CONFERENCE DRAFT, NOT FOR CITATION OR DISTRIBUTION

Running Head: REDUCING BEHAVIOR PROBLEMS IN HEAD START

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

C. Cybele Raver
New York University
Stephanie M. Jones
Fordham University
Christine Li-Grining
Loyola University
Fuhua Zhai
New York University
Molly Metzger
Northwestern University
Bonnie Solomon
University of Illinois at Chicago

Prepared for the Population Association of America (PAA) 2008 Annual Meeting
April 17-19, 2008, New Orleans, LA

Contact Information:
C. Cybele Raver
Steinhardt School of Culture, Education, and Human Development
New York University
246 Greene St, Room 519
New York, NY 10003
Phone: (212) 998-5519
Email: cybele.raver@nyu.edu

Fuhua Zhai
Steinhardt School of Culture, Education, and Human Development
New York University
726 Broadway, Room 529
New York, NY 10003
Phone: (212) 998-5209
Email: fuhua.zhai@nyu.edu

Abstract

The present study evaluated the efficacy of a multi-component classroom-based intervention in reducing preschoolers' behavior problems. The Chicago School Readiness Project model was implemented in 35 Head-Start classrooms, using a clustered RCT design. Results indicate significant treatment effects for teacher-reported and independent observations of children's internalizing and externalizing behavior problems, with effect sizes ranging from d =.53 to d =.89. 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.

Key Words: randomized, trial, classroom-based consultation, teacher-training

Recently researchers and policy makers have expressed the concern that preschoolers' behavior problems may significantly compromise their chances for later success in school (Gilliam, 2005; Raver, 2002). Specifically, young children who are persistently sad, withdrawn or disruptive have been found to receive less instruction, to have fewer opportunities for learning from peers, and to be less engaged and less positive about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, Brown, Meagher, Baker, Dobbs, & Doctoroff, 2006; Raver, Garner, & Smith-Donald, 2007). 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 (Dodge, Pettit, & Bates, 1994; Fantuzzo et al., 1999; Farmer, Stangl, Burns, Costello, & Angold, 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? Preschool classrooms are an increasingly important service setting outside the home (Spoth, Cavanagh, & Dishion, 2002), particularly as 67% of young children in the United States are now enrolled in center-based or non-relative care prior to enrollment in Kindergarten (Innes, Denton, & West, 2001; Webster-Stratton & Taylor, 2001). Path-breaking studies in the last decade suggest that targeting classroom processes can be an effective way to reduce children's behavioral problems (August, Realmuto, Hektner, & Bloomquist, 2001; CPPRG, 1999a; Hawkins, Catalano, Kosterman, Abbott, & Hill, 1999; Ialongo Werthamer, Kellam, Brown, Wang, & Lin, 1999; Lochman & Wells, 2003; Reid, Eddy, Fetrow, & Stoolmiller, 1999; Smolkowski, Biglan, Barrera, Taylor,

Black, & Blair, 2005; Webster-Stratton, Reid, & Hammond, 2004). However, most of those classroom-based studies target 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 those same classroom processes hold for urban preschool settings in which younger children are served, where preschool teachers have substantially lower levels of training, support, salary, and education, on average, than do teachers working 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 (see Gilliam, 2005; and Webster-Stratton, Reid, & Stoolmiller, in press for exceptions). Results of a recent nationally-representative sample suggest that while low-income children experienced short-term academic benefit from having attended preschool, their emotional and behavioral adjustment was placed at substantially greater risk in the long-run (Magnuson, Ruhm, & Waldfogel, 2007). This "tradeoff" 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 while also supporting their emotional, behavioral, and academic adjustment.

Moderating role of economic risk, race/ethnicity, and gender

Recent research in prevention science has highlighted the ways that low-income children of color face significantly higher risks of behavioral difficulty while also facing large disparities in their access to mental health and behavioral health services (Epstein et al., 2005; Yoshikawa & Knitzer, 1997; Fantuzzo et al., 1999). A key step in closing the gap in these behavioral health

disparities is that we need to learn whether interventions demonstrate similar or different levels of efficacy for both boys and girls, and for children in different sociocultural and socioeconomic contexts (Knight & Hill, 1998). Extensive research has demonstrated that girls and boys follow different developmental trajectories of behavior problems from toddlerhood to adolescence (Bierman et al., 2004; Broidy et al., 2003; Keenan & Shaw, 1997; Schaeffer et al., 2006). 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; Reid & Webster-Stratton, 2002; Tolan, Gorman-Smith, & Henry, 2004; Van Zeijl et al., 2006). 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 serving the needs of African American children (Goerge, Dilts, Yang, Wasserman, & Clary, 2007). Accordingly, we also examined the moderating role of children's race/ethnic status when examining 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), often with 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; Conduct Problems Prevention Group, 1999; Kellam, Ling, Merisca, Brown, & Ialongo, 1998). Based on this theoretical framework we provided intensive training in strategies 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; Webster-Stratton, Reid, & Hammond, 2001; Webster-Stratton, et al., in press).

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; Morris-Rothschild & Brassard, 2006). 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. For example, teacher burnout has been found to be negatively related to their willingness to try innovative teaching methods (Brouwers & Tomic, 1998; Woolfolk, Rosoff, & Hoy, 1990; Fuchs, Fuchs, & Bishop, 1992). 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 "coach" in supporting teachers while they try new techniques learned in the teacher training (Donohue, Falk, & Provet, 2000; Gorman-Smith, Beidel, Brown, Lochman, & Haaga, 2003). As an additional program component, mental health consultants also spent a significant portion of the school year (in winter) conducting stress reduction workshops to help teachers to reduce stress and limit burnout. One critique might be that MHCs bring both "an

extra pair of hands" to the classroom in addition to their clinical expertise. To 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's 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 (Shaw, Dishion, Supplee, Gardner, & Arnds, 2006; Campbell, 1995). Exposure to family and community violence may represent a particularly pernicious threat to young children's emotional and behavioral adjustment and school readiness (Margolin & Gordis, 2000). In recent research, as many as 30% of inner-city, low-income preschool and early school-aged children were reported to have been exposed to severe violence that included robbery, threats with a weapon, or shooting or stabbing (Bell & Jenkins, 1993; Randolph, Koblinsky, & Roberts, 1996). Based on 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. While Head Start is mandated to serve children with special needs, children with emotional and behavioral problems remain substantially under-referred and underserved for emotional disturbance (Fantuzzo et al., 1999; Yoshikawa & Knitzer, 1997). In short, this model suggests that mental health consultants could aid parents, teachers, and children by providing both classroom-based and individualized consultative services (Briggs-Gowan, Carter, Moye Skuban, & McCue Horwitz, 2001; Keenan et al., in press; Yoshikawa & Knitzer, 1997; Raver & Knitzer, 2002). 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.

The principal aim of the Chicago School Readiness Project intervention was to marshal these primary programmatic components to improve low-income preschool-aged children's school readiness by increasing their emotional and behavioral adjustment. Our long-term research objectives are to examine whether this emotionally- and behaviorally-focused intervention in preschool has a significant long-term impact on children's academic achievement in 1st grade (as measured by standardized tests, school records, and teacher reports). Our immediate research aim addressed in this paper, is to test whether this multi-component intervention yields short-term benefits by reducing children's behavioral problems in the spring of children's preschool year. Following standards set out by Flay et al (2005), multiple methods are employed to assess the impact of the CSRP intervention on children's behavior problems, including teacher reports collected for the full sample and observational reports collected for a stratified, randomly-selected subsample.

Hypotheses

First, we expected the intervention to decrease children's internalizing and externalizing behavior problems by spring of the Head Start school year. We expected the precision of our estimates of the impact of treatment to 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 to 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. Due to the limited research examining race/ethnic differences in

treatment effects in this context, no specific hypotheses are made. Fourth, we examined whether the intervention had a significantly larger impact on girls compared to 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 are considered as "nested" within classroom and institutional contexts. Numerous prevention trials aimed at supporting low-income youth and the institutions that serve them have demonstrated the importance of disaggregating the potential confounds presented by family demographic characteristics by including child-level covariates of race/ethnic status and proxies for poverty (see for example, Aber, Jones & Brown, 2003; August, Egan, Realmuto, & Hektner, 2003; Schaeffer et al., 2006; Tolan, Gorman-Smith, & Henry, 2004). In addition, recent research suggests that interventions may work quite differently in settings marked by low levels of institutional resources and low levels of teacher education and motivation, compared to settings that are well-run, more resourced, and staffed by teachers with lower levels of stress and higher levels of motivation (Gottfredson, Jones, & Gore, 2002).

Building on recent innovations in educational research, we employed a cluster-randomized experimental design to evaluate the efficacy of the CSRP model. We also took a context-specific "settings-level" approach to our analyses by including a large array of baseline characteristics of teachers (at the classroom level) and of programs (at the site level) when modeling treatment impacts across our sample of preschools (see Yoshikawa & Shinn, in press; Kellam et. al., 1998). Recent school-based studies suggest that including a large number of child-, classroom-, and school-level baseline characteristics increases 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 subject selection. In an effort to balance generalizability and feasibility, preschool sites were selected 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 7 high-poverty neighborhoods that were selected on the basis of a set of criteria including high poverty, exposure to high crime, and lower rates of mobility (see Raver et al., 2008 for a detailed discussion of exclusionary criteria). CSRP staff completed block-by-block surveys of all 7 neighborhoods, identifying all child-serving agencies that might potentially provide Head Start-funded preschool services. All identified sites were telephoned to determine if they met the additional site selection criteria (including receipt of Head Start funding, etc.). CSRP staff members then contacted and met with eligible sites to offer them the opportunity to self-nominate for participation in the research project. Eighteen sites across 7 neighborhoods completed the self-nomination process and were included as CSRP sites. Two classrooms within each site were randomly selected for participation, with a research coordinator and research staff successfully able to recruit 83% of the children enrolled in classrooms between Labor Day and the assigned enrollment cut-off 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 aggressive/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 and quality (see below). 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 including the availability of a fulltime family worker at the Head Start site, the size of the program (as indexed by the number of children served, ages 3-5), the proportion of the site identified as African American, the proportion of the teachers with Bachelors' degrees and the proportion of teacher assistants with some college, the proportion of the families served that were single parent families, employed, and reliant on TANF. 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 9 treatment sites, 2 classrooms participated, for a total of 18 treatment classrooms. Across the 9 control sites, there were 17 classrooms (2 classrooms in 8 sites, and 1 classroom in the remaining site which lost one 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 teaching assistants as described above. Additional information on the number of hours teachers spent in training, classrooms received mental health consultation, and teacher assistants spent in control classrooms are reported in a recent paper describing statistically significant classroom-level treatment impacts (Raver et al., 2008; Li-Grining et al., revised and resubmitted).

The CSRP intervention was implemented for two cohorts of children and teachers, with Cohort 1 participating from fall to spring in 2004-05 and Cohort 2 participating from fall to spring in 2005-06. As with other recent efficacy trials implemented with multiple cohorts across time, regions, or racially segregated neighborhoods, 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 (see, for example, Gross et al., 2003).

Because we planned to model child outcomes as potentially responsive to both the intervention and to teacher- and classroom-level characteristics, teachers were also included as research subjects. As with classrooms, teachers were enrolled in two cohorts which were also pooled into a single dataset ($N = 90$). A total of 87 teachers participated in CSRP 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 one child was requested to leave the Head Start program and one parent opted to withdraw her child from participating in CSRP.

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 revealed no significant differences between children who were in the analytic sample and those children who were excluded due to missing data. Similarly, no significant differences were detected between

teachers who were in the analytic sample and those teachers who were omitted due to children's missing data.

Because of the importance of multi-method, multi-measure approaches to estimates of treatment efficacy (Flay et al., 2005), observational assessments of children's problem behaviors were conducted for a stratified, randomly selected subsample of the CSRP sample, where roughly 6 children per classroom were selected, representing 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 (BPI) (described below). Each child's z-score was centered by child gender and classroom membership (i.e., calculated in comparison with other children of the same gender in the same classroom). The *n* for this subsample was 181 (the sample size for analysis was 172), with approximately even groups of males and females (54% female). Moreover, percentages of each race/ethnic group in the observational subsample matched those in the total CSRP sample.

Design

Intervention. As described earlier, The Chicago School Readiness Project intervention model comprises four specific components: (1) teacher training in behavior management strategies, (2) MHCs' provision of "coaching" to teachers in implementing these strategies in the classroom, (3) MHC's provision of stress reduction workshops, and (4) MHC's provision of targeted, direct services for children with the highest emotional and behavioral problems. The intervention components were embedded within a model of service delivery that 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 "1-on-1"

child-focused services were delivered in this sequence by consultants to classrooms randomized to treatment. The full model was manualized, and staff members involved in service delivery underwent two days of training in implementation, as well as clinical supervision (1x every two weeks) and administrative supervision (1x every week) throughout the year.

Treatment Classrooms.

Teacher Training. Treatment teachers were invited to participate in 30 hours of workshop-style training sessions (across five Saturdays in October through January). These trainings, adapted from the evidence-based Incredible Years Teacher Training Program (Webster-Stratton, 1995; 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/hour), catered lunches, and on-site childcare. To address concerns regarding the relevance and acceptability of the training program for ethnic minority teachers of young African American and Hispanic children, the training component was piloted in two demonstration sites not included in the full study. The training and measures were translated into English and Spanish, and piloting and follow-up focus groups were conducted with 12 teachers in the two community organizations serving low-income African American and low-income Hispanic families. Teachers' feedback 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. Past

intervention studies have found that training alone may not ensure teachers' transfer of learned material to their classrooms, given the everyday challenges of running a classroom smoothly (Atkins et al., 2006; Gorman-Smith et al., 2003; Gross et al., 2003; Jones et al., 2006; Wasik, Bond, & Hindman, 2006). Therefore, during the first 10 weeks (or the first third) of the intervention, MHCs' classroom visits focused on following a set of 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, 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 one-day stress reduction workshop for each Head Start site. MHCs also began devoting time during classroom visits to discuss teachers' experiences of and strategies for relieving stress (e.g., taking breaks, airing concerns). Findings on teaching as a profession suggest that risks associated with low-wage jobs with little administrative support might compromise teachers' ability to maintain emotionally supportive classrooms (Curbow, 1990; Raver, 2004). When given concrete forms of support from a MHC, teachers may be better able to focus their attention on meeting the needs of children in their classrooms.

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 six months of the intervention MHCs identified 3 to 4 children per class (approximately 70 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 of the school year. In addition, MHCs relied on their extensive documentation of children's behavioral and emotional adjustment using matrices to log

the number of school visits spent working with individual children as well as the primary clinical concerns addressed during "1-on-1" sessions with individual children.

Program Satisfaction. Exit survey data (from cohort 1 treatment-group teachers) suggested high acceptability of mental health consultation, with over 80% of teachers reporting 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% reported that 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 if they could "do it over again," 89% of teachers said yes.

Control Classrooms. Classrooms within sites randomly assigned to the control group were paired with Associate's degree-level teaching assistants during the intervention year. Control group teachers were also invited to attend teacher training sessions during the following year. Unlike the MHCs, teacher assistants did not attend the teacher training sessions, nor did they devote classroom visits to coaching, stress reduction, or direct services. Rather, teacher assistants provided an "extra pair of hands and eyes" during everyday classroom activities.

Facilitators and Providers.

Program Coordinator. Working with both control and treatment classrooms, the program coordinator helped ensure that the project ran smoothly across the 35 classrooms and 18 sites participating in CSRP. For example, the program coordinator scheduled and organized all teacher trainings for teachers in the intervention group, acted as a community liaison, and encouraged teachers in the intervention group to register for the training sessions. This included following up with intervention-group teachers regarding their plans to attend, making child care

arrangements for teachers' own children during the sessions, and facilitating timely reimbursement for teachers' training participation. The program coordinator also monitored teacher assistants in the control condition by collecting weekly logs of the hours they spent in control classrooms.

In addition to the logistical and administrative support provided by the Program Coordinator, the MHCs received clinical supervision from a licensed clinical social worker with a Master's degree. The clinical supervisor's specific responsibilities included: leading weekly meetings with the MHCs, and supervising and monitoring the MHCs as they delivered services. In doing so, the clinical supervisor (in consultation with the principal investigator) delineated aspects of the treatment approach the MHCs utilized, and was available for trouble-shooting the unavoidable challenges associated with the implementation of a research-based randomized intervention program. A second licensed clinical social worker was included on the intervention team as a leader of teacher trainings. The teacher training leader used the Incredible Years Teacher Training Program (Webster-Stratton, 1995; Webster-Stratton et al., 2004) and had extensive experience leading parent and teacher groups in community-based child-serving agencies in the Chicagoland area.

Mental Health Consultants. 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. Furthermore, the principal investigator placed a heavy emphasis on MHCs' ability to deliver culturally competent services, requiring that the MHCs had experience working in ethnic minority communities. Additionally, when making hiring decisions, the principal investigator considered languages spoken (e.g., Spanish) and cultural match across MHCs, teachers, and children.

Treatment Fidelity.

Though trainings were offered to all teachers randomized to treatment, teachers attended 3 of the 5 trainings on average. Similarly, even though classroom visits were a main ingredient of the intervention package, the number of visits made by mental health consultants varied from 21 to 40, with an average of 29 visits during the academic year. This resulted in classrooms participating in an average of 128 hours of mental health consultation by the end of the academic year. The most common social services offered to teachers were social support and coaching, with mental health consultants providing social support and coaching during a mean number of 23 and 15 visits, respectively. In addition, preliminary findings suggest that classrooms' receipt of treatment was relatively evenly distributed across programs that varied along a range of classroom and teacher characteristics. For example, our preliminary analyses of take-up of services suggest that teachers who faced more stress were just as likely to participate in training and to put the intervention's behavior management techniques into practice (Li-Grining et al., 2007).

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, with approximately 17 children in each classroom enrolling in CSRP. Consent forms for each child were signed by his or her parent or guardian, who also completed an interview, which included questions regarding child and family demographic characteristics, such as child's gender and race/ethnicity, as well as parent's preferred home language, marital status, education level, and employment status. Parents also reported on household income and number of children in the household.

Children's behavior problems were rated by teachers in both fall and spring of the Head Start year. Teachers were given the Behavior Problems Index (BPI) 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. CSRP research staff (blind to the treatment status of the classroom) dropped off questionnaire packets to teachers with prepaid postage envelopes for easy return, making follow-up phone calls and visits to collect teacher reports within a 6-week window. Teachers were subsequently reimbursed a nominal subject payment (of $20 per packet) for completing questionnaires during each round of assessment.

For the observational subsample of CSRP, children's aggressive/disruptive and withdrawn/disconnected behaviors were observed during the course of the school day by 12 trained observers using the Penn Interactive Peer Play Scale (developed by Fantuzzo, et al.,1995; see also Milfort & Greenfield, 2002). Observers rated each child's behavior with peers for twenty-minute blocks occurring throughout the day. A cover sheet was also included that asked observers to record the child's ID, date and time of the observation, school, classroom, teachers present and their names, number of children present, as well as a brief description of activity taking place during the observation.

To account for classroom-level differences in resources and support for children's social-emotional development, both trained observers and teachers provided classroom-level data in the fall. 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 degree. Of the 12 members, 6 were African American and 6 were Caucasian or Asian,

thus approximately half the time, the observers' race matched the race of most children. Using the ECERS and CLASS, observers rated overall classroom quality as well as dimensions of emotional climate, including teacher sensitivity, behavior management, and negative climate. While conducting observations, staff noted the number of children and adults in the classroom. Teachers leading each classroom, as well as assistant teachers, completed self-reports, providing demographic information such as their age and level of education and rating their psychosocial well-being, in terms of depressive symptoms (Kessler et. al, 2002) and job overload (Curbow, 1990).

In the fall, administrators at each Head Start site provided CSRP with access to site-level characteristics. These characteristics included information regarding staff, such as the average teacher salary, the percentage of teachers with a B.A., the percentage of teaching assistants with any college education, teachers' salary, and the availability of a family support worker on site. Administrators also provided data on children and families, such as total enrollment, percent of children who were African American, and the percentages of families who were led by single parents, who had at least one parent employed, and who received TANF.

Measures

Dependent measures of treatment impact. To assess the impact of the CSRP intervention on children's behavior problems, two measures of children's internalizing and externalizing behavior problems were included. 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 of the Head Start year. The C-TRF was designed to be completed by day care providers, teachers or any other care givers that have observed a child in the previous two months by rating the child's behavioral and emotional problems. The measure consists of 99

items asking the respondent to rate the child on a scale from 0 to 2 (where 0 = not true, 1= somewhat or sometimes true, and 2= very true or often true). Teachers' reports on these 99 items were summed into two domains, Internalizing (α= .90) and Externalizing (α= .97) subscales.

In addition, teachers completed the Behavior Problems Index (BPI) in both fall and spring, a 28-item rating scale originally designed for parent report of child behavior adapted from multiple studies of children's behavior problems by Nicholas Zill and James L. Peterson (Zill, 1990). CSRP modified the original version in several minor ways, including combining two items regarding children's disobedience (e.g. "is disobedient at home" and "is disobedient at school" into a single item). For the purposes of this study items were summed to form two domains, Internalizing (α= .80) and Externalizing (α= .92) following recommendations of the National Longitudinal Survey of Youth (NLSY79) (Zill, 1990).

In order to provide independent observational assessments of children's behavior problems, CSRP received permission from the author, John Fantuzzo, to use the PIPPS as an observational tool. The original measure was altered slightly (e.g., by asking observers to only report whether the child displayed the specific behavior (yes=1) or not (no=0) versus the original measure's use of a 4-point scale (never, seldom, often, always). Preliminary factor analyses of observer scores on the 30 items suggested high levels of inter-item reliability for two out of the three subscales (aggression/disruption $\alpha_{MAY} = .71$, withdrawal/disconnection $\alpha_{MAY} = .64$) with items loading over $\geq .40$. The third subscale of interaction showed inadequate inter-item reliability (alphas <.60) and so was dropped from further analysis. Inter-rater reliability was high, ranging from 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).

Child-level Covariates. Child-level demographic characteristics were also 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 3 poverty-related risks (including mothers' educational attainment of less than high school degree, family income-to-needs ratio for previous year less than half the federal poverty threshold, and mothers' engagement in 10 hours or fewer of employment per week). Previous analyses with large, nationally-representative data sets by the first and second authors suggest that these factors represent the most reduced and informative set of measured indicators for families' exposure to "deep poverty." (Raver, Garner, & Smith-Donald, 2007; Jones et al., submitted). A final child-level covariate was the child's total behavioral problems score (as rated by teachers on the BPI) in fall.

Classroom/Teacher-level Covariates. A set of teacher characteristics were included as proxy assessments of classroom quality, and were assessed through teacher report. These included teachers' reports regarding their age, level of education (teacher's attainment of BA as well as teacher assistant's attainment of BA), as well as their report on several psychosocial characteristics that might affect teachers' perceptions of children's behavioral difficulty (see Anthony et al., 2005). To assess their baseline psychosocial characteristics, teachers' depressive symptoms were briefly assessed at baseline using the *K6*. The K6 is a six-question truncated scale embedded within a ten-question screening scale of psychological distress developed for the redesigned US National Health Interview Survey (Kessler et al., 2002). With simple scoring, we converted the K6 to a 0-24 scale where each of the six questions are coded as 0-4 and then summed ($\alpha = .65$). In addition, teachers' reports of job overload were included, using the "job

demands" and "job control" subscales of the Child Care and Early Education Job Inventory (Curbow et al. 2000). All items in the measure use a 5 point scale ranging from 1= "Disagree", 3= "Neither Agree nor Disagree", and 5= "Agree". The "job demands" scale ($\alpha = .67$) was based on 6 items (e.g., all children need attention at the same time), and "job control" ($\alpha = .56$) was based on 5 items (e.g., control over getting children to do what you want). In order to calculate teachers' demographic and psychosocial characteristics (e.g. age, K6 scores) as classroom-level covariates, scores on each variable were averaged across all teachers in each classroom (e.g., Gerard & Buehler, 2004; Li-Grining et al., 2007).

To capture important classroom-level covariates and to control for the variability in sites' classroom quality, observational measures of classroom quality were also included. Classroom quality was assessed in fall using the CLASS (La Paro et al., 2004) and the Early Childhood Environment Rating Scale-R (ECERS-R, Harms et al., 2003). CLASS indicators included 7-point Likert scores on negative climate, teacher sensitivity, and behavior management. The negative climate score reflected teachers' expressions of anger, sarcasm, or harshness. Teacher sensitivity measured teachers' responsiveness to the children's needs or the extent to which they provided a "secure base" for the children. The final indicator, behavior management, captured teachers' strategies of structuring the classroom so that the children knew what was expected of them, as well as the use of appropriate redirection when children demonstrated challenging behavior. Three-quarters of the observations were double coded "live" by two observers to gauge inter-rater reliability. Because measures were coded on an ordinal scale from 1 to 7, inter-rater reliability was established via calculation of intraclass correlation values (α), which indicated adequate to high levels of inter-observer agreement (negative climate, $\alpha = .70$; teacher sensitivity, $\alpha = .77$; behavior management, $\alpha = .66$).

The ECERS-R (Harms et al., 2003) is a widely used research tool used to measure early childhood classroom quality across a wide range of constructs. Based on 43 items, the ECERS-R provides an observational snapshot of the "use of space, materials and experiences to enhance children's development, daily schedule, and supervision" where each item is scored from one to seven (ranging from 1 = "Inadequate" to 7 = "Excellent"; 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, with 43% of the ECERS-R observations double-coded for purposes of reliability (α =.87 for the ECERS-R Total Score). In addition, the number of children and the number of teachers observed in each classroom in September were also included, to control for the potential confounds of differences in class size or staffing ratios at both time points.

Site-level Covariates. In order to test the role of "settings-level" program characteristics, a limited number of site-level covariates were entered into models, including the availability of a fulltime family worker at the Head Start site, the size of the program (as indexed by the number of children served, ages 3-5), the proportion of the site identified as African American, the proportion of the teachers with Bachelors' degrees and the proportion of teacher assistants with some college, the proportion of the families served that were single parent families, employed, and reliant on TANF.

Results

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 of the intervention year as dependent measures of CSRP program influence. We then repeated our analyses with observational assessments of children's aggressive/disruptive and withdrawn/disconnected

behaviors collected in May for a stratified randomly-selected subsample of children, as an additional set of dependent measures of CSRP program influence. In both sets of analyses, a multi-level strategy is necessary because children in this study are nested in classrooms and sites. Multilevel modeling allows for the simultaneous estimation of variance associated with individual (within-subjects) and population (between-subjects) change based on the specification of fixed- and random-effect variables in the model (Raudenbush & Bryk, 2003). In the present study, the assessment of children in fall and spring of the Head Start year allows for estimates of residualized change in behavior problems as an indicator of treatment impact. With these data it is possible to assess the direct impact of school- or classroom-level (e.g., intervention vs. control) variables, net of both person-level (e.g., demographic and ecological characteristics of children and families), classroom/teacher-level, and site-level characteristics.

The overall impact of intervention was then modeled using 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 and family characteristics, including child's gender, race/ethnicity, and pre-treatment BPI; single-parent family, parents aged of 22 or younger, four or more children in household, Spanish speaking, TANF receipt, and number of family risks (i.e., parents' education, employment, and income). ε_{ijk} is a random error term.

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

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

where $\sum_n \beta_{mnk} C_{mnjk}$ is the sum of n teacher and classroom characteristics, including teacher's BA degree, age, teacher's K6 score, job demand, and job control 3 baseline (fall) CLASS variables and overall baseline ECERS scores, class size, and number of adults in class

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 while $\sum_p \gamma_{m0pk} S$ represents the sum of p site-level characteristics, including whether site had additional family support worker on staff, total number of children ages 3 to 5 by summing racial groups, and percentages of teachers with BA, teacher assistants with any college education, African American/black children, single-parent families, families with at least one parent employed, and families receiving TANF. 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 treatment or control group; γ_{000} is the adjusted mean level of behavioral problems across all control group sites; and γ_{001} is the treatment effect. Though not shown here, γ_{100} - γ_{800} represent the pooled within-site regression coefficients for the Level-1 covariates. The magnitude of treatment impact can then be examined, where γ_{001} represents the average difference between treatment and control sites, controlling for all covariates. Effect sizes are calculated by dividing that difference by the full sample's SD for the measure used as the dependent variable.

We first conducted HLM analyses with a baseline unconditional model to estimate the proportion of variance in children's behavior problems that is attributable to the child-, classroom, and site levels (not shown). We then use HLM analyses, as specified above, to

estimate the treatment effects of CSRP. We further investigated whether the treatment impacts (if any) were moderated by child race/ethnic status, child gender, or child low vs. high exposure to poverty-related risks. Because this study used a cluster-randomized design with a relatively small number of program sites randomized to control and treatment conditions (n=18), and given the known difficulty detecting interaction effects with small sample sizes (McClelland & Judd, 1993), omnibus tests of interactions between treatment and the three possible moderators, regardless of statistical significance, were followed with post-hoc analyses of treatment impact within subgroups. These are described in the results section and are depicted in figures. Finally alternative model specifications were included, with the 8 pairwise site assignments 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 treatment on child behavior problems using alternative specification of the three-level model, whereby estimates of treatment were yielded net of children's enrollment in specific matched pairs of sites, rather than net of the site characteristics used to conduct pairwise matches. Implications of these different specifications of the model of intervention impact are discussed below.

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 treatment 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 treatment impact, as illustrated below.

Tables 2 and 3 present the "Intent to Treat" or ITT estimates of CSRP intervention services on teachers' reports of children's internalizing and externalizing BPI and TRF scores. Model 1 included all 3-level covariates; Model 2 included interaction terms between treatment and child race/ethnicity, gender, and family socioeconomic risks; while in Model 3, site-pair dummy variables were included instead of site-level covariates.

The results in Tables 2 and 3 show that overall, CSRP benefited children in the treatment group, whereby children in the treatment group were reported as having significantly fewer internalizing and externalizing behavior problems than did their control group-enrolled counterparts by spring of their Head Start year. In particular, compared to children in the control group, those in treatment group on average had significantly lower scores on BPI Internalizing (1.80 points in Model 1) and Externalizing (2.92 points), and TRF Internalizing (3.30 points) and Externalizing (5.08 points).

After including the interactions between treatment and child race/ethnicity, gender, and family socioeconomic risks (Model 2), children in the treatment group had even lower BPI (2.15 and 4.50 points for Internalizing and Externalizing scales, respectively) and TRF (3.96 and 8.77 points for Internalizing and Externalizing scales respectively) and TRF scores (with differences equaling 3.96 and 8.77 points on internalizing and externalizing scales, respectively) scores. Estimates of effect sizes for these decreases in children's behavior problems range from d = .53 to d = .89.

Tables 2 and 3 also reveal several important main effects of the Level 1 covariates, such that children with higher levels of behavioral risk in the fall of Head Start year had significantly higher BPI and TRF scores in the spring. Children of single-parents tended to have more externalizing behavior problems measured by BPI (Model 1) and TRF (Models 1 and 2) scales.

In contrast, children whose parents spoke Spanish were less likely to have BPI (Models 1 and 2) and TRF (Model 2) internalizing behavior problems. Finally, the number of children in the household was not significantly related to child BPI or TRF scores.

Among the teacher and classroom-level covariates as predictors of children's behavior problems, teacher's ratings of job demand and job control were both positively associated with child BPI and TRF internalizing and externalizing scores. In addition, while teacher's K6 scores were negatively associated with children's behavior problem scores. In contrast teacher's behavior management was negatively related to child BPI and TRF scores, while teacher's sensitivity showed an unexpected positive association. Similarly, negative climate was negatively associated with child BPI and TRF internalizing scores, while class sizes were associated with both lower internalizing and externalizing scores of BPI and TRF. Other teacher and class covariates, such as teacher with BA, age, ECERS overall classroom quality, and number of the adults in class, were not associated with child BPI or TRF scores.

Among the covariates at the site level, Tables 2 and 3 show that 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. In contrast, children enrolled in larger sites, in sites with higher percentages of African American children, teachers with BAs, and families with at least one parent employed tended to be rated by teachers as having a higher number of behavior problems. Other site-level covariates, such as percentages of single families and families receiving TANF, did not show significant associations with child BPI or TRF scores.

Because of the known difficulty in detecting the effects of moderators, we then graphed point estimates and confidence intervals for all sub-groups, to examine whether the CSRP

treatment demonstrated differential efficacy with some groups of children compared to others. Figures 2-5 show that overall girls and Hispanic children in the treatment group tended to show larger point estimate reductions in behavior problems than did their counterparts in the control group. For example, girls in the treatment group were rated 1.98 and 3.88 points lower in BPI internalizing and externalizing scores, respectively, than girls in the control group; while boys in the treatment group were rated 1.65 and 1.98 points lower, respectively, than boys in the control group. Compared to their same race/ethnic group peers who were in the control group, Hispanic children in the treatment group tended to score 2.09 and 3.45 points lower in BPI internalizing and externalizing scores, respectively; while Non-Hispanic Black children in the treatment group scored 1.36 and 1.81 points lower than Non-Hispanic Black children in the control group. What these figures illustrate 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 treatment.

Results from additional post-hoc sensitivity analyses (Model 3) revealed that treatment effects were no longer statistically significant when binomial codes for pairwise matches between sites were substituted for program characteristics as level-3 covariates. It should be noted that the same site-level covariates were employed to determine the matches. As such, we are confident that the difference between these two specifications of the model is attributable to the sensitivity of the estimates to the inclusion vs. exclusion of specific program pairs. In short, these smaller and non-significant estimates probably reflect heterogeneity of program impact with small samples (n=18 sites in 9 pairs), which limits the statistical power of our analyses.

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 4, there were marginally significant effects for treatment on children's aggressive/disruptive behavior, B = -1.40, $p < .10$. In addition, there was a statistically significant treatment by poverty-related risk interaction, B = .92, $p < .01$. As shown in Figure 6, children with no or one risk tended to benefit from CSRP treatment, while the treatment effect for children with two or more risks was not statistically significant. These analyses suggest sparse evidence for the role of child, teacher/classroom, and site characteristics as significant predictors of observed behavior problems (see Table 4).

Discussion

Results from our analyses suggest that CSRP had a large, statistically significant impact on reducing low-income preschoolers' internalizing and externalizing behavior problems. On average, 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 with effect sizes ranging from d = .89 for teacher-reported Behavior Problems Index scores to d = .62 for Teacher Report from CTRF. The multicomponent CSRP model of intervention also demonstrated efficacy in reducing preschoolers' externalizing behavior problems, including symptoms of aggression and defiance, with d = .64 and d = .53 on the BPI and TRF, respectively.

Results from our observational assessments of children's disruptive behavior confirmed our teacher-reported findings, providing important confirmation of the impact of this intervention on children's outcomes. Among the randomly selected subsample of children in our study, a trend suggested that children in the treatment group showed fewer disruptive behaviors during their school day than did children in the control group. Results from our observational data

suggested that benefits of the intervention were more robust for children facing lower levels of poverty-related risk, with a statistically significant treatment X poverty-related risk interactions for children's disruptive behavior (See Figure 5). Observations of children's disconnected behaviors were in the hypothesized direction, but were not sufficiently large to reach levels of statistical significance.

In short, results indicate that classroom-based multicomponent intervention combining evidence-based teacher-training and mental health consultation is an important avenue 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 support and build on the strengths of low-income families and the institutions that serve them (Berryhill & Prinz, 2003; Dishion volume& Stormshak, 2006; Webster-Stratton & Taylor, 2001; Tolan, Gorman-Smith & Henry, 2004; Yosikawa & Shinn, 2008; Zins, Weissber, Wang, & Walberg, 2004). Our findings suggest that a key source of institutional support of children's communities, namely the Head Start-funded preschool programs in their neighborhoods, can be leveraged to provide important behavioral as well as academic intervention and support.

Placing intervention in the social context of preschools

When considering intervention 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 behavior problems among children in their classrooms. Findings by Anthony et al. (2005) suggest that some teachers reporting exceptionally few children exhibiting difficulty while other teachers report as many as 60% of their children showing serious behavioral difficulty. To disentangle the sources of this tremendous variability, this study sought to include 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 in 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 (having to meet a large number of job demands with little support) was significantly predictive of teachers' reports of higher behavior problems for the children in their classroom. Site-level characteristics such as whether programs had sufficient resources to have a full-time family support staff-member on staff were also 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 both at the level of CSRP's service delivery of the intervention and at the level of our analyses of the impact of our intervention.

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; Gilliam, 2005). 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 their behavior problems (see Jones et al., in press for review). As with those programs, CSRP placed central importance on intervention staff serving as coaches to teachers, to aid in building new relationships with students using more adaptive strategies of management and engagement (Donohue, Falk &

Provet, 2000; Gilliam, 2005; Gorman-Smith et al., 2003; Green, Simpson, Everhart, Vale, & Gettman, 2004). 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 where an increasingly large fraction of preschool children are served. These services show significant promise for Head Start-funded programs serving low-income, ethnic minority preschoolers in neighborhoods of concentrated disadvantage.

Did the CSRP intervention work better for some children, as compared to other children? Significant treatment X gender and treatment X race/ethnic group membership interactions for teachers' reports suggest that Hispanic girls showed the largest reductions in behavior problems, as compared to other children in the sample. While 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 sub-sample. Following McClelland & Judd (1993), we took additional steps to guard against the risk of type II error (that is, concluding null differences when the problem may lie with inadequate statistical power to detect differences in treatment impact for different groups). To guard against that risk, we conducted follow-up analyses of treatment impacts within each subgroup, finding no evidence that treatment impacts were substantially driven by the improvement of children at higher vs. lower risk. Follow-up subgroup analyses suggest, when children in the treatment group were compared to their race/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.

Limitations of the current study

While the present study had several strengths, this study's conclusions are constrained by several limitations. A key limitation is that the current analyses are restricted to assessments of children's behavior problems at the end of the school year in which programs were randomly assigned to treatment or control groups. For more robust estimates of CSRP's efficacy, it will be important to see whether evidence of the benefits of our classroom-based intervention extend into children's kindergarten year (Flay et al., 2005). Analyses are currently underway to detect whether treatment group-enrolled children sustain these improvements as they make transitions to new kindergarten classrooms where teachers are unaware of children's prior involvement in treatment- vs. control-group assigned Head Start programs. With this caveat in mind, our findings are in keeping with previous clinical research that has relied on post-test assessments of children's behavior problems as indicators of intervention efficacy, where short-term improvement in children's behavioral adjustment is an important outcome in its own right (see Berryhill & Prinz, 2003 for review; Gilliam, 2005).

A second 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. Ours was a relatively small-scale classroom-based study with only sufficient statistical power to randomize programs to a control group vs. to receipt of a single "package" of intervention supports. In the meantime, our study represents an important preliminary empirical step, bridging previous research traditions focusing on elementary school-based RCT trials on the one hand and programs and non-randomized 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: On average, the families in our study reported living on incomes below the federal poverty threshold, with 70% of caregivers "making ends meet" as single parents, in neighborhoods with high levels of crime and a high proportion of economically disadvantaged neighbors. An encouragingly large number of comprehensive family-based interventions have demonstrated efficacy in supporting ethnic minority parents and their children in reducing the risk of behavioral and emotional difficulty in urban communities such as Baltimore, Philadelphia, Seattle, and Chicago (Tolan et al., 2003; Webster-Stratton et al., in press). Our findings suggest that children's preschools may represent an additional avenue to pursue in building a comprehensive system of prevention and early intervention.

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., & 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 & 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., Frazier, S. L., Birman, D., Adil, J. A., Jackson, M., Graczyk, P. A., Talbott, E., et al. (2006). School based mental health services for children living in high poverty urban communities. *Administration and Policy in Mental Health and Mental Health Services Research, 33,* 146-159.

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.

Bell, C. C., & Jenkins, E. J. (1993). Community violence and children on Chicago's southside. *Psychiatry: Interpersonal and Biological Processes, 56,* 46-54.

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

Bierman, K. L., Burschi, C., Domitrovitch, C., Fang, G. Y., Miller-Johnson, S., & The Conduct Problems Prevention Research Group. (2004). Early disruptive behaviors associated with emerging antisocial behavior among girls. In M. Putallaz & K. L. Bierman (Eds.) Aggression, antisocial behavior, and violence among girls (pp. 137-161). New York: Guilford Press.

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

Briggs-Gowan, M. J., Carter, A. S., Moye Skuban, E., & McCue Horwitz, S. (2001). Prevalence of social-emotional and behavioral problems in a community sample of 1- and 2-year-old

children. *Journal of the American Academy of Child & Adolescent Psychiatry, 40,* 811-819.

Brody, G. H., Murry, V. M., Chen, Y., Kogan, S. M., & Brown, A. C. (2006). Effects of family risk factors on dosage and efficacy of a family-centered preventive intervention for rural African Americans. *Prevention Science, 7*, 281-291.

Broidy, L. M., Nagin, D. S., Tremblay, R. E., Bates, J. E., Brame, B., & Dodge, K. A. (2003). Developmental trajectories of childhood disruptive behaviors and adolescent delinquency: A six-site, cross-national study. *Developmental Psychology, 39,* 222-245.

Brouwers, A., & Tomic, W. (1998). Student disruptive behavior, perceived self-efficacy, and teacher burnout. *Netherlands Journal of Psychology, 53,* 173-183.

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.

Bryk, A. S., & Raudenbush, S.W. (1992). *Hierarchical linear models: Applications and data analysis methods.* Newbury Park, CA: Sage Publications.

Burchinal, M. R., Bailey, D. B., & Snyder, P. (1994). Using growth curve analysis to evaluate child change in longitudinal investigations. *Journal of Early Intervention, 4,* 403–423.

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.

Dodge, K. A., Pettit, G. S., & Bates, J. E. (1994). Socialization mediators of the relation between socioeconomic status and child conduct problems. *Child Development, 65,* 649-665.

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 riskfor conduct disorder, substance abuse, and school failure. *Clinical Child and Family Psychology Review, 2(1),* 37-53.

Epstein, J. N., Willoughby, M., Valencia, E. Y., Toney, S. T., Abikoff, H. B., Arnol, L. E., & Hinshaw, S. P. (2005). The role of children's ethnicity in the relationship between teacher ratings of Attention-Deficit/Hyperactivity Disorder and observed classroom behavior. *Journal of Consulting and Clinical Psychology, 73,* 424-434.

Fantuzzo, J., Stoltzfus, J., Lutz, M. N., Hamlet, H., Balraj, T., Turner, C., et al. (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.

Farmer, E. M. Z., Stangl, D. K., Burns, B. J., Costello, E. J., & Angold, A. (1999). Use, persistence, and intensity: Patterns of care for children's mental health across one year. *Community Mental Health Journal, 35,* 31-46.

Flay, B. R., Biglan, A., Boruch, R. F., Castro, F. G., Gottfredson, D., Kellam, S., Moscicki, E. K., Schinke, S., Valentine, J. C., & Ji, P. (2005). Standards of evidence: Criteria for efficacy, effectiveness, and dissemination. *Prevention Science, 3*, 151-175.

Fuchs, L. S., Fuchs, D., & Bishop, N. (1992). Instructional adaptation for students at risk. *Journal of Educational Research, 88,* 281–289.

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, NY: 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, Illinois: 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.

Gorman-Smith, D., & Tolan, P. (1998). The role of exposure to community violence and developmental problems among inner-city youth. *Development and Psychopathology, 10,* 101-116.

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(1)*, 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.

Green, B., Simpson, J., Everhart, M., Vale, E., & Gettman, M. (2004). Understanding integrated mental health services in Head Start: Staff perspectives on mental health consultation. *National Head Start Association Dialogue, 7*, 35—60.

Greenberg, M. T. (2004). Current and future challenges in school-based prevention: The researcher perspective. *Prevention Science, 5(1),* 5-13.

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, revised edition.* New York: Teachers College Press.

Hawkins, J. D., Catalano, R. F., Kosterman, R., Abbott, R. D., & Hill, K. G. (1999). Preventing adolescent health-risk behaviors by strengthening protection during childhood. *Archives of Pediatrics & Adolescent Medicine, 153(3)*, 226-234.

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 meeting of the Society for Research in Child Development, Minneapolis, MN.

Jones, S. M., Brown, J., Aber, J. L., & Thomas, G. (2006, June). *Quantity and quality of implementation of the 4Rs program: Links to teacher and classroom characteristics.* Paper presented at the Annual Meeting of the Society for Prevention Research, San Antonio, TX.

Jones, S. M., Raver, C. C., Metzger, M., & Silver, M. (in press). How does trouble come knocking? Using a measurement modeling approach to poverty-related risks and children's behavioral problems over time. *Applied Developmental Science.*

Keenan, K., & Shaw, D. (1997). Developmental and social influences on young girls' early problem behavior. *Psychological Bulletin, 121,* 95-113.

Keenan, K., Wakschlag, L., S., Danis, B., Hill, C., Humphries, M., Duax, J., & Donald, R. (in press). 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.

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., Walters, E.E., & Zaslavsky, A. (2002). Short screening scales to monitor population prevalence and trends in nonspecific psychological distress. *Psychological Medicine*. 32(6), 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: Lawrence Erlbaum.

La Paro, K., Pianta, R., & Stuhlman, M. (2004). Classroom assessment scoring system (CLASS): Findings from the pre-k year. *The Elementary School Journal, 104,* 409–426.

Li-Grining, C. P., Raver, C. C., Smallwood, K. M., Sardin, L., Metzger, M. W., & Jones, S. M. (tentatively accepted for publication). Classroom quality and teacher support in Head Start classrooms: The role of teachers' psychosocial stressors. *Journal of School Psychology*.

Li-Grining, C., Raver, C. C., Madison-Boyd, S., & Jones-Lewis, D. (in preparation). Improving classrooms' emotional climate and preschoolers' socioemotional adjustment: Lessons learned from the Chicago School Readiness Project. *Infant Mental Health Journal.*

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

LoCasale-Crouch, J., Konold, T., Pianta, R., Howes, C., Burchinal, M., Bryant, D., et al. (2007). Observed classroom quality profiles in state-funded pre-kindergarten programs and associations with teacher, program, and classroom characteristics. *Early Childhood Research Quarterly, 22,* 3–17.

Lochman, J. E. (2001). Issues in prevention with middle-aged children: Ongoing intervention refinement, developmental theory, prediction and moderation, and implementation and dissemination. *Prevention & Treatment, 4(1).*

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.*

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

Morris-Rothschild, B., & Brassard, M. R. (2006). Teachers' conflict management styles: The role of attachment styles and classroom management efficacy. *Journal of School Psychology, 44,* 105–121.

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. (2003). *Hierarchical linear models: Applications and data analysis methods.* California: Sage Publications Inc.

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

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.

Raver, C. C., & Knitzer, J. (2002). Ready to enter: What research tells policymakers about strategies to promote social and emotional school readiness among three- and four-year-

old children (Promoting the Emotional Well-being of Children and Families Policy Paper 3). New York, NY: National Center for Children in Poverty, Columbia University Mailman School of Public Health.

Reid, J. B., Eddy, J. M., Fetrow, R. A., & Stoolmiller, M. (1999). Description and immediate impacts of a preventive intervention for conduct problems. *American Journal of Community Psychology, 27,* 483-518.

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

Schaeffer, C. M., Petras, H., Ialongo, N., Masyn, K. E., Hubbard, S., Poduska, J., et al. (2006). A comparison of girls' and boys' aggressive-disruptive behavior trajectories across elementary school: Prediction of young adult antisocial outcomes. *Journal of Consulting and Clinical Psychology, 74*

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: 2-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.

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

Tolan, P. H. (2001). Emerging themes and challenges in understanding youth violence movement. *Journal of Clinical Child Psychology, 30,* 233-239.

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., Koot, H. M., & Alink, L. R. A. (2006b). Terrible ones? Assessment of externalizing behaviors in infancy with the Child Behavior Checklist. *Journal of Child Psychology and Psychiatry, 47,* 801–810.

Van Zeijl, J., Mesman, J., Van Ijzendoorn, M. H., Bakermans-Kranenburg, M. J., Juffer, F., Stolk, M. N., Koot, H. M., & Alink, L. R. A. (2006a). 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. (1995). Parent-training with low-income families: Promoting parental engagement through a collaborative approach. In J. R. Lutzker (Eds.), *Handbook of Child Abuse Research and Treatment* (pp. 183-209). New York: Plenum Press.

Webster-Stratton, C., & Reid, M. J. (2003). The Incredible Years parents, teachers and child training series: A multifaceted treatment approach for young children with conduct problems. In A. Kazdin & J. Weisz, (Eds.), *Evidence-based Psychotherapies for Children and Adolescents.* Guilford Press.

Webster-Stratton, C., Reid, M. J., & Hammond, M. (2001). Preventing conduct problems, promoting social competence: A parent and teacher training partnership in Head Start. *Journal of Clinical Child Psychology, 30,* 238–302.

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.

Webster-Stratton, C., Reid, M. J., & Stoolmiller, M. (in press). Preventing conduct problems and improving school readiness: Evaluation of the Incredible Years Teacher and Child Training Programs in high-risk schools. *Journal of Child Psychology and Psychiatry.* (OnlineEarly Articles).

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

Weissberg, R. P. (2000). Improving the lives of millions of school children. *American Psychologist, 55,* 1360-1373.

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.

Yoshikawa, H., & Shinn, M. (in press). Facilitating change: Where and when should community psychology intervene? In T. A. Revenson, et al. (Eds.), *Community psychology: A quarter century of theory, research in action and social historical context*. New York: Kluwer Academic/Plenum.

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

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

Footnotes

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 the study at baseline and remaining in the study throughout the school year. In addition to the original group at baseline, 59 children entered the study later in the school year, with five of these children subsequently exiting and 54 of them remaining in the study. Thus, there were 509 children (455 children who entered at baseline and 54 children who entered after baseline) participating in the study by the spring. Attrition analyses suggest that differences in exits and entry between treatment and control groups were minimal and unlikely to bias analyses of treatment impact (see Raver et al., 2008).

2) Note: dots represent point estimates; lines represent their 95% confidence intervals. (For figures 2 – 6)

Table 1 CSRP Descriptive Statistics (N = 547)

	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)
TRF internalizing score	4.21 (4.70)	5.46 (5.21)	2.93 (3.71)
TRF externalizing score	7.84 (9.58)	9.94 (11.06)	5.69 (7.21)
PIPPS disruption score†	1.60 (1.88)	1.57 (1.85)	1.64 (1.91)
PIPPS disconnection score†	1.18 (1.52)	1.12 (1.48)	1.23 (1.56)
Child and Family Characteristics			
Boy	0.47 (0.50)	0.53 (0.50)	0.42 (0.49)
Black	0.65 (0.48)	0.66 (0.48)	0.64 (0.48)
Hispanic	0.28 (0.45)	0.28 (0.45)	0.27 (0.45)
Other race	0.07 (0.26)	0.06 (0.24)	0.09 (0.28)
Child BPI fall risk	0.29 (0.45)	0.31 (0.46)	0.26 (0.44)
Parent human capital risk	1.08 (1.00)	1.15 (1.01)	1.00 (0.99)
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	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)
Number of 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)
Number of children aged 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 BAs	0.43 (0.39)	0.50 (0.36)	0.36 (0.41)
Proportion of teacher assistants 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)

† The sample size of PIPPS data was 181; the descriptive statistics within PIPPS sample were similar to those in the full sample, as presented in this table.

Table 2 BPI Internalizing and Externalizing Scores

	BPI Internalizing			BPI Externalizing		
	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
Treatment	-1.81**	-2.15**	-0.19	-2.92*	-4.50**	-0.22
	(0.43)	(0.54)	(0.51)	(0.92)	(1.07)	(1.05)
Child and Family Characteristics						
Boy	-0.14	-0.31	-0.28	1.26**	0.39	0.47
	(0.16)	(0.23)	(0.23)	(0.36)	(0.73)	(0.68)
Treatment*Boy		0.33	0.24		2.30*	1.94+
		(0.32)	(0.32)		(1.03)	(0.94)
Black	0.21	-0.08	-0.28	0.46	-0.49	-0.43
	(0.35)	(0.47)	(0.44)	(0.78)	(1.09)	(1.00)
Treatment*Black		0.82	1.18*		1.72	1.85
		(0.55)	(0.49)		(1.24)	(1.10)
Other race	0.39	0.05	0.32	0.78	0.34	1.40
	(0.39)	(0.52)	(0.55)	(0.88)	(1.70)	(1.52)
Treatment*(other race)		0.64	0.43		1.90	0.55
		(0.73)	(0.78)		(2.27)	(2.07)
HC_RISK	0.12	0.17	0.16	0.08	0.18	0.21
	(0.08)	(0.13)	(0.15)	(0.19)	(0.27)	(0.31)
Treatment*HC_RISK		-0.12	-0.09		-0.26	-0.32
		(0.18)	(0.21)		(0.38)	(0.44)
Fall total BPI risk	1.25**	1.25**	1.24**	3.45**	3.48**	3.45**
	(0.21)	(0.21)	(0.20)	(0.46)	(0.45)	(0.45)
Single parent	0.18	0.21	0.16	0.68+	0.60	0.52
	(0.18)	(0.18)	(0.18)	(0.41)	(0.40)	(0.39)
>=4 kids in HH	0.15	0.17	0.16	-0.44	-0.35	-0.31
	(0.19)	(0.19)	(0.19)	(0.42)	(0.41)	(0.41)
Spanish speaking	-0.58*	-0.59*	-0.60*	-1.00	-1.04	-1.12+
	(0.29)	(0.29)	(0.29)	(0.65)	(0.64)	(0.63)
Teacher and Class Characteristics						
Teacher BA	0.03	0.01	-0.01	0.65	0.77	1.25+
	(0.30)	(0.29)	(0.32)	(0.64)	(0.55)	(0.64)
Teacher age	-0.02	-0.01	0.04**	-0.02	0.00	0.08**
	(0.01)	(0.01)	(0.01)	(0.03)	(0.03)	(0.02)
Teacher K6 score	-0.23**	-0.17+	0.05	-0.33+	-0.15	0.20
	(0.08)	(0.09)	(0.06)	(0.17)	(0.17)	(0.13)
Teacher job demand	0.90**	0.85**	0.23	1.88**	1.95**	0.45
	(0.28)	(0.27)	(0.23)	(0.60)	(0.53)	(0.47)
Teacher job control	0.62**	0.66**	0.47*	0.69+	0.94*	0.15
	(0.19)	(0.18)	(0.20)	(0.40)	(0.35)	(0.39)
Teacher behavior management	-1.23**	-1.24**	-0.75**	-2.33**	-2.64**	-1.87**
	(0.27)	(0.25)	(0.23)	(0.57)	(0.49)	(0.46)
Teacher sensitivity	0.83*	0.70*	-0.38	2.35**	2.57**	0.85
	(0.32)	(0.32)	(0.27)	(0.69)	(0.63)	(0.57)
Negative climate	-0.84**	-0.79**	-0.56*	-0.25	0.07	0.65
	(0.22)	(0.21)	(0.21)	(0.47)	(0.41)	(0.42)
ECTOTAL	-0.25	-0.03	0.86**	0.01	0.46	2.12**
	(0.25)	(0.27)	(0.27)	(0.54)	(0.53)	(0.53)

Class size	-0.33**	-0.29**	-0.16*	-0.80**	-0.73**	-0.66**
	(0.07)	(0.07)	(0.06)	(0.14)	(0.13)	(0.13)
# adults in class	0.18	0.15	-0.16	0.46	0.62	-0.28
	(0.22)	(0.22)	(0.26)	(0.47)	(0.42)	(0.51)
Site Characteristics						
Family support worker on staff	-1.03**	-0.89**		-1.77**	-1.62**	
	(0.21)	(0.22)		(0.46)	(0.44)	
Number of children aged 3-5	0.02**	0.02**		0.03**	0.03**	
	(0.00)	(0.00)		(0.01)	(0.01)	
% African Americans	1.57*	1.00		2.27	1.23	
	(0.68)	(0.71)		(1.46)	(1.48)	
% teachers with BA	1.92*	1.75*		3.04*	2.47+	
	(0.60)	(0.58)		(1.27)	(1.13)	
% teacher assistants with college	-2.50**	-2.37**		-4.39**	-3.94**	
	(0.47)	(0.47)		(1.01)	(0.91)	
% single families	-0.07	0.16		0.71	1.93	
	(0.98)	(1.02)		(2.09)	(2.05)	
% families employed	2.40*	1.97+		5.05*	4.35*	
	(0.97)	(0.98)		(2.05)	(1.87)	
% families with TANF	0.40	0.29		1.28	1.63	
	(0.48)	(0.48)		(1.03)	(0.92)	
Site Characteristics or Matching						
Site pair 01			0.50			2.80+
			(0.71)			(1.47)
Site pair 02			2.12*			5.59*
			(0.80)			(1.61)
Site pair 03			1.51+			3.29*
			(0.66)			(1.38)
Site pair 04			3.43**			6.74**
			(0.67)			(1.38)
Site pair 06			1.15			4.09*
			(0.66)			(1.38)
Site pair 07			0.83			1.81
			(0.76)			(1.58)
Site pair 08			1.69+			4.01+
			(0.89)			(1.83)
Site pair 09			2.80**			4.56*
			(0.69)			(1.46)
Constant	5.65*	4.78+	1.49	-0.44	-0.35	-1.63
	(2.33)	(2.30)	(1.82)	(0.42)	(0.41)	(3.57)

Standard errors in parentheses; + $p < .10$; * $p < .05$; ** $p < .01$

Table 3 TRF Internalizing and Externalizing Scores

	TRF Internalizing			TRF Externalizing		
	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
Treatment	-3.03*	-3.96*	0.08	-5.08*	-8.77**	-1.25
	(0.98)	(1.49)	(1.20)	(1.71)	(2.20)	(2.21)
Child and Family Characteristics						
Boy	0.21	-0.25	-0.28	2.59**	0.61	0.66
	(0.37)	(0.60)	(0.57)	(0.77)	(1.63)	(1.48)
Treatment*Boy		0.89	0.82		5.18*	4.58*
		(0.84)	(0.79)		(2.28)	(2.07)
Black	0.11	-0.06	-0.84	1.01	-0.96	-1.49
	(0.81)	(1.18)	(1.07)	(1.67)	(2.14)	(2.02)
Treatment*Black		1.76	2.79*		4.01	4.90*
		(1.46)	(1.21)		(2.39)	(2.15)
Other race	0.68	0.56	0.44	2.86	1.64	2.75
	(0.91)	(1.30)	(1.24)	(1.85)	(3.36)	(2.96)
Treatment*(other race)		0.83	0.81		3.89	1.93
		(1.88)	(1.77)		(4.51)	(4.10)
HC_RISK	0.22	0.08	0.07	0.41	0.60	0.60
	(0.20)	(0.37)	(0.42)	(0.41)	(0.57)	(0.63)
Treatment*HC_RISK		0.14	0.13		-0.69	-0.70
		(0.53)	(0.59)		(0.80)	(0.89)
Fall total BPI risk	2.67**	2.66**	2.63**	7.54**	7.57**	7.56**
	(0.48)	(0.47)	(0.47)	(0.99)	(0.96)	(0.96)
Single parent	0.54	0.55	0.48	1.67+	1.62+	1.50+
	(0.42)	(0.41)	(0.41)	(0.87)	(0.84)	(0.84)
>=4 kids in HH	0.02	0.09	0.07	-1.28	-1.17	-1.07
	(0.44)	(0.43)	(0.43)	(0.90)	(0.86)	(0.87)
Spanish speaking	-1.03	-1.13+	-1.19+	-1.95	-2.06	-1.98
	(0.67)	(0.66)	(0.66)	(1.38)	(1.35)	(1.35)
Teacher and Class Characteristics						
Teacher BA	0.37	0.35	0.88	1.81	2.05+	3.00*
	(0.69)	(0.65)	(0.76)	(1.20)	(1.16)	(1.32)
Teacher age	-0.04	-0.01	0.09**	-0.07	-0.04	0.12*
	(0.03)	(0.03)	(0.03)	(0.06)	(0.06)	(0.05)
Teacher K6 score	-0.33+	-0.21	0.24	-0.65+	-0.30	0.34
	(0.19)	(0.20)	(0.15)	(0.33)	(0.35)	(0.26)
Teacher job demand	1.42*	1.39*	-0.25	3.42**	3.55**	0.71
	(0.65)	(0.61)	(0.56)	(1.11)	(1.10)	(0.97)
Teacher job control	1.33**	1.45**	0.95+	1.46+	2.07**	0.63
	(0.42)	(0.41)	(0.47)	(0.73)	(0.73)	(0.82)
Teacher behavior management	-2.38**	-2.40**	-1.23*	-4.82**	-5.34**	-3.65**
	(0.61)	(0.57)	(0.55)	(1.07)	(1.04)	(0.96)
Teacher sensitivity	2.07*	1.80*	-0.87	4.73**	4.95**	1.78
	(0.74)	(0.74)	(0.65)	(1.28)	(1.32)	(1.15)
Negative climate	-1.63**	-1.45**	-1.14*	-1.06	-0.52	0.89
	(0.50)	(0.48)	(0.52)	(0.87)	(0.86)	(0.89)
ECTOTAL	-0.79	-0.25	1.52*	-0.47	0.39	3.16**
	(0.57)	(0.61)	(0.63)	(1.03)	(1.10)	(1.11)

Class size	-1.05**	-1.01**	-0.62**	-1.66**	-1.50**	-1.22**
	(0.15)	(0.16)	(0.15)	(0.27)	(0.28)	(0.26)
# adults in class	0.54	0.40	0.09	1.22	1.38	0.31
	(0.51)	(0.51)	(0.61)	(0.88)	(0.86)	(1.04)
Site Characteristics						
Family support worker on staff	-2.28**	-1.98**		-3.81**	-3.49**	
	(0.49)	(0.52)		(0.86)	(0.92)	
Number of children aged 3-5	0.04**	0.03**		0.06**	0.06**	
	(0.01)	(0.01)		(0.01)	(0.01)	
% African Americans	3.68*	1.24		4.89	2.66	
	(1.56)	(1.61)		(2.80)	(3.02)	
% teachers with BA	4.74**	4.13*		4.86+	3.85	
	(1.36)	(1.35)		(2.38)	(2.36)	
% teacher assistants with college	-4.66**	-4.12**		-6.77**	-5.98*	
	(1.08)	(1.07)		(1.89)	(1.90)	
% single families	0.87	2.48		2.15	5.34	
	(2.24)	(2.48)		(3.89)	(4.09)	
% families employed	5.58*	3.56		9.86*	7.79+	
	(2.20)	(2.24)		(3.80)	(3.88)	
% families with TANF	0.66	0.88		3.28	3.99+	
	(1.10)	(1.10)		(1.94)	(1.93)	
Site Characteristics or Matching						
Site pair 01			1.44			4.17
			(1.76)			(2.99)
Site pair 02			3.83+			7.89+
			(1.91)			(3.40)
Site pair 03			3.39+			4.71
			(1.56)			(2.80)
Site pair 04			7.68**			10.32**
			(1.62)			(2.81)
Site pair 06			1.90			7.13*
			(1.61)			(2.78)
Site pair 07			1.71			1.04
			(1.84)			(3.20)
Site pair 08			1.97			2.57
			(2.12)			(3.70)
Site pair 09			5.82**			6.29+
			(1.65)			(2.93)
Constant	15.01*		8.27+	11.66	1.84	0.65
	(5.32)		(4.30)	(9.20)	(9.31)	(7.46)

Standard errors in parentheses; + $p < .10$; * $p < .05$; ** $p < .01$

Table 4 PIPPS Disruption and Disconnection Scores

	Disruption	Disconnection
Treatment	-1.40+	-0.41
	(0.67)	(0.59)
Child and Family Characteristics		
Boy	-0.59	0.12
	(0.41)	(0.36)
Treatment*Boy	0.67	0.16
	(0.56)	(0.49)
Black	0.87	0.49
	(0.63)	(0.56)
Treatment*Black	-0.56	-0.81
	(0.71)	(0.63)
Other race	0.24	-0.20
	(0.93)	(0.81)
Treatment*(other race)	-0.62	0.12
	(1.24)	(1.08)
HC_RISK	-0.37+	0.04
	(0.20)	(0.17)
Treatment*HC_RISK	0.92**	0.02
	(0.27)	(0.23)
Fall total BPI risk	0.76*	0.79**
	(0.32)	(0.27)
Teacher and Class Characteristics		
Teacher BA	-0.17	-0.16
	(0.38)	(0.32)
Teacher age	-0.04*	-0.00
	(0.02)	(0.02)
Teacher K6 score	0.01	-0.05
	(0.09)	(0.08)
Teacher job demand	0.50	0.10
	(0.33)	(0.28)
Teacher job control	-0.00	0.02
	(0.23)	(0.19)
Negative climate	-0.15	0.42+
	(0.28)	(0.24)
ECTOTAL	-0.38	-0.22
	(0.32)	(0.27)
Class size	-0.09	0.05
	(0.09)	(0.08)
# adults in class	0.28	-0.32
	(0.27)	(0.24)
Site Characteristics		
Family support worker on staff	-0.36	-0.17
	(0.24)	(0.21)
Number of children aged 3-5	0.01+	0.00
	(0.00)	(0.00)
% African Americans	1.12	0.07
	(0.80)	(0.69)

% teachers with BA	0.79	0.89
	(0.76)	(0.69)
% teacher assistants with college	0.86	-1.16*
	(0.52)	(0.45)
% families employed	1.74	2.07+
	(1.18)	(1.02)
Constant	-0.36	-1.64
	(2.41)	(2.10)

Standard errors in parentheses; + $p < .10$; * $p < .05$; ** $p < .01$

Figure Captions

Figure 1. CSRP Participant Flow Chart

Figure 2. Effects of CSRP treatment on Child BPI Internalizing Scores

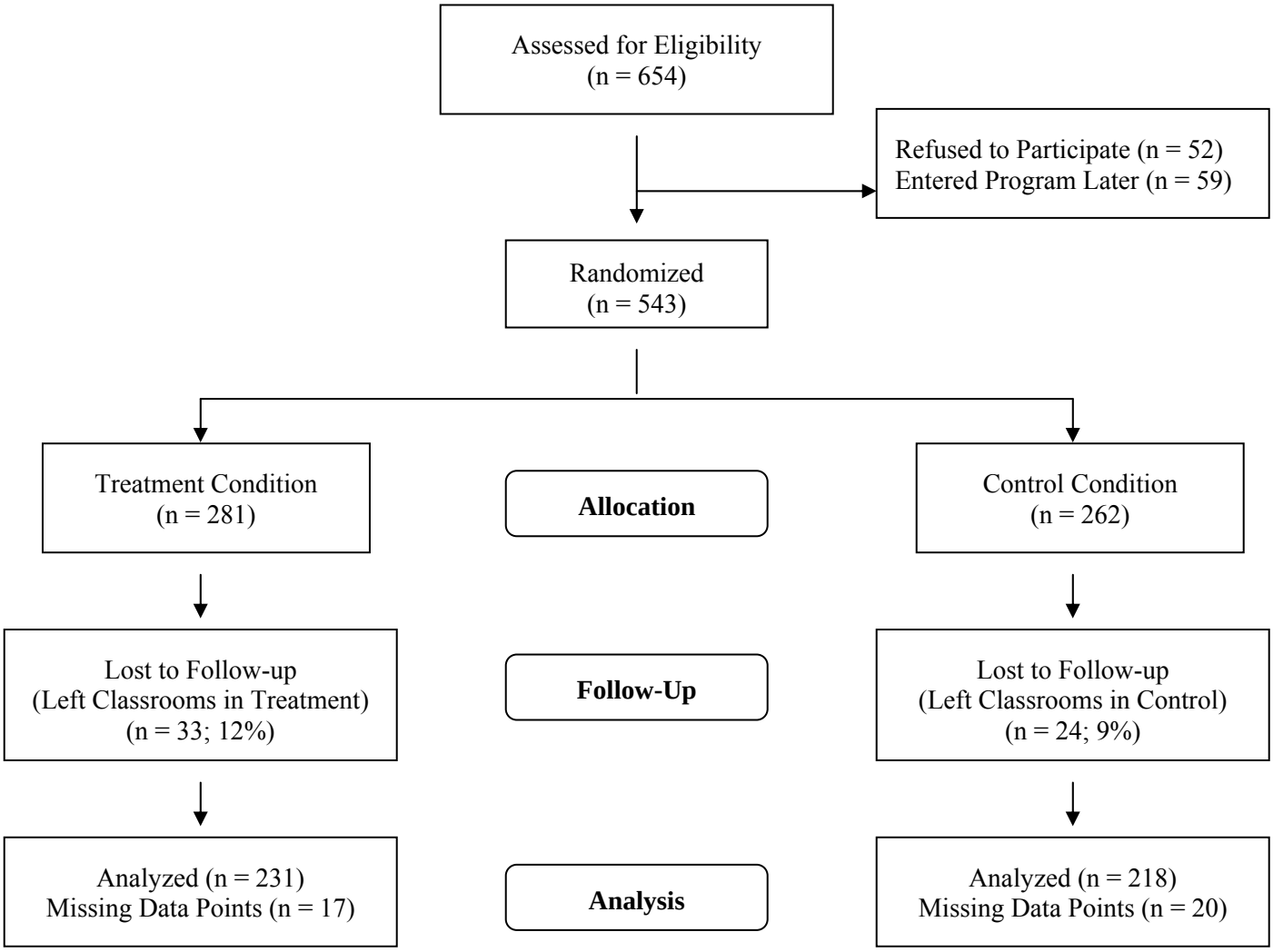
Figure 3.Effects of CSRP treatment on Child BPI Externalizing Scores

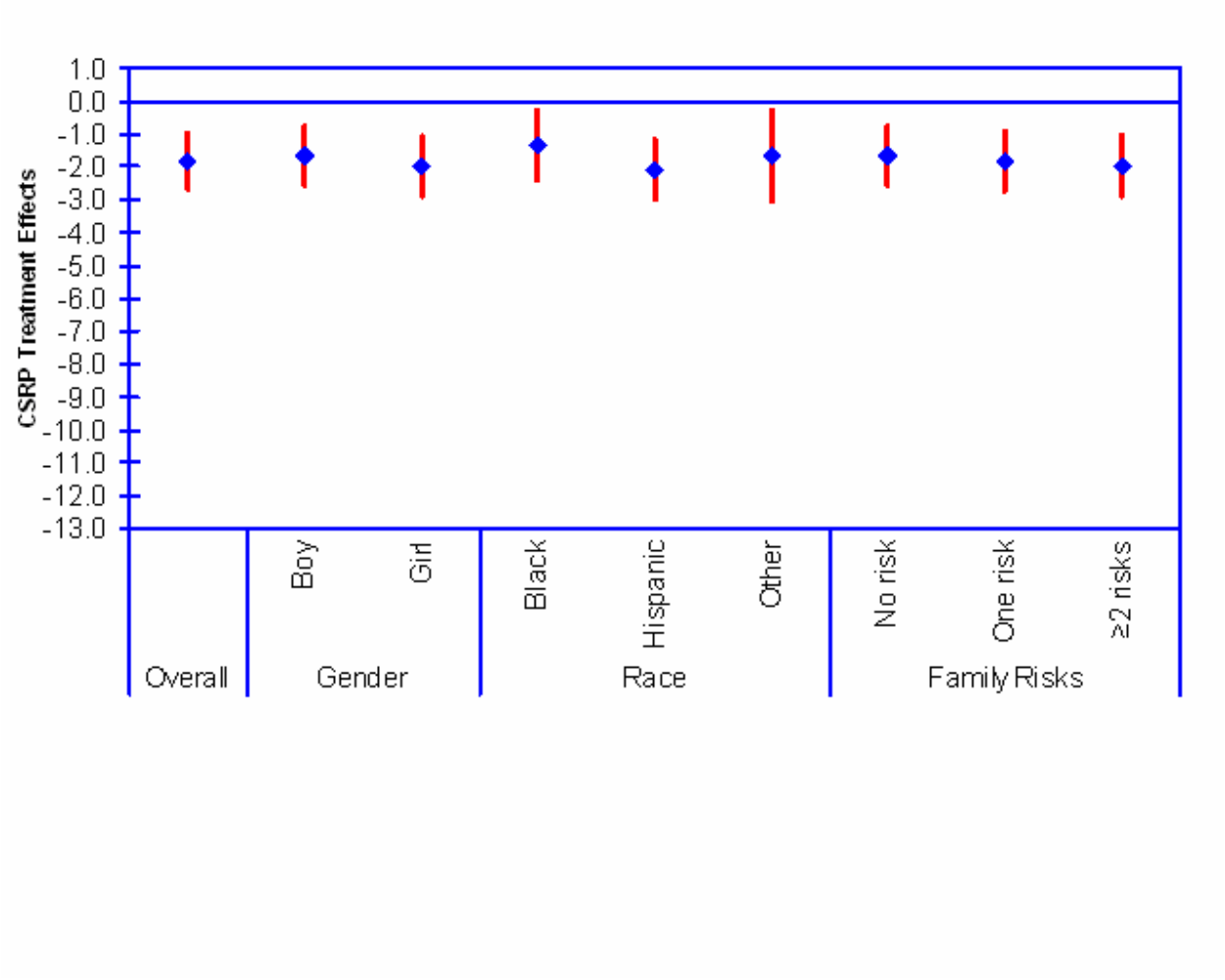
Figure 4.Effects of CSRP treatment on Child TRF Internalizing Scores

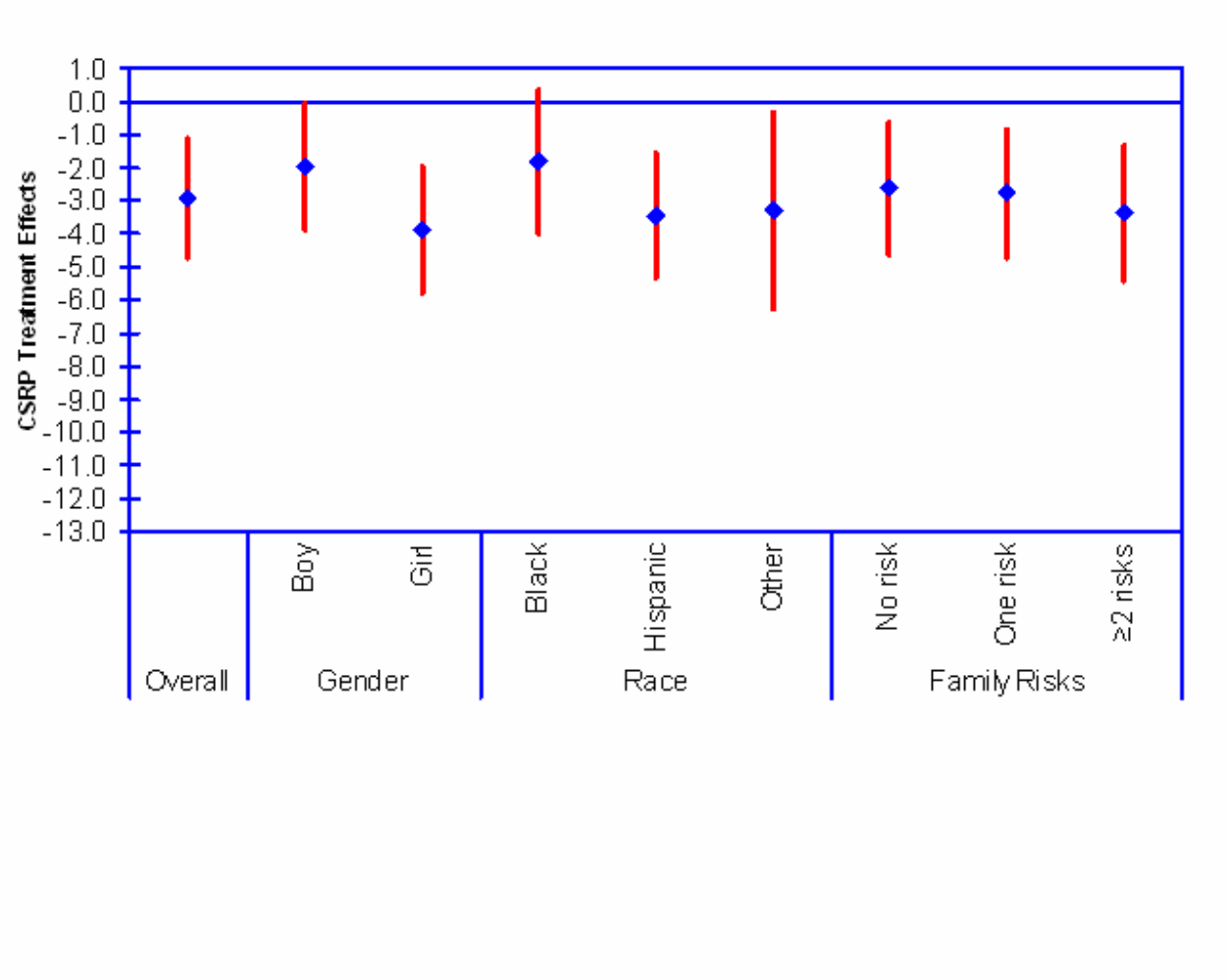
Figure 5.Effects of CSRP treatment on Child TRF Externalizing Scores

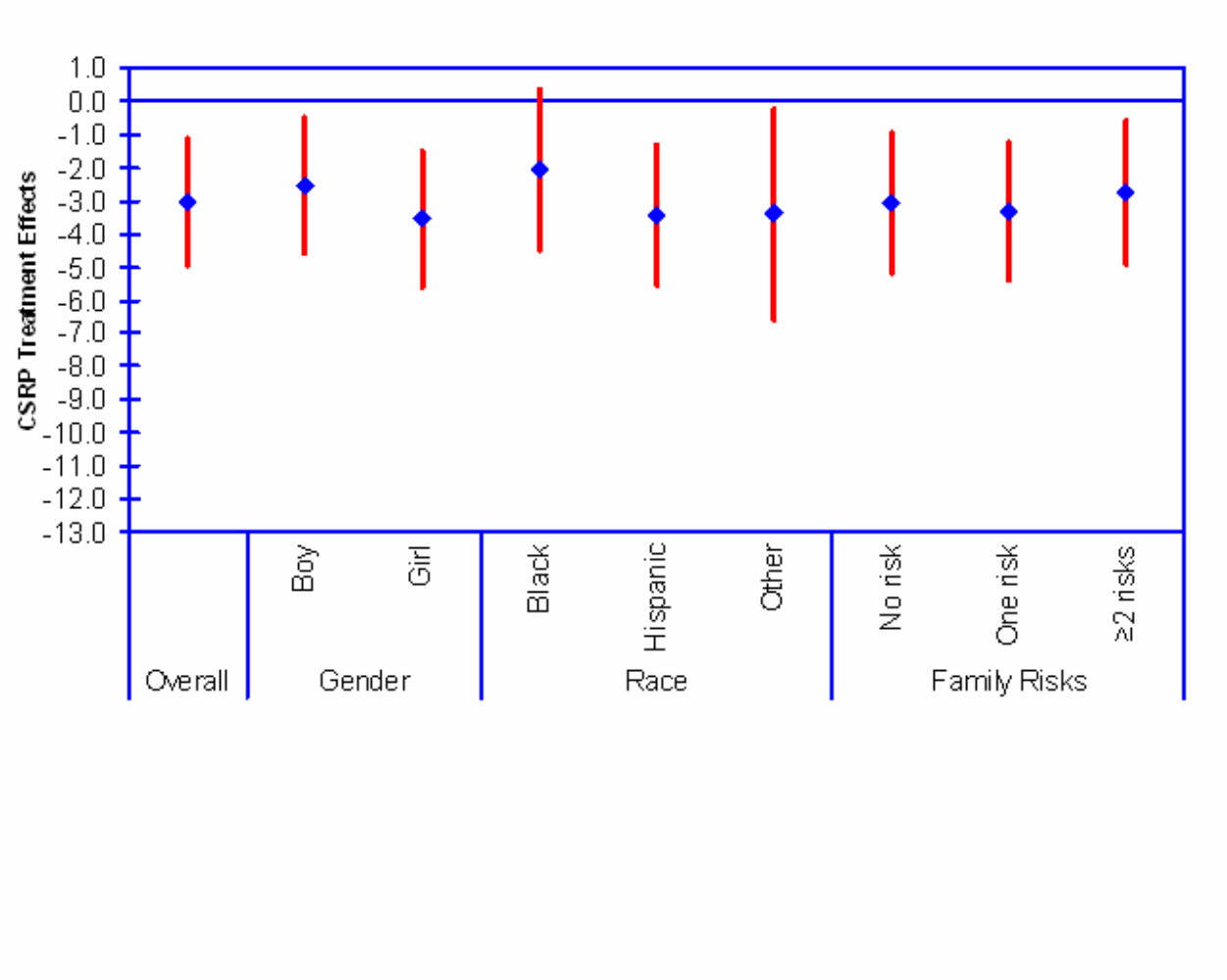
Figure 6.Effects of CSRP treatment on Child PIPPS Disruption Scores

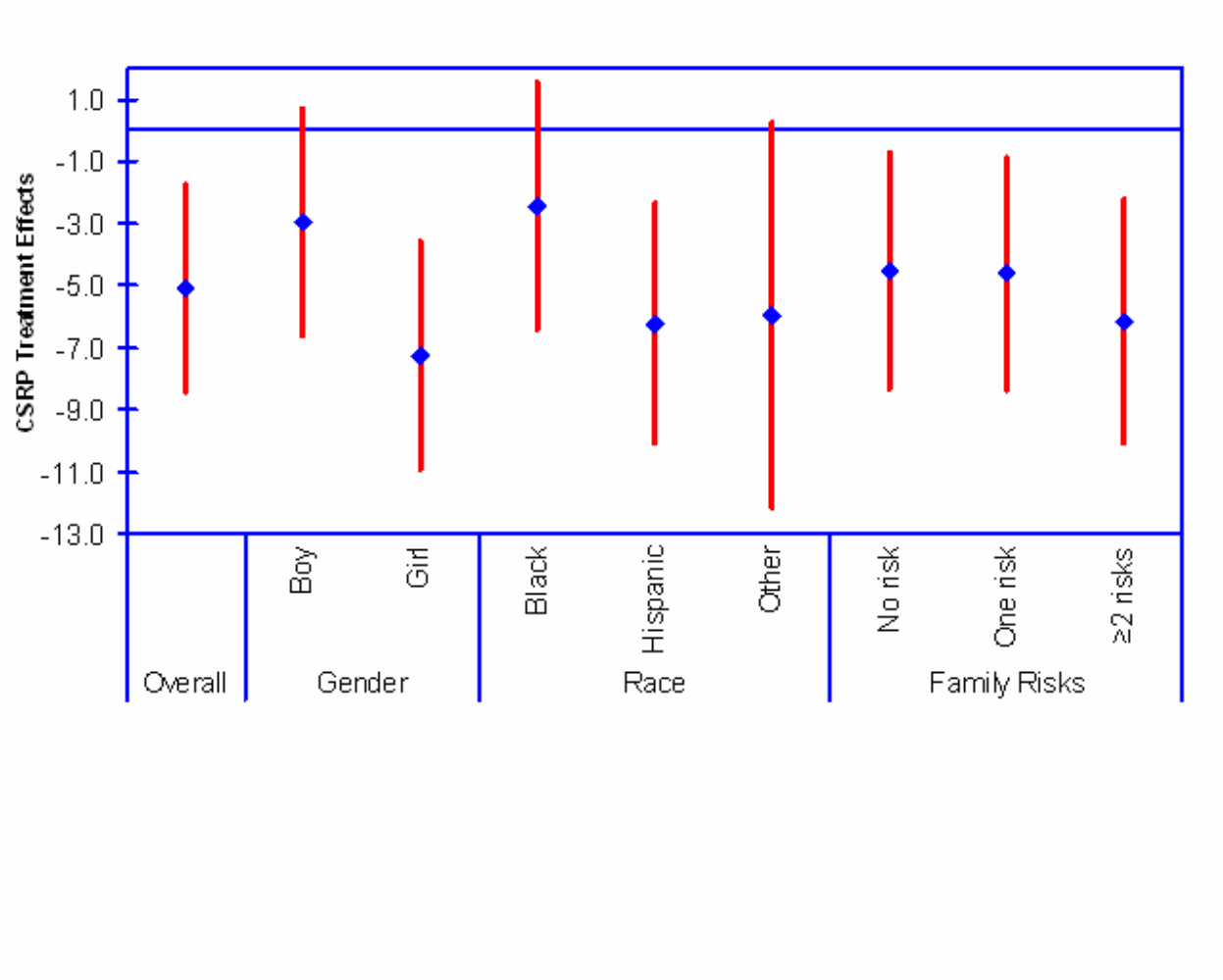
For Figures 2 – 6 See footnote 2

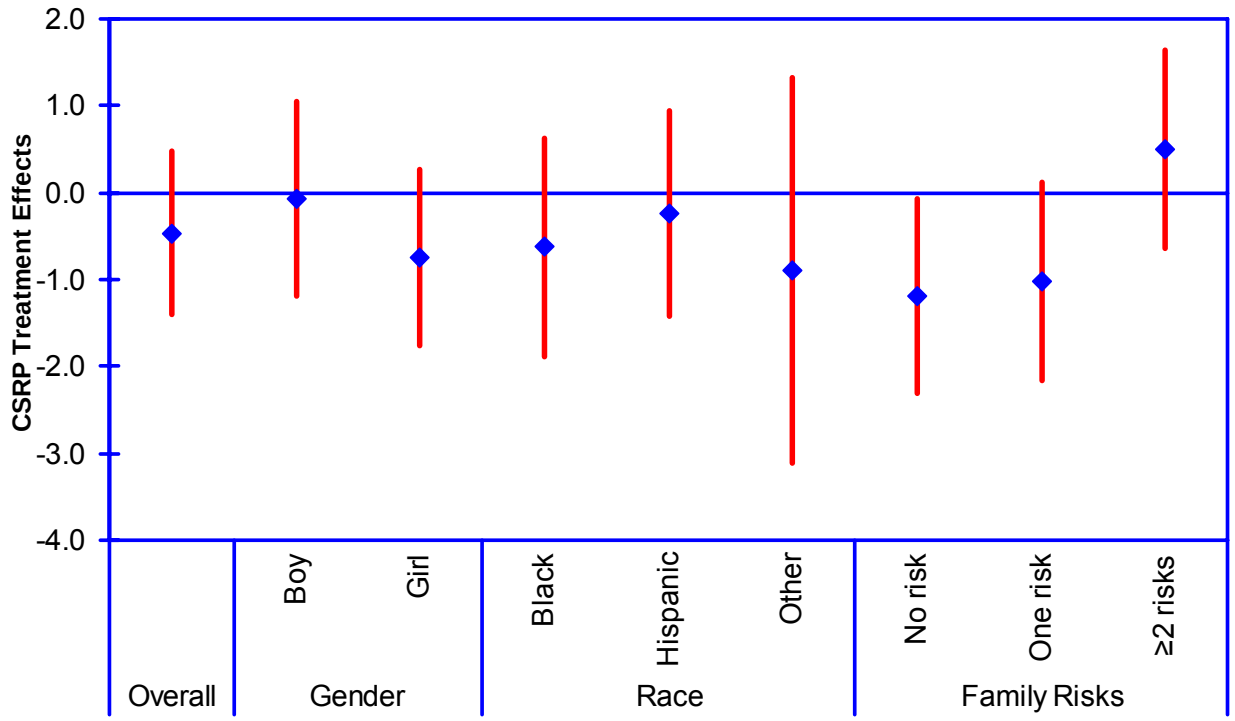

Assessed for Eligibility
(n = 654)
Refused to Participate (n = 52)
Entered Program Later (n = 59)
Randomized
(n = 543)
Treatment Condition
(n = 281)
Allocation
Control Condition
(n = 262)
Lost to Follow-up
(Left Classrooms in Treatment)
(n = 33; 12%)
Follow-Up
Lost to Follow-up
(Left Classrooms in Control)
(n = 24; 9%)
Analyzed (n = 231)
Missing Data Points (n = 17)
Analysis
Analyzed (n = 218)
Missing Data Points (n = 20)


CSRP Treatment Effects
1.0
0.0
-1.0
-2.0
-3.0
-4.0
-5.0
-6.0
-7.0
-8.0
-9.0
-10.0
-11.0
-12.0
-13.0
Overall
Boy
Girl
Gender
Black
Hispanic
Other
Race
No risk
One risk
≥2 risks
Family Risks


CSRP Treatment Effects
1.0
0.0
-1.0
-2.0
-3.0
-4.0
-5.0
-6.0
-7.0
-8.0
-9.0
-10.0
-11.0
-12.0
-13.0
Overall
Boy
Girl
Gender
Black
Hispanic
Other
Race
No risk
One risk
≥2 risks
Family Risks


CSRP Treatment Effects
1.0
0.0
-1.0
-2.0
-3.0
-4.0
-5.0
-6.0
-7.0
-8.0
-9.0
-10.0
-11.0
-12.0
-13.0
Overall
Boy
Girl
Gender
Black
Hispanic
Other
Race
No risk
One risk
≥2 risks
Family Risks


CSRP Treatment Effects
1.0
-1.0
-3.0
-5.0
-7.0
-9.0
-11.0
-13.0
Overall
Boy
Girl
Gender
Black
Hispanic
Other
Race
No risk
One risk
≥2 risks
Family Risks


2.0
1.0
0.0
-1.0
-2.0
-3.0
-4.0
CSRP Treatment Effects
Overall
Boy
Girl
Gender
Black
Hispanic
Other
Race
No risk
One risk
≥2 risks
Family Risks