Journal of Clinical Child and Adolescent Psychology
2004, Vol. 33, No. 2, 279–291

EMPIRICAL ARTICLES

Halting the Development of Conduct Problems in Head Start Children: The Effects of Parent Training

M. Jamila Reid and Carolyn Webster-Stratton
University of Washington

Nazli Baydar
University of Washington
Koc University

We examined parent and child moderators of outcome, program engagement effects, and predictors of engagement in the Incredible Years Parent Training Program. Head Start classrooms (N = 882) were randomly assigned to an intervention condition (that received the Incredible Years program) or to a control condition (that received usual Head Start services). Structural equation modeling (SEM) was used to model the effects of the training program on child outcomes. The analyses showed differential program effects depending on children's initial levels of conduct problems and mothers' initial levels of critical parenting. Children with high baseline levels of conduct problems and children of mothers with high initial levels of critical parenting benefited most from the program. Changes in children's conduct problems were also related to maternal engagement in the program and to intervention mothers' success at reducing their critical parenting.

This research was supported by the Center for Substance Abuse Prevention Predictor Variables Initiative 3 UR6 SP07960–03–1, an Administration for Children and Families Head Start Partnership grant, and Research Scientist Development Award MH00988 from the National Institute of Mental Health. The analysis was funded, in part, by the University of Washington Center for Mind, Brain, and Learning through a generous gift from the Talaris Research Institute, the family foundation of Bruce and Jolene McCaw.

We are grateful to a number of people who assisted in extensive work related to grant preparation, data collection, teacher training, and intervention implementation: Karrin Bianchi, Carmen Brower, Kate Calhoun, Ashley Carter, Kim Chou Pham, Sarah Cook, Diane Elliott, Lois Hancock, Doris Harkness, Nat Houtz, Peter Loft, Kathy Rogers, Margaret Trudeau, and Amy Vore. Finally, we are indebted to the Seattle Public School and Puget Sound Head Start staff and administrators, particularly the family service workers and teachers whose dedication and commitment to families made this program possible.

Carolyn Webster-Stratton has disclosed a potential financial conflict of interest because she disseminates these treatments and stands to gain from a favorable report. Because of this, she has voluntarily agreed to distance herself from certain critical research activities (i.e., recruiting, consenting, primary data handling, and analysis), and the University of Washington has approved these arrangements.

Requests for reprints should be sent to Carolyn Webster-Stratton, University of Washington, School of Nursing, Parenting Clinic, 1107 N.E. 45th Street, Suite 305, Seattle, WA 98105. E-mail: cws@u.washington.edu

Preventing conduct problems in young children has become a focus of early intervention efforts. Estimates show that 7% to 35% of young children meet the diagnostic criteria for oppositional defiant disorder or conduct disorder (Hawkins, Catalano, Kosterman, Abbott, & Hill, 1999) with the highest rates found in low-income populations (Webster-Stratton & Hammond, 1998). Left untreated, children with early conduct problems face increased risk of conduct disorders and school difficulties during early schooling and of school dropout, delinquency, and antisocial behaviors in adolescence and adulthood (Loeber & Farrington, 2000). Maladaptive parenting styles are the most well-researched influences on the development of conduct problems. As a group, parents of children with conduct problems are less positive, more permissive and inconsistent, and use more violent and critical discipline (Dishion, French, & Patterson, 1995; Patterson & Stouthamer-Loeber, 1984). Parent intervention is a major focus for decreasing and preventing conduct problems in childhood, and experts contend that these programs are most effective when targeted to parents of young children (Brestan & Eyberg, 1998). Research in prevention (Gross, Louis, Webster-Stratton, Garvey,

& Grady, 2003; Spoth, Redmond, & Shin, 1998; Yoshikawa, 1994) and treatment (i.e., children diagnosed with oppositional defiant disorder or conduct disorder; Brestan & Eyberg, 1998) contexts has shown that parent training diminishes harsh and inconsistent parenting and increases positive parenting (Taylor & Biglan, 1998). The extent to which parenting change leads to changes in child behavior is important in determining the effectiveness of parent training for preventing the development of conduct problems in children.

A number of factors contribute to the difficulty of measuring prevention program effects on child behavior. Analyses that aggregate information for all families (regardless of baseline problem behaviors) often show stronger effects for changes in parent behavior than for reductions in child conduct problems (August, Realmuto, Hektner, & Bloomquist, 2001; Barrera et al., 2002; Reid, Eddy, Fetrow, & Stoolmiller, 1999; Webster-Stratton, Reid, & Hammond, 2001XX). These aggregate analyses may mask important intervention effects for several reasons. First, not all parents in a targeted prevention sample exhibit maladaptive parenting; parents who have good parenting skills may not change their parenting substantially. Second, not all parents who have maladaptive parenting practices have children who display conduct problems. Lastly, even when parenting skills improve, some children's problem behaviors will persist due to other contributing factors (e.g., learning problems, temperament, deviant peer group).

For these reasons, even a parent program that reliably improves parenting may not have uniformly beneficial effects on all children. Large-scale multimodal prevention programs for elementary-school children have found differential program effectiveness depending on the level of conduct problems. Aggressive, impulsive, and hyperactive children benefited from these programs more than children with low or moderate levels of conduct problems (August et al., 2001; Barrera et al., 2002; Reid et al., 1999). Thus, an evaluation of the effectiveness of a preventive parenting program requires testing the effectiveness of the program in conjunction with the risk status of the parent and the child and the amount of change in parenting.

The Incredible Years Parent Training Program has proven effective in multiple treatment samples (Webster-Stratton, Mihalic, et al., 2001; Webster-Stratton, Reid, & Hammond, 2001XX) as well as in two randomized prevention (Webster-Stratton, 1998b; Webster-Stratton & Reid, 2001) samples in Head Start. Our prevention studies showed that when the Incredible Years Parent Training Program was offered as a selective prevention program to *all* parents enrolled in the experimental Head Start centers, there were significant and strong improvements in parent behavior, with somewhat weaker results for child behavior. These reports focused primarily on analyses examining the effects of the program regardless of initial level of parenting skill or child behavior problems. Parent and child behavior change were considered separately, so changes in child behavior could not be attributed specifically to improvements in parenting behavior. This study builds on our previous reports of this Head Start population by examining parent and child moderators of outcomes, program engagement effects, and predictors of engagement.

The following hypotheses were addressed in this study: (a) Higher levels of program engagement by mothers will predict more benefits for children; (b) parent training will be most effective (in reducing conduct problems and promoting prosocial behaviors) for children who exhibit high initial levels of conduct problems; (c) the parent training program will be most effective among the children whose mothers show highly negative parenting; (d) the improvements in children's conduct problems and prosocial behaviors will be significantly higher for children whose mothers display a substantial improvement in their parenting practices.

Methods

Participants and Procedures

This study combined assessments from three cohorts of Head Start families who participated in previous prevention studies. Two cohorts of participants were enrolled in the study in the 1993 and 1994 (Webster-Stratton & Hammond, 1998). The third cohort was enrolled in 1997 (Webster-Stratton, Mihalic, et al., 2001). Two other publications have used this same combined sample (Baydar, Reid, & Webster-Stratton, 2003; Reid, Webster-Stratton, & Beauchaine, 2001). All cohorts were studied using a quasi-experimental design wherein Head Start centers were matched on several variables (e.g., ethnicity of children, experience of teachers) and randomly assigned to either (a) an experimental condition in which parents were offered the Incredible Years Program (14 centers) or (b) a control condition consisting of the regular Head Start curriculum (9 centers). After centers were matched, two thirds of the centers were randomly assigned to receive intervention and one third were assigned to control. All centers agreed to random assignment and none refused to participate.

A detailed description of the Incredible Years Program content, training process, and integrity checks can be found in other articles (Webster-Stratton, 1998a; Webster-Stratton & Hancock, 1998). The program teaches parents to use child-directed play skills, positive and consistent discipline strategies, strategies for coping with stress, and ways to strengthen chil-

dren's social skills. The first two cohorts were offered the curriculum in weekly 2½-hr sessions for 8 to 9 weeks. The identical curriculum for the third cohort was lengthened to 12 weekly 2-hr sessions (total program content remained the same). All groups were run by a certified Parenting Clinic leader who was paired with a family service worker from the Head Start site. Control sites continued their regular Head Start curriculum.

Teachers received intervention as well as parents. In the first two cohorts, teachers received 2 days of training that exposed them to the material taught in the parent groups, including child-directed play, praise, incentives, and limit setting. In the third cohort, teachers received 6 days of training on the same topics (Webster-Stratton, 1998b; Webster-Stratton & Reid, 2003; Webster-Stratton, Mihalic, et al., 2001). The analyses presented in this article focused on the relation between parenting and child outcomes at home, rather than on child outcomes at school. Although the possibility that the teacher intervention could have had indirect effects on children's behavior at home cannot be ruled out, this generalization from school to home is unlikely. A recent study that assessed the effects of adding teacher training to parent training found no short-term added benefit of the teacher training on children's behavior at home. In the long term, teacher training did add to parent training, but only for children's school behavior (Reid, Webster-Stratton, & Hammond, 2003; Webster-Stratton, Reid, & Hammond, 2001XX).

Families in the intervention and control conditions completed identical assessments pre- and postintervention. Data reported here were obtained from parent and teacher reports and independent home observations of parent–child interactions. All data were collected during two home visits by trained staff who were blind to the family's treatment condition.

Thirty-one percent of the 882 participants were in the control condition, and the remaining 69% (N = 607) were in the intervention group. Fifty-one percent were White, 19% African American, 10% Hispanic, 8% Asian, and 12% of mixed or other races. Almost all (86%) of children were under the age of 5 and 53% were male. Eighty-four percent of families reported a gross annual income of $20,000 or less.

Eyberg Child Behavior Inventory (ECBI). The ECBI (Robinson, Eyberg, & Ross, 1980) is a 36-item inventory of conduct problems for children ages 2 to 16 years. Reliability coefficients for the ECBI scales range from 0.86 (test–retest) to 0.98 (internal consistency). The Intensity score, an indicator of the degree of conduct problems, was used.

Dyadic Parent–Child Interactive Coding System–Revised (DPICS–R). The DPICS–R (Robinson & Eyberg, 1981; Webster-Stratton, 1985c) is an observational measure developed for recording behaviors of children with conduct problems and their parents at home. This measure has been used by this research group for the past 15 years and is described in detail in prior reports (Webster-Stratton, 1998b; Webster-Stratton, Mihalic, et al., 2001). Reliability data were collected on approximately 20% of the home observations at all assessment phases. ICC reliability statistics for the DPICS–R range from .65 to .98. Mothers and children were observed interacting for 30 min at home. The DPICS–R instrument coded children's conduct problems, noncompliance, and positive behavior. The positive and externalizing or noncompliant behavior items were analyzed separately using exploratory factor analyses with varimax rotation. Items that had a loading on the dominant factor exceeding 0.3 were considered for this measure. The resulting DPICS–R Conduct Problems subscale had six items (e.g., hit, whine, yell, smart talk) and an internal reliability of 0.75. All four items (e.g., verbal or physical affection, positive affect) describing positive or prosocial child behaviors loaded on the dominant positive factor (obtained by varimax rotation) by more than 0.3. These items constituted the Positive/Prosocial behaviors scale (α = 0.52). Despite the modest internal consistency, a scale score, rather than individual items, was used for two reasons. First, the low reliability was not expected to have a substantial negative impact on the statistical models of child behavior, because in the structural equation models this measure of Positive/Prosocial behavior was used in conjunction with another assessment of prosocial behavior (see later discussion) in a measurement model that recognized and modeled the measurement error. Second, using individual DPICS–R items was not a viable option because their coarse distribution and limited range would violate normality assumptions of statistical models adopted here. The ECBI intensity scores were positively correlated with the DPICS–R Conduct Problems scores and uncorrelated with the DPICS–R Positive/Prosocial behaviors scores ($r = .20, p < .01$, and $r = -.00, p > .90$, respectively). The two DPICS–R scale scores were correlated by $r = -.14$ ($p < .01$).

Coder Impression Inventory (CII). The CII was adapted from the Oregon Social Learning Centers' Impression Inventory and describes parenting style and child affect and behavior. Coders completed the CII following the 30-min home observation. Conduct problems and Affectionate/Prosocial Behaviors scales were used in these analyses. The CII items indicating conduct problems and prosocial behaviors were analyzed separately using exploratory factor analyses with varimax rotation. Items with dominant factor loadings exceeding 0.3 were included in the resulting scales. The CII Conduct Problems scale had eight items (e.g., child showed anger, noncompliance, physically aggression)

and an internal reliability of 0.82. The CII Affectionate/Prosocial Behaviors scale had six items (e.g., child verbally or physically affectionate, engaged in interaction) and an internal reliability of 0.54. The ECBI intensity scores were positively correlated with the CII Conduct Problems scores and negatively correlated with the CII Affectionate/Prosocial Behaviors scores ($r = .23, p < .01$, and $r = -.16, p > .01$, respectively). These correlations are in magnitude to those reported by other studies that included multi-informant assessments (Feinberg, Neiderhiser, Howe, & Hetherington, 2001). It is likely that there are some aspects of child behavior, noted by observers, that are not shared by the mothers. This could be because of bias in mothers' reporting or other situational factors influencing observers more strongly than the mothers (Stoolmiller, Eddy, & Reid, 1999). In general, observational measures correlated well with each other (.68, $p < .01$, for conduct problems; .47, $p < .01$, for prosocial behaviors). The two CII subscale scores were correlated by –.35 ($p < .01$).

Definition of "indicated" mothers and children. Head Start families are often characterized as at high risk for a host of negative outcomes (including maladaptive parenting, child behavior problems, and school difficulties) due to problems associated with low socioeconomic status. Nevertheless, a substantial proportion of Head Start parents had positive and adaptive parenting practices prior to parent training. These parents were not expected to exhibit less change in their parenting in response to intervention than parents who were harsh, inconsistent, or disengaged. To identify mothers who displayed problematic parenting practices, an observational indicator was used that identified mothers who made 10 or more critical statements to children during the home observation (range 0–119, median 13). These mothers constituted 58% of total sample at baseline and are be referred to as the indicated mother sample. This cutoff, based on independent observations, ensured an indicator that would be unaffected by self-report biases. Prior research showed that frequency of criticisms discriminated between abusive and nonabusive parenting and was highly correlated with child deviant behaviors and noncompliance, $r = .49$, $p < .01$ (Webster-Stratton, 1985a). The cutoff of 10 critical statements was used in previous studies and distinguished between clinical and normative samples (Webster-Stratton, 1985b).

To identify the indicated children, baseline teacher reports of problem behaviors were used. There were several reasons for this choice. First, preintervention parent and observer reports were already included as predictors of the outcome measures in this study and no additional information could be gained from using these as a measure of indicated cases. Second, from a methodological point of view, if mother and observer measures were used to identify the indicated children, the outcome variables would have had severely truncated distributions in multivariate models for indicated and nonindicated children, violating key distributional assumptions of estimation algorithms. Third, teacher reports represented a view of how children behave compared to similar children of the same age.

Two well-validated teacher report measures were used. The first two cohorts of the sample completed the Teacher Report Form (TRF; Achenbach, 1991a, 1991b). The third cohort completed the Social Competence and Behavior Evaluation–Preschool Edition (SCBE; LaFreniere, Dumas, Capuano, & Dubeau, 1992), an 80-item rating scale developed to assess patterns of social competence, emotion regulation and expression, and adjustment difficulties in children ages 30 to 78 months. Unfortunately, no identical teacher-report measures were available for all cohorts in this study. The indicated sample represented children displaying the top 30% of problem behaviors (TRF standard *T* score of 53 or higher and SCBE standard *T* score of 47 or lower). Indicated children in these analyses had conduct problems that were, on average, one standard deviation above the normative mean. Because two different teacher measures were used, their equivalence for the study sample was investigated. All cohorts enrolled in the study had observer reports of externalizing behaviors in school (observer ratings using the Teacher Observation of Classroom Adaptation–Revised, TOCA–R; Werthamer-Larsson, Kellam, & Oveson-McGregor, 1990). The TOCA–R Poor School Readiness scale is a summary score of coders' ratings of 14 items related to child adjustment in the classroom (e.g., yells, fights, lies, breaks rules, harms others). The indicated children using the TRF had an average TOCA–R School Readiness score of 9.3, and the indicated children based on the SCBE had an average TOCA–R School Readiness score of 9.2. Among the nonindicated children based on the TRF, 79% were in the lower 70% of the TOCA–R. Among the nonindicated children based on the SCBE, 75% were in the lower 70% of the TOCA–R ratings. Based on these findings, the two externalizing measures were judged to be adequately equivalent to yield a teacher-report qualitative categorization of indicated and nonindicated children.

In this sample, 34% of families had neither an indicated mother nor child, 42% an indicated mother only, 8% an indicated child only, and 16% both an indicated mother and child.

Program benefit. For some analyses, mothers who benefited from the intervention had to be identified. To indicate benefit, the criterion of a 30% reduction in the number of maternal critical statements between pre- and postintervention assessments was used. We chose maternal critical statements because, in our past work, this variable has discriminated between

abusive and nonabusive parenting (Webster-Stratton, 1985a). Because this measure does not have established norms, we based our use of a 30% reduction on prior studies with children with conduct problems that used 30% reductions as indicators of treatment success using nonnormed behavioral observational measures (Dumas & Wahler, 1983; Patterson, Chamberlain, & Reid, 1982; Webster-Stratton & Fjone, 1989; Webster-Stratton & Hammond, 1997).

Program engagement. Program engagement was assessed by (a) the number of parent training sessions attended by the mother, (b) the percentage of homework completed, and (c) the group leader's rating of the mother's engagement in group discussion. These three measures were combined in a measurement model to represent program engagement. Correlations among these measures were as follows: sessions attended with completed homework, $r = .47, p < .01$; sessions attended with discussion participation, $r = .92, p < .01$; completed homework and discussion participation, $r = .48, p < .01$.

Program attenders. For some analyses, an indicator was constructed that distinguished mothers who attended the intervention program at a "meaningful" level and mothers who did not. It is evident that program effects would differ among the children of the mothers who attended at a level that was meaningful versus those whose mothers did not. From a substantive perspective, three sessions seemed appropriate because those mothers who attended at least three sessions displayed their commitment to make an effort to benefit from the program. Admittedly, that commitment may be an indicator of those attributes that make it possible for the mothers to benefit from the program. In other words, some characteristics that predict program attendance may also predict program benefits. For this reason, it is important to investigate the predictors of attendance. Previous work indicated that program attendance is not associated with those attributes of the mothers that are commonly suspected to facilitate program benefits. For example, mothers with poor parenting skills and those with mental health problems were just as likely and, in some cases, more likely to attend the program than others (Baydar et al., 2003). The validity of the three-session cutoff was also investigated. For example, more than 54% of mothers who attended three or more sessions displayed a reduction of critical statements by 30% or more, compared to 38% who attended zero to two sessions, $\chi^2(1) = 8.7, p < .01$. Based on these considerations and the face validity of the concept of "attendance," a cutoff for meaningful attendance to parent training was defined (admittedly in an ad hoc manner) as three or more sessions. Mothers who attended at least three sessions are referred to as *attenders*. It should be noted that 60% of all mothers in the intervention condition were attenders, and, on average, attenders were present at 7.7 training sessions. Thus benefits experienced by most attenders resulted from many more than the three sessions.

Methods of Analysis

Program effectiveness was modeled using structural equation modeling (SEM). SEM provided four advantages for analyses in comparison to other methods of analysis such as analyses of variance. First, SEM could incorporate measurement submodels, allowing the joint modeling of the measures of child behaviors from two different informants. In doing so, SEM allowed the specification of latent child behavior constructs that (under certain assumptions) did not have any measurement error, an issue that gained relevance in view of some of the measures with moderate internal reliabilities. The latent constructs indicated by several assessments always have a reliability that exceeds the reliability of their best (most reliable) indicator (Hancock & Meuller, 2001). In the models presented here, child conduct problems and prosocial behaviors were assessed by three and two different measures, respectively. Second, SEM could represent measurement models, program impacts, program engagement effects, and sample attrition models all at once. Third, SEM provided the ability to test the equality of program effects across different subgroups of the participants, such as indicated mothers or children, children of attenders, and children of mothers who benefited substantially from parent training. Finally, SEM allowed the inclusion in the statistical modeling of the group members who could not be assessed postprogram. This was especially advantageous for the assessment of the effectiveness of a voluntary intervention because those families who did not participate in the postprogram assessments (i.e., the attrition group) may have been a selective group. Twenty percent of the control and 24% of the intervention participants were in the attrition group.

For the three indicators of child conduct problems available at pre- and postintervention, a measurement model with two latent constructs was specified for each time point (see Figure 1). An overall conduct problems construct at each time point was indicated by all available measures (i.e., maternal and observer reports) at that time. For example, the overall conduct problems construct at preintervention represented the child's negative or antisocial behaviors that were readily agreed on by the mother and the observer at preintervention assessments. In addition, an observer-specific construct was specified. This latter construct represented those aspects of a child's conduct problems that were noted by the objective and subjective observer reports but not corroborated by (i.e., uncorrelated with) the maternal report. Thus, the variance in the observer re-

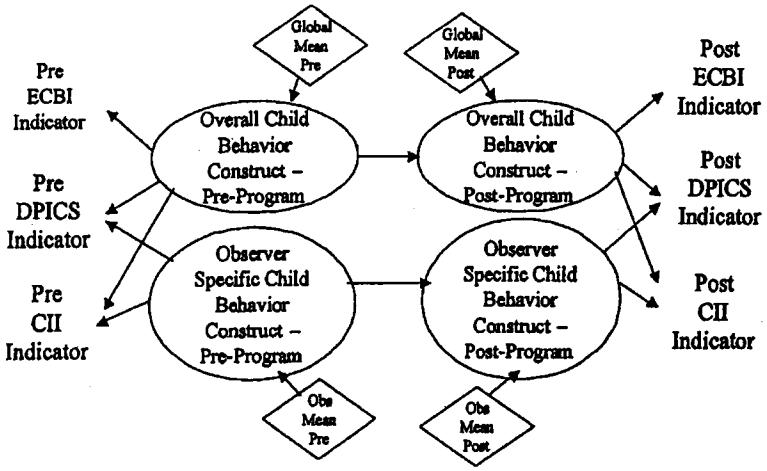

Figure 1. Model structure for estimating the effects of the parent training intervention on children's outcomes (measurement model and Model II).

ports was partitioned at each time point to a component that was correlated with maternal reports and a component that was not (see also Baydar et al., 2003; Feinberg et al., 2001). The observer-specific construct (i.e., those aspects of the observer reports that were uncorrelated with maternal reports) embodied three interpretations. First, it represented those aspects of conduct problems that were unreported by mothers (possibly due to biases in maternal reports) but noted by observers. It was this interpretation of the observer-specific construct that was substantively most compelling and provided a rationale for its inclusion in the structural equation models. Specifically, there is reason to believe that maternal reports may not be particularly sensitive to the changing nature of child behavior. Mothers do not evaluate their children's behaviors on a daily basis. Instead, they form overall (robust) opinions that may be valid when the child behavior is stable but may be less sensitive when measuring intervention effects because of their robust nature. In that case, observer reports may be more sensitive to detect change, because they are formed on direct observations and not long-term opinions. The second and third interpretations of the observer-specific constructs were not of substantive interest. The second interpretation was that the observer-specific construct represented those aspects of conduct problems that were specific to a particular observation session. Third, it represented other factors that were shared across observer reports but not between maternal and observer reports (e.g., the observer bias).

The measurement model for the three indicators of conduct problems for each of the pre- and postintervention assessments was fitted to the data from the control group to assess the fit of the measurement model displayed in Figure 1. The two-construct model fit the data satisfactorily. Furthermore, model testing showed that the loadings of the pre- and postintervention latent constructs on their respective observed indicators could be equated, comparative $\chi^2(3) = 0.3$, $p = .95$. This latter model fit the observed data very well, supporting the two-construct model structure with measurement models for pre- and postintervention that had equal loadings, $\chi^2(5) = 2.3$, $p = .81$, comparative fit index (CFI) = 1.00, root mean squared error of approximation (RMSEA) =. 00. An alternative model with a single construct could not represent the data adequately, $\chi^2(9) = 154.6$, $p = .00$, CFI = .58, RMSEA = .29 (see Figure 1).

There were two observer reports and no maternal reports of child's prosocial behaviors. Therefore, the measurement model for this outcome was substantially simpler, with a single latent construct that was indicated by the two observer assessments. This model could be identified following the standard specification for this type of structural equation model (Arbuckle, 1997). Furthermore, model testing indicated that the loadings of the pre- and postintervention latent constructs on their respective observed indicators could be equated, comparative $\chi^2(1) = 0.4$, $p = .50$. This measurement model for the prosocial behaviors fit the pre and post data from the control group satisfactorily, $\chi^2(2) = 5.7$, $p = .06$, CFI = .97, RMSEA = .09, indicating the acceptability of the constraints imposed on the structure of the measurement model.

The structural model that represented the process of change in child behaviors followed the standard specification suggested by Arbuckle (1997) and had three important attributes. First, the model included multiple groups of the sample, enabling the testing of intervention effects in different groups of program participants. For example, eight subgroups were considered in some models, based on indicated status, attendance, and intervention or control group. These groups are listed later in this article. Second, the structural equation models presented here included a model of the means

and intercepts of the latent constructs (as well as observed indicators). The estimated means and intercepts assessed the following: (a) whether at the initial assessment groups of children (defined on the basis of intervention status, attendance, and indicated status) had comparable levels of conduct problems or prosocial behaviors and (b) whether at the postintervention assessment levels of conduct problems and prosocial behaviors in the intervention groups (at indicated or nonindicated status) were significantly different from the control group. This latter estimate was a key in demonstrating the effectiveness of the parent training program for modifying child behaviors.

Three separate models were presented for each of the two outcomes of interest (i.e., child conduct problems and prosocial behaviors). Model I assessed whether program engagement (defined by three indicators of engagement) predicted improvements in children's behaviors, accounting for preintervention child behavior. The structure of Model I is presented in Figure 2. Model II considered the variability of program effectiveness. The following eight groups were modeled in Model III: (a) control, nonindicated; (b) control, indicated; (c) intervention, indicated attenders; (d) intervention, nonindicated attenders; (e) intervention, nonindicated nonattenders; (f) intervention, indicated nonattenders; (g) nonindicated attrition; and (h) indicated attrition group. The groups were represented with the model structure presented in Figure 1. At the final step, intervention children whose mothers showed 30% improvement in parenting were compared to intervention children whose mothers did not show these improvements (Model III).

Results

Based on the definition of the indicated mothers (more than 10 critical statements during home observation) and children (top 30% of behavior problems on teacher reports) provided earlier, 58% of the mothers and 28% of the children in the sample were indicated. Mothers of indicated and nonindicated children had similar rates of program attendance (61% and 61%, respectively), $\chi^2(1) = 0.0$, $p > .10$. In addition, indicated and nonindicated mothers had similar rates of program attendance (61% and 64%, respectively), $\chi^2(1) = 0.4$, $p > .10$.

Table 1 (Model I) presents the results of the analyses that investigated whether engagement in the intervention program influenced child outcomes. Program engagement reduced conduct problems of children as detected by the observers but not the mothers. One standard deviation increase in program engagement predicted 17% of a standard deviation decline in the observer-specific construct representing postprogram conduct problems. The impact of program engagement on prosocial behaviors was positive, and the effect size was very similar to that of conduct problems. Model I also yielded the effects of preprogram child behaviors on program engagement. The shared opinions of mothers and observers regarding conduct problems significantly predicted program engagement, indicating that mothers whose children exhibited more conduct problems were more likely to become engaged in the program. A standard deviation of increase in conduct problems predicted more than a quarter of a standard deviation of increase in program engagement. Prosocial behaviors did not predict mothers' engagement.

As stated previously, intervention effects were expected to vary, depending on the extent to which the mothers and the children needed an intervention. The multivariate models presented in Table 2 addressed this issue. Models IIa and IIb were estimated for children grouped according to intervention or control status, program attendance, and child (Model IIa) or mother (Model IIb) indicated status. The results of these models presented in Table 2 only include the findings regarding the groups of indicated children and mothers. Parallel results regarding the nonindicated groups are available on request. The estimated program effects displayed in Table 2 are based on the intercept esti-

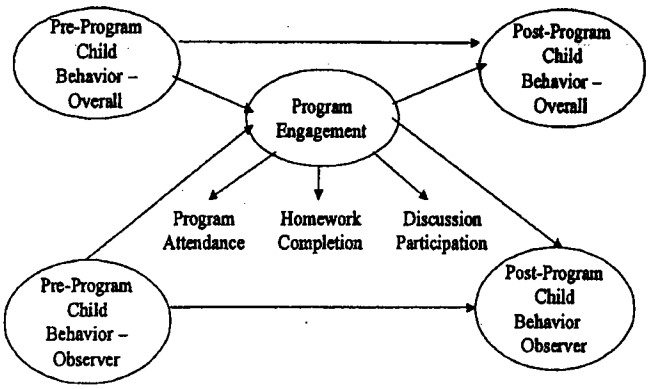

Figure 2. Model structure for the effects of program engagement (Model I).

Table 1. *Results of Model I on Program Engagement (Standardized Parameter Estimates in Parentheses)*

	Estimated Effects (Standardized Effects)		
	Conduct Problems[a]		
	Overall Construct	Observer Specific Construct	Prosocial Behaviors[b,c]
Effects of program engagement on postintervention child behaviors	−.001 (−.012)	−.044** (−.166)	.548** (.164)
Effects of preintervention child behaviors on program engagement	3.502** (.274)	.157 (.040)	−.020 (−.064)

[a]$\chi^2 = 39.2$. $df = 27$. $p = .06$. [b]$\chi^2 = 22.0$. $df = 14$. $p = .08$. [c]This is the only latent construct estimated for prosocial behaviors. It includes only observer measures and no parent report.
*$p < .05$. **$p < .01$.

mates of the latent child behavior constructs. Because these latent constructs are scaled relative to the control group and their measurement scales are arbitrary, and because standardized intercept estimates cannot be computed, it is difficult to compare the effect sizes across different models. To provide some estimates that have comparable scales, the estimated intercepts were expressed as a percentage of the estimated standard deviation of the corresponding latent construct for the control group at preintervention. These percentage effect sizes (listed in the table in parentheses) provide a means of comparing the magnitude of the statistically significant program effects.

Model IIa (Table 2, first panel) estimated the effectiveness of the training program for the indicated children (children with teacher reports of conduct problems in the upper 30th percentile) compared to the nonindicated children. At pretraining, indicated children of attenders showed significantly more conduct problems at home on both the overall and the observer specific constructs, compared to the nonindicated controls (estimated relative latent means of .161 and .378, respectively). These children also exhibited significantly fewer home prosocial behaviors compared to the nonindicated controls (–3.637). The indicated children of nonattenders did not differ significantly from nonindicated controls in their conduct problems (latent means of –.023 and .189); however, they had significantly lower home prosocial behaviors (–5.311). Note that the indicated children were defined by their teacher reports. Thus, if children were identified as having behavior problems at school *and* they were also having problems at home, their mothers were more likely to come to the parent training. Indicated children of the control mothers had higher levels of conduct problems than their nonindicated counterparts, as represented by the positive and significant estimated mean of the overall construct (.066); however, they did not differ in their prosocial behaviors (.075).

Model IIa also yielded the estimates of the posttraining levels of conduct problems and prosocial behaviors separately for indicated children of attenders, nonattenders and controls, as compared to the nonindicated controls. At posttraining, indicated children of attenders did not differ significantly from nonindicated controls in their conduct problems, although they had high levels of conduct problems at preintervention. At postintervention, indicated children of nonattenders again did not differ from nonindicated controls in their conduct problems, similar to preintervention. Thus, the intervention served to normalize the conduct problems for children who were identified by both teachers and parents as having difficulties because those parents were more likely to attend the intervention program. The conduct problems of children who were identified by teachers as having difficulty but whose mothers did not attend the training program (by choice or due to random assignment to the control group) escalated over the treatment period. At posttraining, the prosocial behaviors of indicated children significantly exceeded that of the nonindicated controls (estimated latent intercept of 3.488) by about 29% of a standard deviation unit. Thus, parent training was effective in reducing conduct problems to normative levels and in promoting the growth of prosocial behaviors. The trends predicted by Model IIa for conduct problems and prosocial behaviors are depicted in Figures 3 and 4, respectively.

Model IIb (Table 2, second panel) estimated the effectiveness of the training program for the children of indicated mothers by attendance status. Children of indicated attenders showed significantly higher levels of conduct problems at baseline than nonindicated controls, on overall (.095) and observer specific (.832) constructs. However, the children of indicated nonattenders did not significantly differ from the nonindicated controls on the overall construct representing conduct problems (–.024), but only on the observer specific construct (.473). What distinguished the attenders from nonattenders among indicated mothers was their perception of their children's behaviors. In summary, highly critical mothers were more likely to attend parent training if they perceived their children had conduct problems. Similar results were obtained

Table 2. *Results of SEMs Regarding the Effectiveness of the Intervention Program*

	Estimated Intercept		
	Conduct Problems		Prosocial Behaviors
	Overall Behavior Construct	Observer Specific Construct	Observer Specific Construct
Model IIa: Effectiveness of the parent training program for indicated and nonindicated children relative to those in the nonindicated control group			
Goodness-of-fit statistics	$\chi^2(74) = 89.3, p = .11$ CFI = .99. RMSEA = .02		$\chi^2(32) = 36.2, p = .28$ CFI = .99, RMSEA = .01
Differences at baseline: Estimated mean for the pretraining latent construct (nonindicated control group set to zero)			
Indicated children of attenders	.161**	.378**	–3.637*
Indicated children of nonattenders	–.023	.189	–5.311*
Indicated children, control moms	.066*	–.054	.075
Program effects: Estimated intercept for the posttraining latent construct (nonindicated control group set to zero)			
Indicated children of attenders	–.059	.067	3.488* (29.1%)[a]
Indicated children of nonattenders	–.009	.167	–1.417
Indicated children, control moms	–.137	.070	.084
Model IIb: Effectiveness of the parent training program for the children of indicated and nonindicated mothers relative to those in the nonindicated control group			
Goodness-of-fit statistics	$\chi^2(69) = 70.9, p = .42$ CFI = 1.00. RMSEA = .01		$\chi^2(31) = 45.5, p = .05$ CFI = .99. RMSEA = .02
Differences at baseline: Estimated mean for the pretraining latent construct (nonindicated control group set to zero)			
Children of indicated attenders	.095**	.832**	–3.817**
Children of indicated nonattenders	–.024	.473**	–3.175
Children of indicated control moms	.044*	.478**	–.222
Program effects: Estimated intercept for the posttraining latent construct (nonindicated control group set to zero)			
Children of indicated attenders	.014	–.088	2.699* (23.8%)
Children of indicated non attenders	–.006	.209	–3.697* (–32.5%)
Children of indicated control moms	–.014	.286* (35.0%)	–1.662 (–14.6%)
Model III: Relative effectiveness of the parent training program for the children whose mothers benefited from the training program and those who did not benefit			
Goodness of fit statistics	$\chi^2(16) = 21.0, p = .18$ CFI = .99, RMSEA = .03		$\chi^2(8) = 12.0, p = .15$ CFI = .99, RMSEA = .04
Estimated intercept of the posttraining latent construct for the mothers who benefited from the program (intercept for those who did not benefit set to zero)[b]	–.009	–.486** (–63.5%)	3.149* (29.4%)

Note: SEMs = structural equation models; CFI = comparative fid index; RMSEA = root mean squared error of approximation.
[a]Percentages express the estimated intercept as a percentage of the estimated standard deviation of the corresponding latent construct for the control group at preintervention. These percentage effect sizes were provided only for intercepts that were statistically significant from zero.
*$p < .05$ (one-tail test). **$p < .01$ (one-tail test).

for prosocial behaviors. Indicated attenders had children who were significantly less prosocial at baseline as compared to nonindicated controls (–3.817). The children of indicated nonattenders did not differ from the nonindicated controls in their prosocial behaviors (–3.175).

The posttraining findings pertaining to the children of indicated mothers had patterns that were largely similar to those for indicated children discussed previously. At posttraining, children of indicated attenders did not differ significantly from nonindicated controls in their conduct problems (latent intercept estimates of .014, and –.088). The children of indicated controls, however, had escalating conduct problems as detected by the observer-specific construct (.286). At posttraining, the prosocial behaviors of children of indicated attenders exceeded that of the nonindicated controls (2.699). The prosocial behaviors of children of indicated nonattenders, however, significantly diminished (–3.697). These findings showed that training critical mothers halted the progression of conduct problems

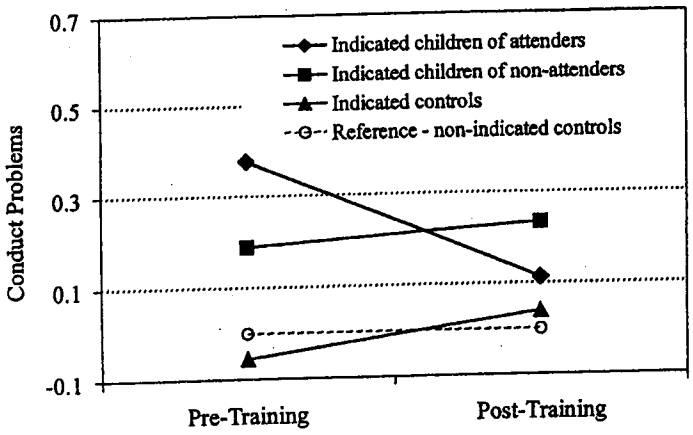

Figure 3. Changes in conduct problems of indicated children.

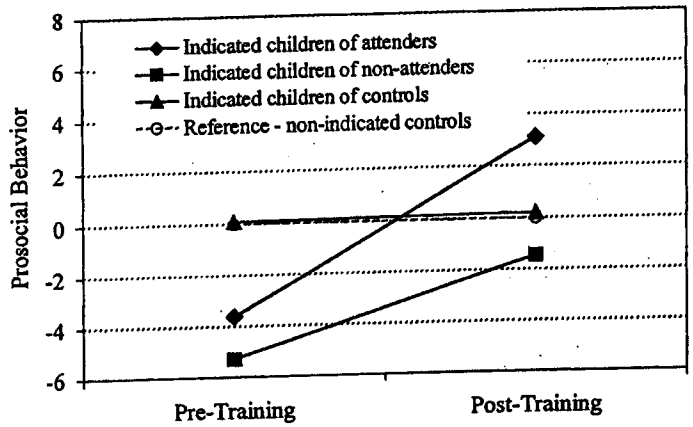

Figure 4. Changes in prosocial behaviors of indicated children.

and actively promoted the growth of the prosocial behaviors. Children of highly critical mothers who did not attend parent training became less prosocial, probably in response to critical parenting. The trends predicted by Model IIb for conduct problems and prosocial behaviors are depicted in Figures 5 and 6.

Model III (Table 2, last panel) assessed whether improvements in maternal behaviors were associated with improvements in child behaviors. Two groups of intervention children were compared: children whose mothers showed a 30% decline in their critical statements (baseline critical statements for this group per 30 min: $M = 23.89$, $SD = 18.$) and those whose mothers did not show this decline (baseline critical statements for this group: $M = 13.76$, $SD = 14.04$). The preintervention levels of conduct problems of these two groups did not differ. Children whose mothers benefited from the training (by reducing their critical statements) had a significant and substantial decline in conduct problems (by two thirds of a standard deviation) represented by the observer-specific construct (estimated intercept of –.486). These children also had a significant increase in their prosocial behaviors (one third of a standard deviation; estimated intercept of 3.149).

Discussion

Analyses presented here highlight differential effects of the Incredible Years Parent Training depending on initial levels of conduct problems and critical parenting. Changes in conduct problems were also related to maternal engagement in intervention and to whether mothers receiving intervention reduced their critical parenting. These differential program results support other preventive trials (August et al., 2001; Barrera et al., 2002; Kellam, Ling, Merisca, Brown, & Ialongo, 1998; Reid et al., 1999) suggesting that analyses of variability of effectiveness are crucial when evaluating prevention programs delivered to large heterogeneous high-risk samples.

On average, mothers who perceived their children as most problematic were more likely to participate in the program. Participation in the parent training was associated with positive child outcomes at home as

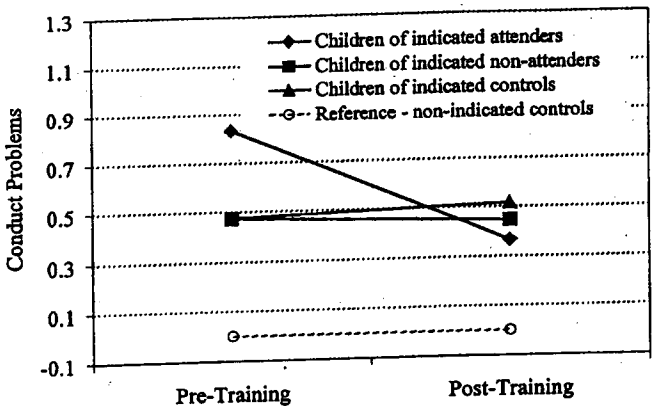

Figure 5. Changes in conduct problems of children of indicated mothers.

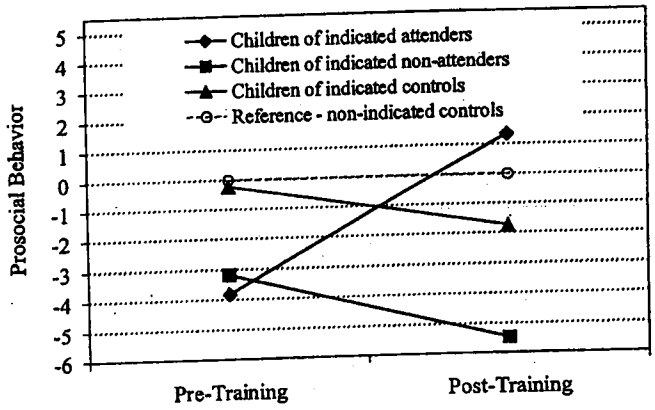

Figure 6. Changes in prosocial behaviors of children of indicated mothers.

measured by increases in child prosocial behaviors and decreases in child conduct problems. Furthermore, the analyses also linked changes in parenting behaviors to reductions in conduct problems and increases in prosocial behaviors, reasserting that parent interventions are, indeed, indicated for high-risk children.

Examination of attendance patterns provides valuable information about who becomes engaged in and who benefits from preventive parent training. For both indicated mothers and indicated children, mothers who did not attend the program, for the most part, were those whose children were not displaying high levels of conduct problems at home. Thus the mothers may not have perceived a need for a training that focused on improving child behaviors. However, teacher reports showed that these children were presenting somewhat elevated conduct problems at school. Over the 6-month period between assessments, when mothers did not attend the parent training, indicated children and children whose mothers were highly critical at home showed diminishing prosocial behaviors at home. Thus, independent observations of mothers' critical statements and teacher reports of child problems in the classroom were good indicators of risk status for undesirable changes in child home behaviors.

For the indicated samples, the parenting program acted to return high levels of conduct problems to normative levels. When indicated mothers attended training, their children's posttreatment conduct problems did not differ significantly from children of nonindicated control group mothers. Likewise, when mothers of indicated children attended the parent groups, their children did not differ significantly from nonindicated control children at the postassessment.

These findings are encouraging, because it seems that more than two thirds of the highest risk indicated mothers were motivated and interested in receiving the parent training program. Moreover, their children benefited substantially from the intervention. This information also highlights a need to focus recruitment efforts on the small group of high-risk mothers and children who may not perceive a need for the parenting program (about one third of the indicated group), even when independent home observations and teacher reports indicate that a problem exists. Teacher reports seemed to accurately identify children whose behavior was at risk. In light of this, teachers would seem to be key informants to help

identify higher risk children and may play a key role in enlisting these children's parents in parent training to prevent later problems.

These analyses also indicate that the short-term observational measures were more sensitive to program effects than mother-report measures. As noted previously, it may be that mothers reported on child behavior based on their overall (robust) opinions that may be valid when the child behavior is stable but may be less sensitive at detecting intervention effects. Without the independent observations, some intervention effects would have been overlooked. Although it is true that observations are obtained at great cost, based on our findings, it would be short-sighted to give up the collection of such data. It should be noted that internal reliability is modest for the positive child behavior scales of the two observational measures. It is possible that positive behaviors are more difficult to observe reliably during relatively short observational periods (30 min) than negative behaviors. It is also possible that positive behaviors are more difficult to clearly define or more varied than negative behaviors. The relatively low reliabilities of the positive behavior assessments make it even more important that analyses use those methods, such as SEM, that can recognize measurement errors and integrate information across informants and methods.

These results also emphasize the importance of including a control group and looking at indicated subgroups within the control and intervention conditions. The comparison with the indicated nonattenders and indicated controls revealed the stable or escalating negative trajectories that occurred when those who needed intervention did not receive it.

In summary, these analyses provide important information about the effects of the Incredible Years Parent Training Program when used as a preventive intervention in a sample of low-income Head Start families. Program effects on conduct problems and prosocial behaviors were found for children of highly critical mothers and for children with elevated initial levels of conduct problems in the classroom. In most cases, it was encouraging that the most critical mothers and the mothers of indicated children chose to attend the parent program. Also encouraging, but not surprising, was that engagement in the program was associated with beneficial change in children's conduct problems and prosocial behaviors and that these effects were directly related to reductions in maladaptive parenting. These results confirm findings from other recent prevention trials (August et al., 2001; Barrera et al., 2002; Reid et al., 1999) that suggest that evaluation of program effects must take into consideration initial levels of both parent and child behaviors as well as the extent to which the parents and children have attended the intervention program at substantively meaningful levels. Only then can we answer questions about whether prevention programs work and for whom.

References

Achenbach, T. M. (1991a). *Manual for the Child Behavior Checklist/4–18 and 1991 Profile.* Burlington: University of Vermont, Department of Psychiatry.

Achenbach, T. M. (1991b). *Manual for the Teacher Report Form and 1991 Profile.* Burlington: University of Vermont, Department of Psychiatry.

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.

Barrera, M., Biglan, A., Taylor, T. K., Gunn, B., Smolkowski, K., & C. B. (2002). Early elementary school intervention to reduce conduct problems: A randomized trial with Hispanic and non-Hispanic children. *Prevention Science, 3,* 83–94.

Baydar, N., Reid, M. J., & Webster-Stratton, C. (2003). The role of mental health factors and program engagement in the effectiveness of a preventive parenting program for Head Start mothers. *Child Development, 74,* 1433–1453.

Brestan, E. V., & Eyberg, S. M. (1998). Effective psychosocial treatments of conduct-disordered children and adolescents: 29 years, 82 studies, and 5,272 kids. *Journal of Clinical Child Psychology, 27,* 180–189.

Dishion, T. J., French, D. C., & Patterson, G. R. (1995). The development and ecology of antisocial behavior. In D. Cicchetti & D. J. Cohen (Eds.), *Developmental psychopathology: Vol. 2. Risk disorder and adaptation* (pp. 421–471). New York: Wiley.

Dumas, J. E., & Wahler, R. G. (1983). Predictors of treatment outcome in parent training: Mother insularity and socioeconomic disadvantage. *Behavioral Assessment, 5,* 301–313.

Feinberg, M., Neiderhiser, J., Howe, G., & Hetherington, E. M. (2001). Adolescent, parent, and observer perceptions of parenting: Genetic and environmental influences on shared and distinct perceptions. *Child Development, 72,* 1266–1284.

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

Hancock, G. R., & Meuller, R. O. (2001). Rethinking construct reliability within latent variable systems. In R. Crude, S. duToit, & D. Sorbom (Eds.), *Structural equation modeling: Present and future* (pp. 195–216). Lincolnwood, IL: Scientific Software International.

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

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

LaFreniere, P. J., Dumas, J. E., Capuano, F., & Dubeau, D. (1992). Development and validation of the Preschool Socioaffective Profile. *Psychological Assessment, 4,* 442–450.

Loeber, R., & Farrington, D. P. (2000). Young children who commit crime: Epidemiology, developmental origins, risk factors, early interventions, and policy implications. *Developmental Psychopathology, 12,* 737–762.

Patterson, G. R., Chamberlain, P., & Reid, J. B. (1982). A comparative evaluation of a parent training program. *Behavior Therapy, 13,* 638–650.

Patterson, G. R., & Stouthamer-Loeber, M. (1984). The correlation of family management practices and delinquency. *Child Development, 55,* 1299–1307.

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,* 483–517.

Reid, M. J., Webster-Stratton, C., & Beauchaine, T. P. (2001). Parent training in Head Start: A comparison of program response among African American, Asian American, Caucasian, and Hispanic Mothers. *Prevention Science, 2,* 209–227.

Reid, M. J., Webster-Stratton, C., & Hammond, M. (2003). Follow-up of children who received the Incredible Years intervention for oppositional defiant disorder. *Behavior Therapy, 42,* 471–491.

Robinson, E. A., & Eyberg, S. M. (1981). The Dyadic Parent–Child Interaction Coding System: Standardization and validation. *Journal of Consulting and Clinical Psychology, 49,* 245–250.

Robinson, E. A., Eyberg, S. M., & Ross, A. W. (1980). The standardization of an inventory of child conduct problem behaviors. *Journal of Clinical Child Psychology, 9,* 22–28.

Spoth, R., Redmond, C., & Shin, C. (1998). Direct and indirect latent variable parenting outcomes of two universal family-focused preventive interventions: Extending a public health-oriented research base. *Journal of Consulting and Clinical Psychology, 66,* 385–399.

Stoolmiller, M., Eddy, J. M., & Reid, J. (1999). Detecting and describing prevention intervention effects in a universal school-based randomized trial targeting violent and delinquent behavior. *Journal of Consulting and Clinical Psychology, 68,* 296–306.

Taylor, T. K., & Biglan, A. (1998). Behavioral family interventions for improving child-rearing: A review of the literature for clinicians and policy makers. *Clinical Child and Family Psychology Review, 1,* 41–59.

Webster-Stratton, C. (1985a). Comparison of abusive and nonabusive families with conduct-disordered children. *American Journal of Orthopsychiatry, 55,* 59–69.

Webster-Stratton, C. (1985b). Comparisons of behavior transactions between conduct-disordered children and their mothers in the clinic and at home. *Journal of Abnormal Child Psychology, 13,* 169–184.

Webster-Stratton, C. (1985c). *Dyadic Parent–Child Interactive Coding System–Revised (DPICS–R): Manual.* Seattle: University of Washington.

Webster-Stratton, C. (1998a). Parent training with low-income clients: Promoting parental engagement through a collaborative approach. In J. R. Lutzker (Ed.), *Handbook of child abuse research and treatment* (pp. 183–210). New York: Plenum.

Webster-Stratton, C. (1998b). Preventing conduct problems in Head Start children: Strengthening parenting competencies. *Journal of Consulting and Clinical Psychology, 66,* 715–730.

Webster-Stratton, C., & Fjone, A. (1989). Interactions of mothers and fathers with conduct problem children: Comparison with a control group. *Public Health Nursing, 6,* 218–223.

Webster-Stratton, C., & Hammond, M. (1997). Treating children with early-onset conduct problems: A comparison of child and parent training interventions. *Journal of Consulting and Clinical Psychology, 65,* 93–109.

Webster-Stratton, C., & Hammond, M. (1998). Conduct problems and level of social competence in Head Start children: Prevalence, pervasiveness, and associated risk factors. *Clinical Child and Family Psychology Review, 1,* 101–124.

Webster-Stratton, C., & Hancock, L. (1998). Parent training: Content, methods, and processes. In E. Schaefer (Ed.), *Handbook of parent training* (2nd ed., pp. 98–152). New York: Wiley.

Webster-Stratton, C., Mihalic, S., Fagan, A., Arnold, D., Taylor, T. K., & Tingley, C. (2001). *Blueprints for violence prevention, book eleven: The Incredible Years: Parent, Teacher, and Child Training Series.* Boulder, CO: Center for the Study and Prevention of Violence.

Webster-Stratton, C., & Reid, M. J. (2001, XXXXXXX). *Preventing conduct problems, promoting social competence: A parent and teacher training partnership in Head Start.* Paper presented at the International Society for Research in Child and Adolescent Psychopathology, Vancouver, British Columbia.

Webster-Stratton, C., & Reid, M. J. (2003). The Incredible Years Parents, Teachers, and Children Training Series: A multifaceted treatment approach for young children with conduct problems. In A. E. Kazdin & J. R. Weisz (Eds.), *Evidence-based psychotherapies for children and adolescents* (pp. xx–xx). New York: Guilford.

Webster-Stratton, C., Reid, M. J., & Hammond, M. (2001a). Preventing conduct problems, promoting social competence: A parent and teacher training partnership in Head Start. *Journal of Clinical Child Psychology, 30,* 283–302.

Webster-Stratton, C., Reid, M. J., & Hammond, M. (2001b). *Treating children with early-onset conduct problems: Intervention outcomes for parent, child, and teacher training.* Manuscript submitted for publication.

Werthamer-Larsson, L., Kellam, S. G., & Oveson-McGregor, K. E. (Eds.). (1990). *Teacher interview: Teacher Observation of Classroom Adaptation–Revised (TOCA–R).* Baltimore, MD: Johns Hopkins University.

Yoshikawa, H. (1994). Prevention as cumulative protection: Effects of early family support and education on chronic delinquency and its risks. *Psychological Bulletin, 115,* 28–54.

Received December 9, 2002
Accepted October 25, 2003