale of 1–7 (Harms et al., 2003). The ECERS–R data were collected during the fall of each year by the same cadre of observers who collected the CLASS data, and 43% of the ECERS–R observations were double coded for purposes of reliability ($\alpha = .87$ for the ECERS–R total score). The number of children and the number of teachers observed in each classroom in September were included to control for the potential confounds of differences in class size or staffing ratios.

Site-level covariates. As a test of the role of "settings-level" program characteristics, a limited number of site-level covariates were entered into models. The covariates included the availability of a full-time family worker at the Head Start site; the size of the program (i.e., the number of children ages 3–5 served); the proportion of the site identified as African American; the proportion of teachers with BA degrees and the proportion of teacher assistants with some college; and the proportion of families served that were single-parent families, employed, and reliant on Temporary Assistance for Needy Families.

Analytic Approach

In this study, we first employed teachers' reports of children's internalizing and externalizing behavior problems on the BPI and the C–TRF in May as dependent measures of CSRP program influence. Across all of these analyses, we set alpha to equal .05 for assessing statistical significance. We then repeated our analyses, with the goal of finding converging evidence of CSRP intervention influence, using observational PIPPS assessments of children's externalizing/disruptive and internalizing/disconnected behaviors collected in May for a stratified, randomly selected subsample of children ($n = 172$). Given that our power to detect the benefits of the intervention was substantially reduced with the smaller subsample, alpha for detecting statistical significance was set to .10 for these additional analyses. In both sets of analyses, a multilevel strategy was necessary because children in this study were nested in classrooms and sites. Multilevel modeling allows for the simultaneous estimation of variance associated with individual (within-participants) and population (between-participants) change based on the specification of fixed- and random-effect variables in the model (Raudenbush & Bryk, 2002). With these data it was possible to assess the direct impact of school- or classroom-level (e.g., intervention vs. control) variables, net of person-level (e.g., demographic and ecological characteristics of children and families), classroom/teacher-level, and site-level characteristics.[2]

The overall impact of intervention was then modeled with three equations, with the equation at Level 1 (child level) specified in the following way:

$$Y_{ijk} = \pi_{0jk} + \sum_{m} \pi_{mjk} X_{mijk} + \varepsilon_{ijk}$$

where Y_{ijk} is the behavioral problem score of child i in classroom j within CSRP site k; $\sum_{m} \pi_{mjk} X_{mijk}$ represents the sum of m child characteristics, such as gender, race/ethnicity, and pre-intervention BPI score, as well as family characteristics (e.g., household size) and number of other family risks (i.e., low level of parental education). ε_{ijk} is a random error term.

Correspondingly, Level 2 (classroom level) was specified in the following way:

[2] Our first set of models includes teacher assessment of children using the BPI in fall and spring of the Head Start year. This allowed for estimates of residualized change in children's BPI scores as an indicator of intervention impact. Additional models include teacher-reported C–TRF scores and observer-reported disruptive behaviors in spring as dependent variables, with children's fall BPI score as a covariate. These analyses yield estimates of intervention-control differences in children's behavioral difficulty, net of baseline behavioral risk (i.e., fall BPI scores).

$$\pi_{mjk} = \beta_{m0k} + \sum_{n} \beta_{mnk} C_{mnjk} + r_{mjk}$$

where $\sum_{n} \beta_{mnk} C_{mnjk}$ is the sum of n teacher characteristics (e.g., whether the teacher has a BA degree) and classroom characteristics such as baseline classroom quality (e.g., ECERS–R scores).

A third equation specifying Level 3 (site level) is then written as

$$\beta_{m0k} = \gamma_{m00} + \gamma_{001} T_k + \sum_{p} \gamma_{m0pk} S + u_{m0k}$$

where T_k is treatment/control assignment and $\sum_{p} \gamma_{m0pk} S$ represents the sum of p site-level characteristics, such as whether the site had additional family support worker on staff. β_{00k}, the adjusted mean level of child behavior problems in site k, varies as a function of whether or not the site was assigned to the intervention or control group; γ_{000} is the adjusted mean level of behavioral problems across all control group sites; and γ_{001} is the intervention effect. Though not shown here, $\gamma_{100} - \gamma_{800}$ represent the pooled within-site regression coefficients for the Level 1 covariates. The magnitude of intervention impact can then be examined, where γ_{001} represents the average difference between intervention and control sites, controlling for all covariates. Effect sizes were calculated by dividing that difference by the full sample's standard deviation for the dependent variable.

We first used HLM analyses, as specified above, to estimate the intervention effects of CSRP. We further investigated whether the intervention impacts (if any) were moderated by child racial/ethnic status, child gender, or children's low versus high exposure to poverty-related risks. Given our study's small sample size (i.e., $n = 18$ sites randomized to control and intervention conditions) and the known difficulty in detecting interaction effects with small sample sizes (McClelland & Judd, 1993), omnibus tests of interactions between intervention and the three possible moderators were followed with post hoc analyses of intervention impact within subgroups (see Figure 2). Finally, a third alternative model specification was included, in which the eight pairwise site assignments were entered as dummy-coded binomial variables at Level 3 of the HLM models in place of the large list of site-level covariates. This provided us with a sensitivity test of the impact of intervention on child behavior problems. In the third alternative model specification, the estimate of intervention is yielded net of the role of specific matched pairs of sites, which represent the observed and unobserved heterogeneity among children enrolled in different types of CSRP Head Start sites with differing resources, across different neighborhoods. Implications of these different model specifications are discussed below.

Results

Table 1 presents descriptive statistics for all predictors of children's behavior problems at the site, classroom, and child levels. As can be seen from the descriptive statistics in Table 1, many measures of poverty-related risk and of children's behavioral problems were higher in intervention than in control sites at baseline, though our analyses suggest that these differences are not statistically significant (Raver et al., 2008). This heterogeneity among sites and classrooms reinforces the importance, however, of including classroom- and site-based covariates when analyzing intervention impact. Mean levels of teacher-reported and observed behavior problems among CSRP-enrolled children, as well as the percentages of children whose fall C–TRF t scores were at or above a "clinically elevated" cutoff score of 60, are also presented in Table 1.

Preliminary, unconditional models were first run to ascertain the portions of variance in each dependent variable that could be attributed to classroom and site levels (see Table 2). Three sets of models were then run to yield the "intent to treat" estimates of CSRP intervention services on children's internalizing and externalizing behavior problems. For each dependent variable, Model 1 included all Level 3 covariates; Model 2 included interaction terms between intervention and child race/ethnicity, gender, and family socioeconomic risks; Model 3 included site-pair dummy variables instead of site-level covariates. For ease of interpretation, results from Model 2 are presented in Table 3.

With Internalizing and Externalizing BPI scores as the dependent variables, the results from Models 1 and 2 suggest that, overall, CSRP benefited children in the intervention group (i.e.,

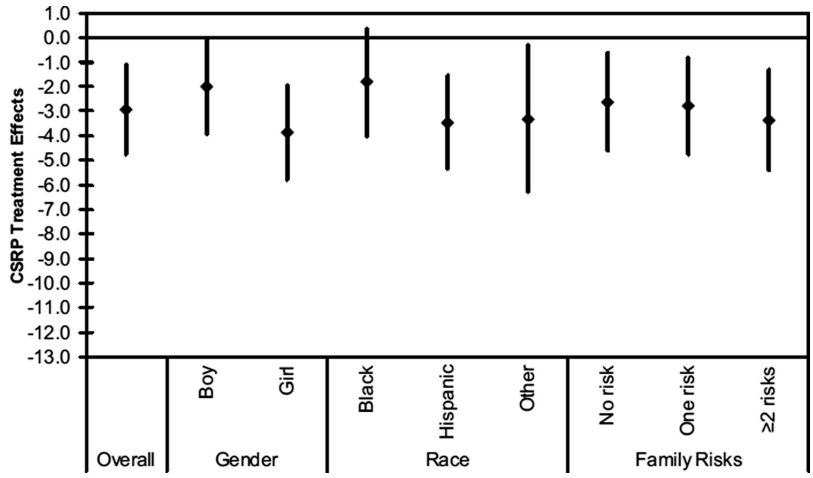

Figure 2. Effects of CSRP treatment on child BPI Externalizing scores. Dots represent point estimates; lines represent their 95% confidence intervals. CSRP Chicago School Readiness Project; BPI = Behavior Problems Index.

Table 1
CSRP Descriptive Statistics (N = 547)

Variable	Overall sample	Treated group	Control group
	Outcome variables		
BPI Internalizing score	1.59 (1.98)	1.98 (2.19)	1.18 (1.65)
BPI Externalizing score	4.18 (4.57)	5.09 (5.16)	3.26 (3.65)
C–TRF Internalizing raw score	4.21 (4.70)	5.46 (5.21)	2.93 (3.71)
% with elevated Internalizing C–TRF *t* score	5.5%	8.5%	2.6%
C–TRF Externalizing raw score	7.84 (9.58)	9.94 (11.06)	5.69 (7.21)
% with elevated Externalizing C–TRF *t* score	9.1%	11.9%	6.3%
PIPPS Disruption score[a]	1.60 (1.88)	1.57 (1.85)	1.64 (1.91)
PIPPS Disconnection score[a]	1.18 (1.52)	1.12 (1.48)	1.23 (1.56)
	Child and family characteristics		
Child age (months) at spring	58.17 (7.47)	58.32 (7.45)	58.01 (7.50)
Child gender: boy	0.47 (0.50)	0.53 (0.50)	0.42 (0.49)
Child race/ethnicity (Black)	0.65 (0.48)	0.66 (0.48)	0.64 (0.48)
Child race/ethnicity (Hispanic)	0.28 (0.45)	0.28 (0.45)	0.27 (0.45)
Child BPI fall risk	0.29 (0.45)	0.31 (0.46)	0.26 (0.44)
Family poverty-related risks[b]	1.08 (1.00)	1.15 (1.01)	1.00 (0.99)
Less than high school degree	0.26 (0.44)	0.27 (0.44)	0.25 (0.43)
Income less than half the poverty line	0.43(0.50)	0.46 (0.50)	0.39 (0.49)
Less than 10 hr of work per week	0.41 (0.49)	0.43 (0.50)	0.39 (0.49)
Single families	0.69 (0.46)	0.70 (0.46)	0.67 (0.47)
Four or more children in household	0.25 (0.43)	0.25 (0.43)	0.25 (0.44)
Parent Spanish speaking	0.22 (0.42)	0.19 (0.39)	0.25 (0.44)
	Teacher and class characteristics		
Teacher BA	0.63 (0.48)	0.66 (0.47)	0.61 (0.49)
Teacher age in years	40.44 (11.66)	37.63 (12.08)	43.30 (10.50)
Teacher K6 score	2.50 (1.99)	3.13 (1.63)	1.86 (2.12)
Teacher job demand	2.71 (0.59)	2.88 (0.63)	2.54 (0.48)
Teacher job control	3.26 (0.68)	3.34 (0.69)	3.19 (0.67)
Teacher behavior management	4.89 (1.04)	4.62 (1.08)	5.17 (0.93)
Teacher sensitivity	4.85 (1.03)	4.62 (0.94)	5.08 (1.06)
Class negative climate	1.99 (0.97)	2.13 (1.08)	1.86 (0.82)
Classroom overall quality	4.72 (0.78)	4.47 (0.72)	4.97 (0.77)
Class size	16.47 (2.59)	16.68 (2.54)	16.25 (2.62)
No. adults in classroom	2.41 (0.69)	2.55 (0.79)	2.27 (0.54)
	Site characteristics		
Family support worker on staff	1.20 (2.35)	0.39 (0.49)	2.03 (3.09)
No. children ages 3–5	111.96 (115.62)	96.18 (48.57)	128.03 (155.30)
Proportion of African Americans	0.71 (0.38)	0.68 (0.38)	0.74 (0.37)
Proportion of teachers with BA	0.43 (0.39)	0.50 (0.36)	0.36 (0.41)
Proportion of TAs with college	0.49 (0.37)	0.36 (0.32)	0.62 (0.36)
Proportion of single families	0.86 (0.16)	0.84 (0.15)	0.88 (0.15)
Proportion of families employed	0.74 (0.26)	0.81 (0.22)	0.67 (0.29)
Proportion of families receiving TANF	0.31 (0.34)	0.24 (0.29)	0.39 (0.37)

Note. Parenthetical data are standard deviations. CSRP = Chicago School Readiness Project; BPI = Behavior Problems Index; C–TRF = Caregiver–Teacher Report Form; PIPPS = Penn Interactive Peer Play Scale; BA= Bachelor's Degree; TAs = Teacher Aides; TANF = Temporary Assistance for Needy Families.
[a] The descriptive statistics within the PIPPS (n = 181) sample are similar to those in the full sample. [b] Poverty-related risk is an aggregate comprising mother-reported income, education, and employment.

children in the intervention group were reported as having significantly fewer internalizing and externalizing behavior problems than were their control-group-enrolled counterparts by spring). In particular, compared with children in the control group, children in the intervention group on average had significantly lower scores on the BPI Internalizing scale (−1.81 points in Model 1, $t = -4.20$, $p < .01$, $d = -0.89$) and the BPI Externalizing scale (−2.92 points, $t = -3.17$, $p < .01$, $d = -0.64$) scores. As shown in Table 3, after inclusion of the interactions between intervention and child race/ethnicity, gender, and family socioeconomic risks (Model 2), children in the intervention group had even lower scores on the BPI Internalizing scale (−2.15 points, $t = -4.00$, $p < .01$, $d = -1.06$) and the BPI Externalizing scale (−4.50 points, $t = -4.21$, $p < .01$, $d = -0.98$).

Table 3 also shows several important associations between a number of the Level 1 and Level 2 covariates and teachers' reports of children's BPI scores in the spring. For example, children with higher levels of behavioral risk in the fall of the Head Start year

Table 2
Intraclass Correlations (ICCs) in CSRP Control Group

Measure	ICC in fall Head Start			ICC in spring Head Start		
	Child level	Class level	Site level	Child level	Class level	Site level
BPI Internalizing	0.74	0.16	0.10	0.78	0.11	0.11
BPI Externalizing	0.68	0.13	0.20	0.68	0.25	0.07
C–TRF Internalizing				0.57	0.20	0.23
C–TRF Externalizing				0.74	0.14	0.12
PIPPS Disruption				0.99	0.00	0.01
PIPPS Disconnection				0.89	0.04	0.07

Note. ICCs were estimated in the control group using unconditioinal three-level hierarchical linear modeling. BPI = Behavior Problems Index; C–TRF = Caregiver–Teacher Report Form; PIPPS = Penn Interactive Peer Play Scale.

were reported to continue to show significantly higher BPI scores in the spring (1.25 points, $t = 6.10$, $p < .01$, $d = 0.62$ on the Internalizing scale; 3.48 points, $t = 7.70$, $p < .01$, $d = 0.75$ on the Externalizing scale). Conversely, children whose parents spoke Spanish were reported by teachers to have lower levels of internalizing problems (-0.59 points, $t = -2.02$, $p < .05$, $d = -0.29$). As shown in Table 3, other child and family demographic characteristics were not significantly related ($p > .05$) to child BPI scores in spring.

Among the Level 2 covariates as predictors of children's behavior problems, teacher's ratings of job demand and of job control were both positively associated with child BPI Internalizing and Externalizing scores (see Table 3). In addition, teacher's K6 scores were weakly associated with children's lower behavior problem scores. In regard to Level 2 classroom-level covariates, teacher's behavior management and negative climate in fall were negatively related to BPI scores across both models (e.g., higher behavior management in fall predicted a 1.24-point decrease in spring BPI Internalizing scores, $t = -4.91$, $p < .01$, $d = -0.61$, and a 2.64-point decrease in BPI Externalizing scores, $t = -5.36$, $p < .01$, $d = -0.54$). Teacher's sensitivity and negative climate were associated with BPI scores in unexpected directions (see Table 3). Other teacher and class covariates were not associated with BPI scores.

In regard to the covariates at the site level, children who were enrolled at sites with more family support workers on staff and a higher percentage of teacher assistants with college degrees on average tended to be rated as having fewer behavior problems (see Table 3). In contrast, children who were enrolled in larger sites or in sites with higher percentages of African American children, who had teachers with BAs, or who came from families with at least one parent employed tended to be rated by teachers as having a higher number of behavior problems (see Table 3). Other site-level covariates did not show significant associations with child BPI scores.

All analyses were repeated with teacher-reported C–TRF scores (assessed in spring) as the dependent measure. These additional models suggest that children in the intervention group were reported as having significantly fewer internalizing and externalizing behavior problems on the C–TRF scales than were their control-group-enrolled counterparts, net of children's initial levels of behavioral risk (see Table 3). Again, effect sizes of intervention impact were larger when interactions between intervention and child race/ethnicity, gender, and family socioeconomic risk were included in Model 2. For example, children in the intervention group had lower C–TRF Internalizing scores (-3.96 points, $t = -2.66$, $p < .05$, $d = -0.82$) and lower C–TRF Externalizing scores (-8.77 points, $t = 3.98$, $p < .01$, $d = -0.92$) than did children in the control group. Inspection of results suggests the same pattern and magnitude of associations between covariates at Levels 1, 2, and 3 and C–TRF scores as were found when BPI scores were considered as the dependent variable (see Table 3).

Because of the known difficulty in detecting the effects of moderators, we then graphed point estimates and confidence intervals for all subgroups for both the BPI and the C–TRF teacher-reported data to examine whether the CSRP intervention demonstrated differential efficacy with some groups of children compared to others. Figure 2 is provided as an illustration and shows that, overall, girls and Hispanic children in the intervention group tended to show larger point estimate reductions in externalizing behavior problems than did their counterparts in the control group. For example, girls in the intervention group were rated 3.88 points lower in BPI Externalizing scores than were girls in the control group; whereas boys in the intervention group were rated 1.98 points lower than were boys in the control group. Compared with their peers from same racial/ethnic group peers who were in the control group, Hispanic children in the intervention group tended to score 3.45 points lower in BPI Externalizing scores than did their Hispanic counterparts in the control group, whereas Hispanic Black children in the intervention group scored 1.81 points lower than did non-Hispanic Black children in the control group. What these figures illustrated for us, however, is that no one group "carried" the effect of the CSRP intervention on children's behavior problems (as indexed by teacher report). In addition, all children, regardless of family socioeconomic risks, tended to benefit from CSRP intervention.

Results from additional post hoc sensitivity analyses conducted with the BPI and C–TRF teacher reports as dependent variables (Model 3) revealed that intervention effects were no longer statistically significant when binomial codes for pairwise matches between sites were substituted for program characteristics as Level 3 covariates (table available from C. Cybele Raver upon request). It should be noted that the same site-level covariates that were included in Model 2 were employed to determine the matches.

Finally, with regard to independent observational assessments of children's aggressive/disruptive and withdrawn/disconnected behaviors, the best fitting models (Model 2) were rerun for the stratified, randomly selected subsample. As indicated in Table 3, there were statistically significant effects for intervention on PIPPS assessments of children's externalizing/disruptive behavior (-1.40 points, $t = -2.09$, $p = .06$, $d = 0.77$). In addition, there was a statistically significant intervention by poverty-related risk interaction (0.92 points, $t = 3.36$, $p < .01$, $d = 0.51$. Children with no or one risk tended to benefit from CSRP intervention, whereas the intervention effect for children with two or more risks was not statistically significant. These analyses suggest converging evidence for the role of child, teacher/classroom, and site character-

Table 3
Parameter Coefficients (and Standard Errors) of the Effects of CSRP on BPI, C–TRF, and PIPPS

	BPI		C–TRF		PIPPS	
Variable	Internalizing	Externalizing	Internalizing	Externalizing	Externalizing/ disruptive	Internalizing/ disconnected
Treatment	−2.15** (0.54)	−4.50** (1.07)	−3.96* (1.49)	−8.77** (2.20)	−1.40† (0.67)	−0.41 (0.59)
			Child and family characteristics			
Boy	−0.31 (0.23)	0.39 (0.73)	−0.25 (0.60)	0.61 (1.63)	−0.59 (0.41)	0.12 (0.36)
Treatment × Boy	0.33 (0.32)	2.30* (1.03)	0.89 (0.84)	5.18* (2.28)	0.67 (0.56)	0.14 (0.49)
Black	−0.08 (0.47)	−0.49 (1.09)	−0.06 (1.18)	−0.96 (2.14)	0.87 (0.63)	0.49 (0.56)
Treatment × Black	0.82 (0.55)	1.72 (1.24)	1.76 (1.46)	4.01 (2.39)	−0.56 (0.71)	−0.81 (0.63)
Other race	0.05 (0.52)	0.34 (1.70)	0.56 (1.30)	1.64 (3.36)	0.24 (0.93)	−0.20 (0.81)
Treatment × Other Race	0.64 (0.73)	1.90 (2.27)	0.83 (1.88)	3.89 (4.51)	−0.62 (1.24)	0.12 (1.08)
Family poverty risk[a]	0.17 (0.13)	0.18 (0.27)	0.08 (0.37)	0.60 (0.57)	−0.37† (0.20)	0.04 (0.17)
Treatment × Poverty Risk[a]	−0.12 (0.18)	−0.26 (0.38)	0.14 (0.53)	−0.69 (0.80)	0.92** (0.27)	0.02 (0.23)
Fall total BPI risk	1.25** (0.21)	3.48** (0.45)	2.66** (0.47)	7.57** (0.96)	0.76* (0.32)	0.79** (0.27)
Single parent	0.21 (0.18)	0.60 (0.40)	0.55 (0.41)	1.62† (0.84)		
4+ children in household	0.17 (0.19)	−0.35 (0.41)	0.09 (0.43)	−1.17 (0.86)		
Spanish speaking	−0.59* (0.29)	−1.04 (0.64)	−1.13† (0.66)	−2.06 (1.35)		
			Teacher and class characteristics			
Teacher BA	0.01 (0.29)	0.77 (0.55)	0.35 (0.65)	2.05† (1.16)	−0.17 (0.38)	−0.16 (0.32)
Teacher age	−0.01 (0.01)	0.00 (0.03)	−0.01 (0.03)	−0.04 (0.06)	−0.04* (0.02)	−0.00 (0.02)
Teacher K6 score	−0.17 (0.09)	−0.15 (0.17)	−0.21 (0.20)	−0.30 (0.35)	0.01 (0.09)	−0.05 (0.08)
Teacher job demand	0.85** (0.27)	1.95** (0.53)	1.39* (0.61)	3.55** (1.10)	0.50 (0.33)	0.10 (0.28)
Teacher job control	0.66** (0.18)	0.94* (0.35)	1.45** (0.41)	2.07** (0.73)	−0.00 (0.23)	0.02 (0.19)
Behavior management	−1.24** (0.25)	−2.64** (0.49)	−2.40** (0.57)	−5.34** (1.04)		
Teacher sensitivity	0.70* (0.32)	2.57** (0.63)	1.80* (0.74)	4.95** (1.32)		
Negative climate	−0.79** (0.21)	0.07 (0.41)	−1.45** (0.48)	−0.52 (0.86)	−0.15 (0.28)	0.42† (0.24)
ECERS–R	−0.03 (0.27)	0.46 (0.53)	−0.25 (0.61)	0.39 (1.10)	−0.38 (0.32)	−0.22 (0.27)
Class size	−0.29** (0.07)	−0.73** (0.13)	−1.01** (0.16)	−1.50** (0.28)	−0.09 (0.09)	0.05 (0.08)
No. adults in class	0.15 (0.22)	0.62 (0.42)	0.40 (0.51)	1.38 (0.86)	0.28 (0.27)	−0.32 (0.24)
			Site characteristics			
Family support worker	−0.89** (0.22)	−1.62** (0.44)	−1.98** (0.52)	−3.49** (0.92)	−0.36 (0.24)	−0.17 (0.21)
No. children ages 3–5	0.02** (0.00)	0.03** (0.01)	0.03** (0.01)	0.06** (0.01)	0.01+ (0.00)	0.00 (0.00)
% African Americans	1.00 (0.71)	1.23 (1.48)	1.24 (1.61)	2.66 (3.02)	1.12 (0.80)	0.07 (0.69)
% teachers with BA	1.75* (0.58)	2.47† (1.13)	4.13* (1.35)	3.85 (2.36)	0.79 (0.76)	0.89 (0.69)
% TAs with college	−2.37** (0.47)	−3.94** (0.91)	−4.12** (1.07)	−5.98* (1.90)	0.86 (0.52)	−1.16* (0.45)
% single families	0.16 (1.02)	1.93 (2.05)	2.48 (2.48)	5.34 (4.09)		
% families employed	1.99† (0.98)	4.35* (1.87)	3.56 (2.24)	7.79† (3.88)	1.74 (1.18)	2.07† (1.02)
% families using TANF	0.29 (0.48)	1.63 (0.92)	0.88 (1.10)	3.99† (1.93)		
Constant	4.78† (2.30)	−1.01 (4.53)	13.33* (5.36)	1.84 (9.31)	−0.36 (2.41)	−1.64 (2.10)

Note. $df = 416$ for Level 1, $df = 23$ for Level 2, and $df = 8$ for Level 3 (with $df = 16$ for the interactions between treatment and Level 1 covariates) in the models presented above. CSRP = Chicago School Readiness Project; BPI = Behavior Problems Index; C–TRF = Caregiver–Teacher Report Form; PIPPS = Penn Interactive Peer Play Scale; BA= Bachelor's Degree; TAs = Teacher Aides; TANF = Temporary Assistance for Needy Families.
[a] Poverty-related risk is an aggregate comprising mother-reported income, education, and employment.
† $p < .10$. * $p < .05$. ** $p < .01$.

istics as significant predictors of observed externalizing behaviors (see Table 3).

Discussion

Children exposed to the multiple family and neighborhood stressors associated with poverty are at substantially increased risk for behavioral difficulty, and Head Start teachers report children's behavioral dysregulation and disruptive, aggressive behaviors as being among their top concerns (Cai, Kaiser, & Hancock, 2004; Morales & Guerra, 2006). In response to this pressing policy concern, the Chicago School Readiness Project was designed to support children's emotional and behavioral regulation and to reduce their risk of behavioral difficulty. The CSRP targeted classroom processes in low-income, urban, community-based preschools in answer to recent calls to focus not only on low-income children themselves but on settings that might alternately ameliorate or exacerbate children's behavioral risk (Atkins, Graczyk, Frazier, & Abdul-Adil, 2003; Cai et al., 2004).

Results from our analyses suggest that CSRP had a large, statistically significant impact on reducing low-income preschoolers' internalizing and externalizing behavior problems. On aver-

age, children enrolled in CSRP classrooms were reported by teachers to manifest significantly fewer signs of sadness and withdrawal than were children in the control group. Effect sizes ranged from $d = 0.89$ for teacher-reported BPI scores to $d = 0.62$ for teacher-reported behavior problems on the C–TRF. This effect size would translate to a reduction of one to two internalizing behavior problems for the intervention group as compared with the control group on the BPI, for example. The CSRP model of intervention also demonstrated efficacy in reducing preschoolers' externalizing behavior problems, including children's symptoms of aggression and defiance, with $d = 0.64$ and $d = 0.53$ on the BPI and C–TRF, respectively. This effect size would translate to a reduction of two to three externalizing behavior problems for children in the intervention group as compared with their control group counterparts (see Figure 2 for illustration of effect sizes plotted as point estimates on teacher-reported BPI Externalizing scores).

Results from our independently assessed classroom observations (using an observer rating version of the PIPPS) provided important converging evidence of the impact of this intervention on children's externalizing behavioral outcomes. Results from our observational data suggested that fewer instances of externalizing, disruptive behavior (e.g., physical and verbal aggression) were observed among children in intervention-assigned classrooms as compared with their counterparts in control-enrolled classrooms; findings were strongest for children facing lower levels of poverty-related risk. Observations of children's internalizing, disconnected behaviors (e.g., wandering aimlessly, withdrawing from play) were in the hypothesized direction but were not sufficiently large to reach levels of statistical significance.

In short, results indicate that our classroom-based intervention offers a highly promising model for supporting the emotional and behavioral development of low-income preschool children exposed to a large number of poverty-related risks. Our results are consistent with emerging research on ways to build on the strengths of low-income families and the institutions that serve them (Berryhill & Prinz, 2003; Dishion & Stormshak, 2007; Tolan et al., 2004). Our findings provide powerful evidence that a key source of institutional support of children's communities, namely, their Head Start–funded preschool programs, can be leveraged to provide important behavioral as well as academic intervention and support.

How do our findings stack up to previously published results of classroom-based intervention? Our findings are consistent with other recent studies on the efficacy of teacher-training programs with older children. Systematic review of the literature revealed few cluster-based randomized trials that combined extensive teacher-training and mental health consultation models for children in this age range (for exceptions, see Dumas, Prinz, Smith, & Laughlin, 1999). That said, CSRP drew from the strengths of previous models emphasizing the importance of providing significant adults in children's lives with the knowledge, skills, and support to effectively support children's self-regulation and reduce children's behavior problems. As with those programs, CSRP placed central importance on intervention staff serving as coaches, to aid teachers in building new relationships with students through use of more adaptive strategies of management and engagement (Donohue et al., 2000; Gorman-Smith et al., 2003). In sum, our results contribute to a growing literature in prevention research that suggests ways teacher training and mental health consultation efforts can be extended "downward" to settings in which an increasingly large fraction of preschool children are served.

Did the CSRP intervention work better for some children than for others? Significant Intervention × Gender and Intervention × Race/Ethnic Group Membership interactions for teachers' reports suggest that, compared with other children in the sample, Hispanic girls showed the largest reductions in behavior problems. Although moderating analyses of the full sample did not yield statistically significant evidence of the moderating role of poverty-related risk, evidence of moderation by poverty-related risk was found for the observational subsample. Following McClelland and Judd (1993), we took additional steps to guard against the risk of Type II error (i.e., concluding null differences when the problem may lie with inadequate statistical power to detect differences in intervention impact for different groups). To guard against that risk, we conducted follow-up analyses of intervention impacts within each subgroup. Follow-up subgroup analyses suggest, when children in the intervention group were compared with their racial/ethnic- and gender-matched control-group peers, the intervention led to significant, albeit smaller, reductions in boys' and African American children's behavior problems as well.

Placing Intervention in the Social Context of Classrooms

When intervention is considered from a classroom-based perspective, results of this study highlight recent observations by Anthony et al. (2005) that classroom teachers vary widely in their reports of the prevalence of students' behavior problems. Findings by Anthony et al. (2005) suggest that some teachers report exceptionally few children as exhibiting difficulty and that other teachers report up to 60% of their children as showing serious behavioral difficulty. To disentangle the sources of this tremendous variability, we included key child-level factors, such as children's initial level of behavior problems (collected at baseline) and children's exposure to high versus low levels of poverty-related risk, when estimating the impact of the intervention on children's adjustment. Our models also took into account teacher and classroom characteristics that might covary with teachers' perceptions of children's behavioral difficulty. For example, our results suggest that teachers' own feelings of job overload were significantly predictive of teachers' reports of higher behavior problems for the children in their classroom. Site-level characteristics, such as whether programs could afford to have a full-time family support staff member on staff, also were related to teachers' reports of whether children's behavior problems worsened or improved over the course of the school year. Our results suggest that it is very important to comprehensively attend to teacher and classroom dynamics in our analyses of classroom-based intervention impact.

Limitations of the Current Study

Although the present study had several strengths, this study's conclusions are constrained by several limitations. For example, each of the measurement approaches used in this study has its strengths and drawbacks. On one hand, teacher reports offer a relatively low cost and ecologically valid means of assessing children's emotional and behavioral dysregulation that is offset by the risk of reporter bias, given teacher knowledge of the program's intervention status. On the other hand, the benefits of indepen-

dently assessed classroom observation are balanced against high cost, low feasibility, and the risk of insufficient power to detect intervention impact, given the "low base rate problem" for young children's observed aggressive and disruptive behaviors (Hughes, White, Sharpen, & Dunn, 2000; Milfort & Greenfield, 2002, p. 593; Miller et al., 2004). With those limitations in mind, we were heartened to find that the observational findings regarding children's externalizing behaviors that used the adapted version of the PIPPS were consistent with (though not confirmatory of) our teacher report findings.

A second key limitation is that the generalizability of our findings of CSRP's efficacy is constrained in a number of ways. For example, our findings show CSRP's efficacy in reducing children's behavioral difficulty only through the end of the preschool year. Analyses are currently under way to detect whether CSRP-enrolled children sustain these improvements as they make transitions to new kindergarten classrooms the following fall. Additionally, we faced significant analytic trade-offs while meeting our goal of testing the efficacy of an intervention in an early phase of development, implemented in a single set of "real-world" community settings. For example, we chose to include a large number of covariates in Models 1 and 2 to take into account the specific characteristics of the participants and places in our study. But this choice significantly limits our ability to generalize our findings to other places and to other groups of children. This limitation is offset by the opportunity to review estimates of program impact when our models were more broadly specified (e.g., Model 3).[3] With these caveats in mind, our findings are in keeping with previous clinical research that used children's behavior problems as indicators of intervention efficacy, in which short-term improvement in children's behavioral adjustment is an important outcome in its own right (see Berryhill & Prinz, 2003, for review).

A third limitation is that we cannot "unpack" the benefits of CSRP intervention to detect which program components may have made the most difference in reducing children's behavior problems. This limitation extends through issues of intervention dosage and fidelity, for which the quality and quantity of service provision were assessed through ratings by MHCs and teachers but not through independent ratings. Propensity score matching and other sophisticated statistical approaches offer significant promise in helping us to evaluate which types of intervention support may have offered greatest benefit to classrooms and children. In the meantime, our study represents an important preliminary empirical step, in that it bridges previous research traditions focusing on elementary school–based RCT trials, on the one hand, and programs and nonrandomized preschool mental health consultation studies, on the other.

Clinical and Research Implications

Previous research in child development and psychopathology has sounded a recurrent, clear alarm regarding the prevalence, persistence, and severity of behavior problems among young children facing conditions of income poverty and inequality (Campbell, 1995; Shaw et al., 2006). The children living in CSRP's low-income neighborhoods face many of these risks: Many of the families in our study reported living on incomes substantially below the federal poverty threshold and trying to "make ends meet" with lean financial resources. The children enrolled in CSRP were reported and observed to demonstrate a range of behavioral difficulties in fall and spring of their preschool year. These difficulties indicated substantial reason for concern for the emotional and behavioral adjustment of young children early on in their educational trajectories.

In the context of these multiple risks, it is important to highlight the socioemotional competence and capacity for behavioral improvement of low-income children: The CSRP offers a highly promising approach for reducing children's risk of behavioral difficulty in classroom settings. Recent innovative studies in prevention science have built on family and community strengths to support low-income children's emotional and behavioral adjustment in urban communities such as Baltimore, Philadelphia, Seattle, and Chicago (Tolan, Gorman-Smith, & Henry, 2003; Van Zeijl et al., 2006). Our findings suggest that children's preschools represent a key component in building a comprehensive system of prevention and early intervention.

[3] The inclusion of a large number of covariates in Model 1 essentially suggests that the variance around treatment versus control group intercepts can be "fixed" or apportioned to a set of covariates at the child, classroom, and site levels (see Raudenbush & Bryk, 2002). Alternately, Model 3 represents a more generalizable model using a "random effects" approach, in which pairwise dummy variables are expected to account for observed and unobserved heterogeneity at the site level. Difference in our results between these two model specifications may be due to (a) limited statistical power, (b) ways that program pairs may not adequately capture the observed and unobserved heterogeneity in our sample, and to (c) Model 3's sensitivity to outliers.

References

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

Achenbach, T. M., Edelbrock, C., & Howell, C. T. (1987). Empirically based assessment of the behavioral/emotional problems of 2- and 3-year-old children. *Journal of Abnormal Child Psychology, 15,* 629–650.

Achenbach, T. M., & Rescorla, L. A. (2001). *Manual for the ASEBA school-age forms & profiles.* Burlington: University of Vermont Research Center for Children, Youth, & Families.

Anthony, B., Anthony, L. G., Morrel, T. M., & Acosta, M. (2005). Evidence for social and behavioral problems in low-income, urban preschoolers: Effects of site, classroom, and teacher. *Journal of Youth and Adolescence, 34,* 31–39.

Arnold, D. H., Brown, S. A., Meagher, S., Baker, C. N., Dobbs, J., & Doctoroff, G. L. (2006). Preschool-based programs for externalizing problems. *Education and Treatment of Children, 29,* 311–339.

Arnold, D. H., McWilliams, L., & Arnold, E. H. (1998). Teacher discipline and child misbehavior in day care: Untangling causality with correlation data. *Developmental Psychology, 34,* 276–287.

Atkins, M. S., Graczyk, P. A., Frazier, S. L., & Abdul-Adil, J. (2003). Toward a new model for promoting urban children's mental health: Accessible, effective, and sustainable school-based mental health services. *School Psychology Review, 32,* 503–514.

August, G. J., Egan, E. A., Realmuto, G. M., & Hektner, J. M. (2003). Parceling component effects of a multifaceted prevention program for disruptive elementary school children. *Journal of Abnormal Child Psychology, 31,* 515–527.

August, G. J., Realmuto, G. M., Hektner, J. M., & Bloomquist, M. L. (2001). An integrated components preventive intervention for aggressive elementary school children: The "Early Risers" program. *Journal of Consulting and Clinical Psychology, 69,* 614–626.

Berryhill, J. C., & Prinz, R. J. (2003). Environmental interventions to enhance student adjustment: Implications for prevention. *Prevention Science, 4,* 65–87.

Bloom, H. S. (2005). *Learning more from social experiments: Evolving analytic approaches.* New York: Russell Sage Foundation.

Brouwers, A., & Tomic, W. (2000). A longitudinal study of teacher burnout and perceived self-efficacy in classroom management. *Teaching and Teacher Education, 16,* 239–253.

Cai, X., Kaiser, A. P., & Hancock, T. B. (2004). Parent and teacher agreement on Child Behavior Checklist items in a sample of preschoolers from low-income and predominately African American families. *Journal of Child and Adolescent Psychology, 33,* 303–312.

Campbell, S. B. (1995). Behavior problems in preschool children: A review of recent research. *Journal of Child Psychology and Psychiatry, 36,* 113–149.

Carter, A. S., Briggs-Gowan, M. J., Jones, S. M., & Little, T. D. (2003). The Infant–Toddler Social and Emotional Assessment (ITSEA): Factor structure, reliability, and validity. *Journal of Abnormal Child Psychology, 31,* 495–514.

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.

Cook, T. D. (2005). Emergent principles for the design, implementation and analysis of cluster-based experiments in social science. *Annals of the American Academy of Political and Social Sciences, 599,* 176–198.

Curbow, B. (1990). Job stress in child care workers: A framework for research. *Child & Youth Care Quarterly, 19,* 215–231.

Curbow, B., Spratt, K., Ungaretti, A., McDonnell, K., & Breckler, S. (2000). Development of the child care worker job stress inventory. *Early Childhood Research Quarterly, 15,* 515–536.

Dishion, T. J., & Stormshak, E. A. (2007). *Intervening in children's lives: An ecological, family-centered approach to mental health care.* Washington, DC: American Psychological Association.

Donohue, P., Falk, B., & Provet, A. G. (2000). *Mental health consultation in early childhood.* Baltimore: Brookes.

Dumas, J. E., Prinz, R. J., Smith, E. P., & Laughlin, J. (1999). The EARLY ALLIANCE prevention trial: An integrated set of interventions to promote competence and reduce risk for conduct disorder, substance abuse, and school failure. *Clinical Child and Family Psychology Review, 2,* 37–53.

Fantuzzo, J., Stoltzfus, J., Lutz, M. N., Hamlet, H., Balraj, V., Turner, C., & Mosca, S. (1999). An evaluation of the special needs referral process for low-income preschool children with emotional and behavioral problems. *Early Childhood Research Quarterly, 14,* 465–482.

Fantuzzo, J. W., Sutton-Smith, B., Coolahan, K. C., Manz, P. H., Canning, S., & Debnam, D. (1995). Assessment of preschool play interaction behaviors in young low-income children: Penn Interactive Peer Play Scale. *Early Childhood Research Quarterly, 10,* 105–120.

Flay, B. R., Biglan, A., Boruch, R. F., Castro, F. G., Gottfredson, D., Kellam, S., et al. (2005). Standards of evidence: Criteria for efficacy, effectiveness, and dissemination. *Prevention Science, 3,* 151–175.

Gerard, J. M., & Buehler, C. (2004). Cumulative environmental risk and youth problem behavior. *Journal of Marriage and Family, 66,* 702–720.

Gilliam, W. S. (2005). *Prekindergarteners left behind: Expulsion rates in state prekindergarten programs (FCD Policy Brief Series No. 3).* New York: Foundation for Child Development.

Goerge, R. M., Dilts, J., Yang, D., Wasserman, M., & Clary, A. (2007). *Chicago children and youth 1990–2010: Changing population trends and their implications for services.* Chicago: University of Chicago, Chapin Hall Center for Children.

Gorman-Smith, D., Beidel, D., Brown, T. A., Lochman, J., & Haaga, A. F. (2003). Effects of teacher training and consultation on teacher behavior towards students at high risk for aggression. *Behavior Therapy, 34,* 437–452.

Gottfredson, G. D., Jones, E. M., & Gore, T. W. (2002). Implementation and evaluation of a cognitive–behavioral intervention to prevent problem behavior in a disorganized school. *Prevention Science, 3,* 43–56.

Granger, R. C., & Marx, E. (1992). The policy implications of job satisfaction ratings for recruiting and retaining early childhood teachers. *Child and Youth Care Forum, 21,* 229–249.

Gross, D., Fogg, L., Webster-Stratton, C., Garvey, C., Julion, W., & Grady, J. (2003). Parent training with multi-ethnic families of toddlers in day care in low-income urban neighborhoods. *Journal of Consulting and Clinical Psychology, 71,* 261–278.

Harms, T., Clifford, R. M., & Cryer, D. (2003). *Early Childhood Environment Rating Scale* (Rev. ed.). New York: Teachers College Press.

Hughes, C., White, A., Sharpen, J., & Dunn, J. (2000). Antisocial, angry, and unsympathetic: "Hard-to-manage" preschoolers' peer problems and possible cognitive influences. *Journal of Child Psychology and Psychiatry, 41,* 169–179.

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

Innes, F. K., Denton, K. L., & West, J. (2001, April). *Child care factors and kindergarten outcomes: Findings from a national study of children.* Paper presented at the biennial meeting of the Society for Research in Child Development, Minneapolis, MN.

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

Keenan, K., Wakschlag, L. S., Danis, B., Hill, C., Humphries, M., Duax, J., & Donald, R. (2007). Further evidence of the reliability and validity of DSM–IV ODD and CD in preschool children. *Journal of the American Academy of Child and Adolescent Psychiatry, 46,* 457–468.

Kellam, S. G., Ling, S., Merisca, R., Brown, C. H., & Ialongo, N. (1998). The effect of the level of aggression in the first grade classroom on the course and malleability of aggressive behavior into middle school. *Development and Psychopathology, 10,* 165–185.

Kessler, R. C., Andrews, G., Colpe, L. J., Hiripi, E., Mroczek, D. K., Normand, S.-L. T., et al. (2002). Short screening scales to monitor population prevalence and trends in nonspecific psychological distress. *Psychological Medicine, 32,* 959–976.

Knight, G. P., & Hill, N. E. (1998). Measurement equivalence in research involving minority adolescents. In V. C. McLoyd & L. Steinberg (Eds.), *Studying minority adolescents: Conceptual, methodological, and theoretical issues* (pp. 183–210). Mahwah, NJ: Erlbaum.

La Paro, K., Pianta, R., & Stuhlman, M. (2004). Classroom Assessment Scoring System (CLASS): Findings from the prekindergarten year. *Elementary School Journal, 104,* 409–426.

Li-Grining, C., Madison-Boyd, S., Jones-Lewis, D., Smallwood, K. M., Sardin, L., Metzger, M. W., et al. (2007). *Implementing a classroom-based intervention in the "real world": The role of teachers' psychosocial stressors.* Paper presented at the biennial meeting of the Society for Research in Child Development, Boston, MA.

Lochman, J. E., & Wells, K. C. (2003). Effectiveness of the Coping Power program and of classroom intervention with aggressive children: Outcomes at a 1-year follow-up. *Behavior Therapy, 34,* 493–515.

Madison-Boyd, S., Raver, C. C., Aufmuth, E., Barden, K., Jones-Lewis, D., & Williams, M. (2006). *The Chicago School Readiness Project Mental Health Consultation Manual.* Unpublished manual.

Magnuson, K. A., Ruhm, C., & Waldfogel, J. (2007). The persistence of preschool effects: Do subsequent classroom experiences matter? *Early Childhood Research Quarterly, 22,* 18–38.

Margolin, G., & Gordis, E. B. (2000). The effects of family violence on children. *Annual Review of Psychology, 51,* 445–479.

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

Milfort, R., & Greenfield, D. B. (2002). Teacher and observer ratings of Head Start children's social skills. *Early Childhood Research Quarterly, 17,* 581–595.

Miller, A. L., Gouley, K. K., Seifer, R., Dickstein, S., & Shields, A. (2004). Emotions and behaviors in the Head Start classroom: Associations among observed dysregulation, social competence, and preschool adjustment. *Early Education and Development, 15,* 147–165.

Morales, J. R., & Guerra, N. G. (2006). Effects of multiple context and cumulative stress on urban children's adjustment in elementary school. *Child Development, 77,* 907–923.

NICHD Early Child Care Research Network. (1999). Child outcomes when child care center classes meet recommended standards for quality. *American Journal of Public Health, 89,* 1072–1077.

Randolph, S. M., Koblinsky, S. A., & Roberts, D. D. (1996). Studying the role of family and school in the development of African American preschoolers in violent neighborhoods. *Journal of Negro Education, 65,* 282–294.

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

Raver, C. C. (2002). Emotions matter: Making the case for the role of young children's emotional development for early school readiness. *Social Policy Report, 16,* 3–6.

Raver, C. C. (2004). Placing emotional self-regulation in sociocultural and socioeconomic contexts. *Child Development, 75,* 346–353.

Raver, C. C., Jones, S. M., Li-Grining, C. P., Metzger, M., Champion, K. M., & Sardin, L. (2008). Improving preschool classroom processes: Preliminary findings from a randomized trial implemented in Head Start settings. *Early Childhood Research Quarterly, 23,* 10–26.

Ritchie, S., & Howes, C. (2003). Program practices, caregiver stability, and child–caregiver relationships. *Journal of Applied Developmental Psychology, 24,* 497–516.

Shaw, D. S., Dishion, T. J., Supplee, L., Gardner, F., & Arnds, K. (2006). Randomized trial of a family-centered approach to the prevention of early conduct problems: Two-year effects of the family checkup in early childhood. *Journal of Consulting and Clinical Psychology, 74,* 1–9.

Spoth, R. L., Kavanagh, K. A., & Dishion, T. J. (2002). Family-centered preventive intervention science: Toward benefits to larger populations of children, youth, and families. *Prevention Science, 3,* 145–152.

Tolan, P. H., Gorman-Smith, D., & Henry, D. B. (2003). The developmental ecology of urban males' youth violence. *Developmental Psychology, 39,* 274–291.

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

Van Zeijl, J., Mesman, J., Van Ijzendoorn, M. H., Bakermans-Kranenburg, M. J., Juffer, F., Stolk, M. N., et al. (2006). Attachment-based intervention for enhancing sensitive discipline in mothers of 1- to 3-year-old children at risk for externalizing behavior problems: A randomized controlled trial. *Journal of Consulting and Clinical Psychology, 74,* 994–1005.

Wasik, B. A., Bond, M. A., & Hindman, A. (2006). The effects of language and literacy intervention on Head Start children and teachers. *Journal of Educational Psychology, 98,* 63–74.

Webster-Stratton, C., Reid, M. J., & Hammond, M. (2004). Treating children with early-onset conduct problems: Intervention outcomes for parent, child, and teacher training. *Journal of Clinical Child and Adolescent Psychology, 33,* 105–124.

Woolfolk, A. E., Rosoff, B., & Hoy, W. K. (1990). Teachers' sense of efficacy and their beliefs about managing their students. *Teaching and Teacher Education, 6,* 137–148.

Yoshikawa, H., & Knitzer, J. (1997). *Lessons from the field: Head Start mental health strategies to meet changing needs.* New York: National Center for Children in Poverty.

Zill, N. (1990). *The Behavior Problems Index* [descriptive material]. Washington, DC: Child Trends.

Received February 29, 2008
Revision received December 29, 2008
Accepted January 12, 2009 ■