

PHILIP A. COWAN *University of California*

CAROLYN PAPE COWAN *University of California**

MARSHA KLINE PRUETT *Smith College School for Social Work***

KYLE PRUETT *Yale School of Medicine****

PETER GILLETTE *University of California*****

Evaluating a Couples Group to Enhance Father Involvement in Low-Income Families Using a Benchmark Comparison

Following an earlier randomized clinical trial, now with broadened sample criteria, 236 low-income White, Mexican American, and African American couples participated in 16-week Supporting Father Involvement couples groups, with assessments at baseline, 2-, and 13-months postintervention. Because couples in the earlier control condition experienced no benefits and some declines in adaptation, a control condition was not offered. Data from the original couples groups (n = 96) and

controls (n = 98) served as benchmarks for evaluating the current results. Of 11 measures in this study, 10 revealed positive baseline–post 2 changes. Father involvement increased for current couples group participants, though not as much as for benchmark couples group participants: they showed statistically similar positive changes on six measures (decline in parenting stress, stability in couple relationship satisfaction, children’s hyperactivity, social withdrawal, psychological symptoms, increased income), and greater positive change on two of three measures (reductions in parents’ violent problem solving, children’s aggression).

Institute of Human Development, University of California, 1131 Tolman Hall – 1690, Berkeley, CA 94720-1690 (pcowan@berkeley.edu).

*Institute of Human Development, University of California, 1131 Tolman Hall – 1690, Berkeley, CA 94720-1690.

**Smith College School for Social Work, 310 Lilly Hall, Northampton, MA 01063.

***Yale School of Medicine, 230 Frontage Rd., P.O. Box 207900, New Haven, CT 06520.

****Hinshaw Lab, Institute of Human Development, University of California, 1131 Tolman Hall – 1690, Berkeley, CA 94720-1690.

Key Words: fatherhood, parenting, prevention/intervention, preventing marital distress, child outcomes.

Studies using randomized clinical trial designs (RCTs) to evaluate effects of intervention programs are relatively rare, even rarer in programs to promote fathers’ involvement with their children (Knox, Cowan, Cowan, & Bildner, 2011). Typically, once an RCT has produced evidence of efficacy, it is adopted by social service agencies without further systematic examination of whether the effects can be replicated in agency settings with more diverse clients.

In 2009, four of the authors of this article and a fifth colleague published a report of the Supporting Father Involvement (SFI) project—an RCT study of two variations of a group-based approach to strengthening relationships in low-income families, with a focus on encouraging fathers' involvement with their children (P. A. Cowan, Cowan, Pruett, Pruett, & Wong, 2009). That study compared three conditions: one 3-hour meeting in which couples discussed the importance of fathers (low-dose control), a 16-week couples group, and a 16-week fathers group, all led by the same mental health professionals. We found specific advantages for participants in the fathers group compared with those in the low-dose control condition, and even more advantages for couples group participants. Before moving beyond the efficacy study to launch a larger trial of the SFI approach in community settings, we set out to establish whether (a) the initial positive results could be replicated, (b) the intervention was effective with a more diverse set of families, and (c) participant characteristics on entering the study predicted postintervention changes.

We faced a dilemma about offering a randomized control condition in this study because our earlier results revealed no improvement and some negative changes over time for participants meeting only once. In the literature on the ethics of medical research, the principle of "clinical equipoise" endorses the use of randomized no-treatment, placebo, or low-dose intervention controls only when there is uncertainty about whether the experimenter-preferred treatment has potential advantages for the participants (Freedman, 1987). In consultation with the intervention staff, we concluded that there was an ethical issue in recruiting participants for a study condition in which the reasonable expectation was that they would experience no change or negative change over time. We considered a randomized control condition of parents assigned to community treatment as usual, but because there were no community services designed to enhance father involvement in the counties involved, we had no treatment-as-usual condition with which to compare the present intervention. We chose to use a "benchmark" approach by comparing the results of our new replication with already published data from our earlier Supporting Father Involvement RCT.

The Benchmark Approach

Benchmarking is one assessment strategy recommended as an approach for evaluating outcomes of interventions implemented in community settings (i.e., an effectiveness design; Hunsley & Lee, 2007). The *benchmark* term refers to the use of comparator values that serve as best practice standards for judging the impact of a service, in which either a single-point estimate or a range of acceptable values can be used. Hunsley and Lee strongly recommended that benchmarks be derived from sources such as meta-analyses that combine treatment results from multiple studies, though it is acknowledged that using data from a single RCT may be warranted in some circumstances. At the time the current study was conducted, the earlier SFI study was the only existing couple-focused intervention designed to promote father involvement and evaluate outcomes for the children, so multiple studies were not available for comparison.

The best chance of replication occurs when a new trial of an intervention approach is implemented with fidelity to a successful RCT. On the one hand, the more similar the features of the new study to those of the benchmark study, the more directly relevant the benchmark findings will be to the current evaluation. On the other hand, there is little justification for conducting a new study if the features of the two are identical. The need, then, is to find a balance between exact replication of the benchmark and include some variation to test the generalizability of the intervention results. Many central features of the benchmark RCT study were replicated exactly in the current study: all but one setting for this study were used in the benchmark study and included similar populations; a majority of the staff conducted both studies, with ongoing supervision provided by the creators of the SFI intervention; and 10 of 11 measures in this study were used in the benchmark study. Variations in this study involved adding one new site with African American couples, extending the age range of the youngest child from age 0 to 7 years (benchmark) to 0 to 11 years, and including couples in which one partner was not the biological parent of the youngest child.

The Benchmark Study

Participants in the initial iteration of the SFI project (P. A. Cowan et al., 2009)—the

benchmark study – included 279 pairs of Mexican American and White coparents residing in four California counties, two thirds of whom had household incomes below 200% of the federal poverty line. Mothers and fathers who were biological parents of their youngest child were invited to take part in one of three randomly assigned conditions: (a) a one-time informational meeting (3 hours), (b) a 16-week (32-hour) group for couples, or (c) a 16-week (32 hour) group for fathers. At each site, all three versions of the intervention were conducted by the same male–female group leaders, with an almost identical curriculum for the couples and fathers groups. In Weeks 5 and 13, mothers and fathers in the couples and fathers groups met separately so that fathers and male coleaders could do some hands-on activities with the youngest child, while mothers met with the female coleaders. All couples, including those in the single informational meeting, were also assigned a case manager during the 18-month study to facilitate attendance, make referrals when additional services were needed, and conduct pre- and postintervention assessments with each parent individually.

Fathers and mothers who participated in the one-time meeting showed no positive changes and some negative changes; at the 18-month follow-up assessment, their satisfaction as a couple had declined significantly, and they reported increased externalizing and internalizing behaviors in their children. By contrast, participants in the fathers and couples groups showed significant increases in fathers' involvement in the care of the children and no increase in children's behavior problems over the course of the study. Mothers and fathers from the couples groups showed the same positive benefits of the fathers groups, along with significant reductions in parenting stress, and no decline in satisfaction as a couple. Because the couples groups had more widespread effects on the families, we chose not to repeat the fathers-only group format.

*Rationale for the Supporting Father
Involvement Couples Group Intervention
Approach*

Father involvement can be defined as direct physical and emotional engagement, and active physical caring involving the tasks of childrearing. The positive impact of fathers' involvement comes not from the amount of time they are

directly involved in childrearing but primarily from the quality of the relationship established with the child (Parke, 2002).

In the last half of the twentieth century there were two approaches to designing interventions to enhance father–child relationships. First, parenting skills classes have always “welcomed” fathers, although they are usually conducted and attended primarily by women. Second, conventional wisdom suggested that the best way to promote father involvement is to create men's groups with male leaders, where men discuss how to overcome motivational and societal barriers to becoming active participants in children's lives. Only in the past two decades have intervention programs to promote father involvement been evaluated systematically. The first generation of programs, which focused largely on helping men to increase their economic self-sufficiency to pay child support, yielded disappointing results (see Knox et al., 2011). As fatherhood programs began to include a focus on family relationships, the systematically evaluated ones found significant effects on father involvement and family relationship quality (Caldwell et al., 2011; Fagan, 2008).

The structure of the SFI intervention, its curriculum, and the assessments for this study were based on a family systemic approach to children's development in which risk and protective factors from five domains interact in a dynamic fashion (P. A. Cowan et al., 2009): each parent's individual well-being and mental health, three generational patterns of adaptation and dysfunction, couple relationship quality, parent–child relationship quality, and life stress and social support outside the nuclear family. Fathers tend to be more engaged with their children when (a) they are less depressed and anxious; (b) relationships in their family of origin are positive; (c) the relationship between father and mother is collaborative, regardless of whether they are married, cohabiting, separated, or divorced; (d) their relationship with the child includes warmth and appropriate limits; and (e) social supports help balance life stresses. The same factors, separately and cumulatively, predict different indices of children's well-being (Parke, 2002; Tamis-LeMonda & Cabrera, 2002).

In designing the SFI involvement intervention, we were influenced by the consistent finding that across the economic spectrum, the single best predictor of fathers' family involvement is the quality of the father's relationship with the

mother (Carlson, Pilkauskas, McLanahan, & Brooks-Gunn, 2011), which holds for married, cohabiting, separated, and divorced coparents (Pruett & Johnston, 2004). Furthermore, when couples are more satisfied with their relationship as partners, their parenting is more sensitive and attuned to the needs of their children (e.g., Adler-Bader et al., 2013). The implication of these findings is that an intervention format that includes both parents and focuses on improving the relationship between them could be expected to have positive effects on fathers' involvement in family life and on the quality of both parents' relationship with the child.

Very few studies have evaluated interventions for couples in low-income families, and, until recently, even fewer have investigated the potential impact of couples groups on father involvement or children's well-being. Three relatively recent large-scale published studies adopted a couples group approach to encouraging father involvement. Building Strong Families (Wood, McConnell, Quinn, Clarkwest, & Hsueh, 2010) randomly assigned low-income unmarried couples at eight U.S. sites to groups that emphasized couple communication skills or to a no-intervention condition. Although an intention-to-treat analysis found no overall significant effects on the participants 18 months postintervention, one site did show statistically significant positive effects of the couples group intervention on father involvement. Rienks, Wadsworth, Markman, Einhorn, and Etter (2011) randomly assigned participants to couples groups, groups attended by only one of the parents, or a no-treatment control condition. The curriculum adapted the Premarital Relationship Education Program (PREP; Markman, Stanley, & Blumberg, 2010), which focuses on couple communication skills. Father involvement increased more for couples group participants than for no-treatment control families or men whose partners attended groups alone. Notably, as suggested by prior correlational studies, an increasingly strong alliance between the parents was associated with increased father involvement. The third couples-based approach to enhancing fathers' family role is the SFI approach, described earlier as the benchmark study.

The Current Study

The current study was designed as a field test of the SFI intervention to answer three concerns:

(a) Would the results of the initial couples group RCT hold up in a new sample? (b) Could the results be replicated in a sample that included African American as well as White and Mexican American couples and nonbiological parents with a youngest child between birth and age 11, and were there differences in patterns of change attributed to the wider inclusion criteria? (c) Did the participants' baseline adaptation (e.g., couple relationship satisfaction, parenting stress) affect the course of change? Specifically, were initial levels of distress associated with more or less positive postintervention change?

Based on two couples group intervention trials conducted with middle-class families (C. P. Cowan & Cowan, 2000; P. A. Cowan, Cowan, Ablow, Johnson, & Measelle, 2005) and the benchmark study of low-income families (P. A. Cowan et al., 2009), the curriculum for the current 16-week group intervention included modules covering topics in all five family domains. For this article, the measurement focus was narrowed to the couple, parent-child, and child outcome domains, in which there were direct intervention effects in the benchmark study on at least one measure in each domain. Stimulated by the consequences of U.S. economic difficulties in 2006 to 2009, we also examined whether household income changed over time.

We predicted that, as in the benchmark study, this intervention would show positive results in terms of couple relationship quality, parent-child relationship quality, and child outcomes. Without the benefit of a low-dose control condition, it was important to define what positive results would mean in the comparison between pre- and postintervention assessments. For some measures, stability over time represents a positive outcome. More than 50 studies in many Western countries show that without intervention, satisfaction with the couple relationship declines for parents of young children over periods ranging from 1 to 15 years (Twenge, Campbell, & Foster, 2003). Thus, stable relationship satisfaction over time in couples group participants would represent a positive finding. Because parents in the benchmark control condition showed increases in the behavior problems ascribed to their children and ongoing group participants did not, we predicted stability over time on the four measures of children's behavior problems. Based on benchmark study data, we predicted that parenting stress would decline over time.

Because benchmark couples' violent problem solving and harsh parenting ideas were stable over time in intervention and nonintervention participants, and because we had not previously examined family income, we did not make specific predictions about change in these measures in this study, nor did we specify hypotheses about who benefits most—a post hoc exploratory aspect of this study.

METHOD

Participants

The SFI project and staff were located within Family Resource Centers in five California counties, four of which were rural and had provided the SFI services in the benchmark study. Staff in each setting included a project director, two group leaders (one male, one female), two case managers, and a data coordinator. At each site, project staff recruited participants through direct referrals from other services in the Family Resource Center, other county agencies, talks at community organizational meetings, ads in the local media, and information tables at sports events, malls, and community events attended by fathers.

Case managers used a brief screening interview with 440 couples who applied to SFI to assess whether parents met four criteria identical to those in the benchmark study: both partners agreed to participate, the participants were raising a young child together, neither parent suffered from a mental illness or drug or alcohol abuse problems that interfered with their daily functioning at work or in caring for their child(ren), and the family had no current or recent (within 12 months) open child or spousal protection case or instance of spousal violence or child abuse that had been reported to a child welfare agency. Because this latter criterion might increase the risks for child abuse or domestic violence, these families were referred for other appropriate services and not included in the study.

In this study, in addition to a fifth site that included African American participants, two selection criteria represented a more inclusive approach: the family's youngest child could range between 0 (pregnancy) and 11 years (the benchmark was 0–7), and coparents were included in which only one parent was the biological parent of the youngest child and the

other was involved in raising that child (e.g., mother or father figure, grandparent).

Over the five sites, group leaders met with each of the 429 eligible couples for a joint 1.5-hour initial interview that introduced topics in five aspects of family life (individual, couple, parent–child, three-generational, life stress and social support) that underlie the risk and protective model (C. P. Cowan et al., 2000) underlying the program and assessments. At the end of the interview, the couples were offered a chance to participate in a couples group and 347 (81%) accepted. Each parent who signed the consent forms was then scheduled for a 1.5- to 2.5-hour baseline assessment consisting of questionnaires administered orally in English or Spanish by a case manager. Of the 347 couples assigned to a couples group, 319 (638 partners; 92%) completed the baseline assessment. Of these, 246 couples (492 partners; 77%) completed post 1, and 236 (472 partners; 96% of post 1 respondents) went on to complete post 2.

Of the participants who completed the baseline assessment, 50% described themselves as Mexican American, 31% White, 11% African American, 5% Asian American or Native American, and 5% mixed race. On entering the study, 62% of the couples completing baseline assessments were married, 27% were cohabiting, 8% were living separately and raising a child together (separated, divorced, or never-married, never-cohabiting couples), and 3% were noncoupled coparents (e.g., father–aunt or father–grandmother). Median household income was \$29,700 per year, with 64% of the sample falling below twice the federal poverty line (\$40,000 yearly household income for a family of four), although a few parents (3%) had yearly incomes ranging from \$100,000 to \$200,000. A majority of fathers (77%) and many mothers (44%) had worked for pay during the week prior to their baseline assessment; 60% of the fathers and 47% of the mothers had completed high school or beyond. At baseline, the number of children in the household ranged from 0 (mother pregnant) to 7 (median = 3.0). The mean age of the youngest child was 3.14 ($SD = 2.71$); median age was 2.00, with a range from 3 months prepartum to 11 years, and the distribution was highly skewed toward the younger ages. Although we were open to accepting couples in which one partner was not a biological parent of the youngest child, only

nine such couples actually participated – too few to analyze whether this affected the results.

Couples Groups

Across the five county sites, there were 39 16-week couples groups (five to nine couples per group), 19 conducted in Spanish, 30 in English. Child care was provided to allow parents to focus on their family issues undisturbed. Each meeting involved a structured curriculum of exercises, discussions, short presentations, and an open-ended time in which participants were free to raise issues and concerns for discussion and problem solving. The structured part of each meeting followed a curriculum focused on our family systems theoretical model, and each meeting focused on one of five domains of family life discussed earlier that constitute risk or protective factors affecting family functioning and children's adaptation.

The clinically experienced coleaders' questions and exercises encouraged participants to discuss how they felt about themselves and what they hoped to change (e.g., discrepancies between actual and ideal self-descriptions), parenting ideas and strategies (e.g., defining and role playing different parenting styles), communicating as a couple (e.g., developing collaborative problem-solving strategies), three generational family patterns (e.g., to repeat or modify in their current family), and supports for dealing with life stresses (e.g., compiling a list of helpful personal and community resources).

On a continuum of intervention styles ranging from open-ended group therapy (e.g., Yalom, 1995) to psychoeducational teaching of specific communication skills (e.g., Markman et al., 2010), our approach occupies a middle ground. Structure and model fidelity were maintained by the curriculum, leaders' reports of topics covered each week, monthly conference calls with a member of the curriculum development team for each staff function at all sites (project directors, group leaders, case managers, data coordinators), and twice yearly all-site meetings. As in the earlier study, all couples received the services of a case manager who made referrals as needed for assistance with family, medical, housing, employment, or legal issues over their 18 months of participation. Case managers also followed up with participants when they missed a group meeting and maintained contact with them between the three assessment interviews.

Parents were not paid for attending groups, but each partner received \$50 for completing the baseline assessment interview, \$50 for the post 1 assessment, and \$100 for the post 2 assessment (a possible total of \$400 per family over 18 months).

Measures

To assess changes in the quality of the parent-child relationships, the couple relationship, and children's adaptation from baseline to the 18-month follow-up, we chose 10 measures identical to those examined in the benchmark study and household income as an eleventh measure.

Parent-Child Relationships.

1. The Pie (C. P. Cowan & Cowan, 1990b) allowed participants to represent graphically their psychological investment in various aspects of their lives by dividing a circle 4" in diameter to represent the salience or importance of that aspect of self, not the amount of time spent in the role. This study focuses on the degrees of the circle they labeled *father* or *parent*.
2. Who Does What? (C. P. Cowan, 1990c) is an 11-item questionnaire administered to both parents to assess fathers' relative involvement in the care of their youngest (target) child (e.g., feeding, getting up with the child at night), using a 1 to 9 scale ranging from 1 (*she does it all*), 5 (*we do it about equally*), to 9 (*he does it all*). Item reliabilities at baseline were high ($\alpha = .80$ for fathers and $.81$ for mothers). Correlations between fathers' and mothers' descriptions at the three assessment points ranged from $.62$ to $.74$, suggesting that both partners described their division of family labor similarly, though not identically.
3. Each parent's level of stress in parenting the youngest child was assessed with a 38-item revised version of the Parenting Stress Index (PSI; Abidin, 1997). Parents indicated the extent of their agreement or disagreement with statements describing themselves as stressed, their child as difficult to manage, and a lack of fit between what they expected and the child they have. The scale has been validated by comparing parents who do and do not have known childrearing stressors

(developmental delay, oppositional defiance, or difficult temperaments).

- In the Ideas About Parenting questionnaire (Heming, Cowan, & Cowan, 1990), fathers and mothers indicated the extent of their own agreement or disagreement with each of 40 items and what they believed were their partners' opinions. We used one of the three factors (12 items), Authoritarian Parenting, which describes harsh parenting with high structure and demands for conformity. Items included "Some children can only be made to obey by scolding and punishment," "A child should not talk back to a parent," and "Parents should keep a firm hold on their child's expression of angry feelings." The authoritarian parenting scale differentiated parenting in another study of families with girls with attention and hyperactivity disorders from girls diagnosed as inattentive but not hyperactive (Hinshaw, 2002).

Couple Relationship Quality.

- The Quality of Marriage Index (QMI; Norton, 1983), a 6-item questionnaire renamed the Quality of Relationship Index for the participants, with one global estimate and five specific questions about couple relationship satisfaction, was used to measure each partner's satisfaction with the couple relationship. This one-factor scale has high overlap with longer, traditional measures of marital quality (Heyman, Sayers, & Bellack, 1994).
- The Couple Communication Questionnaire (C. P. Cowan & Cowan, 1990a) provided a 6-item scale of the degree to which the couples use violent behavior during conflicts or disagreements. Items included "I yell or insult my partner," "I throw something," and "I push, grab or shove my partner" ($\alpha = .93$ for fathers and $.94$ for mothers).

Children's Behavior Problems. Each parent filled out a 54-item adaptation of the 106-item Child Adaptive Behavior Inventory (Cowan, Cowan, & Heming, 1995). We composited the scores into four dimensions based on a factor analysis of the scale: Externalizing-Aggression, Externalizing-Hyperactivity, Internalizing-Shy Withdrawn, and Internalizing-Anxiety/Depression. In previous studies (Gottman & Katz, 1989), the

interitem consistencies of these composite dimensions filled out by teachers were very high (alphas in the .80s and .90s) and those filled out by parents were moderate (.60s and .70s). In this study, the alphas for parents' descriptions ranged between .71 (hyperactivity) and .85 (externalizing-aggression) for mothers and fathers; all of the reliabilities for parents' ratings in the current low-income study were higher than those for the corresponding scales in a middle-class sample. Correlations between mothers' and fathers' descriptions on each of the four dimensions were consistently moderate to high at each assessment: baseline .35 to .45, post 1 .35 to .54, and post 2 .35 to .53.

RESULTS

Retention and Group Attendance

Using a mixed model generalized linear model (GLM) analysis of baseline measures (sex of parent x 11 baseline measures x completion vs. noncompletion of post 2 assessment), we tested for a possible retention bias attributable to the completers having had more adaptive baseline scores than couples who later dropped out. There was no statistically significant interaction between sex of parent and completion versus noncompletion. Compared with completers, parents who dropped out of the study reported lower relationship satisfaction, $F(1, 296) = 8.59$, $p < .01$, more violent problem solving, $F(1, 296) = 7.25$, $p < .01$, harsher parenting ideas, $F(1, 296) = 5.92$, $p < .05$, and lower family income at baseline, $F(1, 296) = 4.36$, $p < .05$. There were no statistically significant differences in the other seven measures. As in the benchmark study, parents in this study who dropped out by post 2 had described their families more negatively at baseline on some, but not all, measures.

Of the participants who completed the baseline assessments, 62% of the White couples, 54% of the African American couples, and 82% of the Mexican American couples completed the posttests; the Mexican American participation rate was significantly higher than that of the other two groups. Despite some differences in retention rates attributable to initial adaptation status and ethnicity, there was a full range of scores in all categories on each of the 11 measures.

In the couples groups, 19% of the mothers had perfect attendance (32 hours), 53% attended

more than 25 hours, 72% more than 19 hours, and 80% more than 13 hours, with a median attendance rate of 81% of the sessions attended. Fathers in the couples groups had a very similar distribution. Once couples attended the first or second meeting of their group, the median attendance rate was close to 90%. Attendance rates did not differ significantly by ethnic group.

Changes Over Time in the Current Sample

Table 1 presents means and standard deviations for each measure at baseline (preintervention), post 1 (2 months after groups ended, 7 months after baseline), and post 2 (18 months after baseline), separately for fathers and mothers. To determine whether there were statistically significant changes between baseline measures and post 1 and post 2 assessments, we conducted MANOVAs on each of 11 measures, with time (pre–post 1–post 2) and sex (mother–father) as within-subject effects. We used Bonferroni corrections to adjust for multiple statistical tests. There were no statistically significant interactions between sex of parent and measure. In interpreting the data presented in Table 1, note our definitions of what constitutes positive results (see Introduction).

Of the four measures of parent–child relationship quality, only The Pie failed to show positive changes over time. The size of the “slice” labeled *father* remained stable in this study rather than increasing over time as it did for fathers in the benchmark couples groups. The remaining three measures all showed statistically significant positive changes: Both parents’ reports of father involvement on Who Does What? increased significantly from baseline to the 18-month follow-up, whereas parents’ harsh authoritarian Ideas About Parenting (controlling, harsh, low warmth) and scores on the PSI (“My child does a few things which bother me a great deal”) showed significant declines over the same time period.

The two measures of couple relationship quality also showed positive changes. We noted earlier that in light of the decline in relationship quality in almost all longitudinal studies (Twenge et al., 2003), maintenance of couple relationship satisfaction over 18 months (Quality of Marriage Index) represents a positive outcome for the current couples. This interpretation is supported by the fact that in this study, couples’ violent problem-solving

strategies (yelling, throwing things, hitting) on the Couple Communication Questionnaire decreased significantly over 18 months.

Based on our earlier findings, we predicted that the four measures of child behavior problems on the Child Adaptive Behavior Inventory (aggression, hyperactivity, social isolation, internalizing psychological symptoms) would show no statistically significant baseline to post 2 change. In fact, parents’ ratings of their child’s hyperactivity showed a significant *decrease* over time, while the other child measures met our expectations.

Comparing the Current and Benchmark Findings

Our attempts to increase the diversity of participants in the study by choosing one new site and broadening the acceptance criteria met with partial success. By adding a new site, the current couples groups were able to include 69 African American participants (there were none in the benchmark couples groups). Expanding the children’s age range resulted in a significant increase in the mean age from 2.50 years in the benchmark study to 3.19 years, $t(329) = 2.23, p < .05$, in this study, though there was considerable overlap in age between the samples, with a skewing toward young children in both.

To compare the current and benchmark samples in terms of similarities at baseline and change over time, we conducted MANOVAs on each of 11 measures, with condition (benchmark controls, benchmark couples group participants, and current study couples group participants) as between-subject effects, and Time (pre–post 1–post 2) and Sex (mother–father) as within-subject effects. Again, we used Bonferroni corrections to adjust for multiple statistical tests. There were no statistically significant differences at baseline on any of the 11 assessment measures among the benchmark controls, benchmark couples group, and current couples group participants.

The main question addressed here is how the baseline to post 2 changes in the current couples group participants compare with the changes obtained in the benchmark RCT.

Mean scores at each assessment of the current couples group participants are presented in Table 1. Mean scores for the benchmark control couples and couples group participants were published in Tables 1 and 2 of the earlier P. A.

Table 1. Pre, Post 1, Post 2: Means, Standard Deviations, and F Tests of Change¹

	Pre	Post 1	Post 2	F _{time}
Measure (possible scores)				
Parent-Child				
Fathers' <i>Pie</i> (0-360 degrees)	105.79 (50.75)	111.80 (54.61)	109.87 (55.64)	ns
<i>Who Does What?</i> Father report (1-9)	3.86 (1.00)	3.91 (1.00)	3.95 (1.05)	8.44**
<i>Who Does What?</i> Mother report (1-9)	3.37 (1.04)	3.58 (1.09)	3.66 (1.20)	
Father's authoritarian <i>Ideas About Parenting</i> (9-108)	69.89 (10.82)	69.71 (10.90)	69.49 (11.19)	6.19**
Mother's authoritarian <i>Ideas About Parenting</i> (9-108)	68.45 (12.20)	67.91 (12.68)	66.54 (13.21)	
Father's Parenting Stress Index (5-180)	70.71 (18.94)	68.18 (18.03)	68.79 (19.50)	5.84*
Mother's Parenting Stress Index (5-180)	70.05 (19.24)	70.00 (20.05)	67.81 (18.15)	
Couple				
Father's Quality of Marriage Index (6-45)	36.77 (7.10)	36.94 (7.84)	35.86 (9.35)	ns
Mother's Quality of Marriage Index (6-45)	34.41 (9.00)	35.20 (8.39)	34.39 (9.56)	
Father's violent problem-solving	1.01 (1.57)	0.81 (1.45)	0.93 (1.63)	8.12**
Mother's violent problem-solving	1.30 (1.66)	0.94 (1.29)	0.98 (1.42)	
Children's adaptation				
Aggression: Father (1-4)	1.80 (.56)	1.80 (.53)	1.82 (.49)	ns
Aggression: Mother (1-4)	1.80 (.51)	1.83 (.50)	1.82 (.45)	
Hyperactivity: Father (1-4)	2.10 (.61)	2.09 (.62)	2.00 (.58)	4.09*
Hyperactivity: Mother (1-4)	2.06 (.60)	2.05 (.62)	2.03 (.59)	
Social isolation: Father (1-4)	.42 (.89)	.25 (.87)	.36 (.90)	ns
Social isolation: Mother (1-4)	.30 (.90)	.35 (.87)	.34 (.92)	
Internal Psych Symptoms: Father (1-4)	1.54 (.47)	1.50 (.43)	1.54 (.41)	ns
Internal Psych Symptoms: Mother (1-4)	1.49 (.45)	1.49 (.43)	1.57 (.47)	
Family income				
Yearly income	39,507.00 (1753)	41,814.00 (1861)	43,078.00 (2008)	17.03***

* $p < .05$, ** $p < .01$, *** $p < .001$.

Cowan et al. (2009) study. To facilitate a direct comparison of the two studies, Table 2 of the current study reports t tests of the change between baseline and post 2 for families in each of the three conditions, derived from post-hoc analyses of the Condition \times Time \times Sex MANOVAs. Because there were no significant differences in fathers' and mothers' patterns of change in any of the MANOVAs, we present the data for fathers and mothers combined. A plus sign for the t tests indicates an increase from baseline, a minus sign, a decrease; whether this represents a positive or negative outcome depends on the direction in which the variable was scored.

The first step in the current analyses was to determine whether adding data from the current couples group sample to the benchmark analyses would maintain the significant Condition \times Time interactions in the benchmark comparison between control and group participants. In effect, this strategy allowed us to determine whether there were statistically significant differences

in the t values representing baseline to post 2 change within each row in Table 2. The second step was to conduct follow-up analyses specifically to compare baseline to post 2 changes in couples group participants in the benchmark and current studies. Table 2 includes F tests for the overall interaction (three conditions) and for the comparison of the current and benchmark couples groups (two conditions). The far right column of Table 2 details the outcome of each analysis in terms of whether baseline-post 2 changes in this intervention participants fail to replicate, match, or exceed outcomes in the earlier study.

Four measures focused on changes in the parent-child relationship. In the men's size of the piece of The Pie labeled *father* there was a main effect increase over time, $F(2,421) = 5.28$, $p < .01$, but the Condition \times Time interaction was not statistically significant. Changes in parents in the benchmark controls, benchmark couples groups, and current couples groups were not statistically different from each other. That is, with the data added from the

Table 2. Pre vs. PO2 *t* Tests: Benchmark and This Study

	Benchmark Controls	Benchmark Couples Groups	Current Study Couples groups	Overall $F_{\text{Condition} \times \text{Time}}$	Benchmark vs. Current Couples Groups $F_{\text{Condition} \times \text{Time}}$	Outcomes for Current Couples vs. Benchmark Couples
Number of couples	98	95	236	429	331	
Parent-child						
Fathers' Pie <i>t</i>	+1.01	+1.99*	+1.00	ns	ns	Failure to replicate
Who Does What? <i>t</i>	+1.67	+4.10***	+2.88**	3.81**	6.12***	Less positive
Authoritarian Ideas about Parenting <i>t</i>	+ .21	-.87	-2.25*	ns	ns	More positive but not sig.
Parenting Stress Index <i>t</i>	-.11	-3.98***	-2.38*	2.74**	ns	Equally positive
Couple						
Quality of Marriage Index <i>t</i>	-4.83***	-1.20	-1.16	6.16***	ns	Equally positive
Violent problem solving <i>t</i>	-.15	-1.28	-2.46**	3.67**	3.25***	More positive
Child						
aggression <i>t</i>	+2.17*	+1.86	+ .82	3.13**	5.82***	More positive
Hyperactivity <i>t</i>	+2.18*	+ .11	-1.99*	3.46**	ns	Equally positive
Social Isolation <i>t</i>	+2.44*	+ .49	-.13	ns	ns	Equally positive
Internal. Psych. Symptoms <i>t</i>	+2.27*	+1.88	+1.75	ns	ns	Equally positive
Family income						
Yearly income	+3.48**	+3.42**	+4.10***	ns	ns	Equally positive

Note: + increase from Pre to PO2, - decrease from Pre to PO2.

* $p < .05$, ** $p < .01$, *** $p < .001$.

current replication, we cannot conclude that participation in a couples group increases men's psychological sense of themselves as father. This was the only measure in this study that failed to replicate earlier positive results. By contrast, when we examined men's hands-on involvement in caring for their children (Who Does What?), mothers' and fathers' descriptions of father involvement in the benchmark couples groups, $t(93) = 4.10, p < .001$, and the current couples groups, $t(234) = 2.88, p < .01$, showed statistically significant increases over time, but the increase in the current study was smaller, $F(2, 660)_{\text{Condition} \times \text{Time}} = 6.12, p < .001$.

The current couples reduced their harsh and critical ideas about how to treat their children (Ideas about Parenting) after participating in the intervention, $t(234) = 2.25, p < .05$, but the absence of a significant Condition \times Time interaction showed that the improvement was not significantly greater than that of parents in the benchmark couples group or control condition. That is, this measure may change over time, but the change cannot be attributed to participating in the SFI intervention. For couples group fathers and mothers in the benchmark and current studies, reductions in the stress that they

experienced as parents (PSI) were significantly greater than for parents in the benchmark control condition who attended only one meeting, $F(4/854)_{\text{Condition} \times \text{Time}} = 2.74, p < .01$. The reduction in parenting stress experienced by current group couples was equal to that experienced by benchmark group couples.

The current couples group participants fared quite well in terms of the quality of their relationship. The stability over time of relationship satisfaction (Quality of Marriage Index) in current and benchmark couples group participants contrasted significantly with the decline in relationship satisfaction shown by the couples in the benchmark control condition, $F(4/854)_{\text{Condition} \times \text{Time}} = 6.16, p < .001$. Furthermore, the benchmark couples group participants' had fewer positive outcomes than the current couples group participants, who reported a greater reduction in violent problem solving (pushing and hitting each other, Couple Communication Questionnaire, $F(2, 660)_{\text{Condition} \times \text{Time}} = 3.25, p < .001$).

We reported previously (P. A. Cowan et al., 2009) that the four measures of problematic child behavior showed significant increases over time in parents' reports of children's behavior

in the benchmark control parents, but not in those of the benchmark couples group parents. That article acknowledged that the t tests were conducted as an exploratory analysis, despite the fact that the Condition \times Time interaction was not statistically significant. Adding the current results to the analysis showed that in two of the four measures (parents' perception of their children's aggression and hyperactivity on the Child Adaptive Behavior Inventory), there were now statistically significant differences in change over time among groups: aggression, $F(4/854)_{\text{Condition} \times \text{Time}} = 3.13$, $p < .01$, and hyperactivity $F(4/854)_{\text{Condition} \times \text{Time}} = 3.46$, $p < .01$. In the eyes of their parents, children of benchmark controls increased in aggressive and hyperactive behavior over 18 months, whereas children of couples group parents in the benchmark and current studies remained relatively stable. Follow-up comparisons of the two ongoing intervention conditions revealed that children of current couples group participants showed more stability in aggressive behavior than the children of the benchmark couples group participants, $F(2/660)_{\text{Condition} \times \text{Time}} = 5.82$, $p < .001$ (i.e., more positive results), whereas the stability of hyperactive behavior in the current couples group participants was equivalent to that seen by benchmark intervention parents (i.e., also positive results).

In this study, as in the benchmark study, the overall Condition \times Time interaction was not statistically significant for children's social isolation and internalizing psychological symptoms, even though simple post hoc t tests replicated the pattern found in our earlier study in which children of parents in the control condition increased in problem behavior while the behavior of children of parents in the couples group remained stable over 18 months.

Finally, the statistically significant increases in income reported by couples in all three conditions were not different from each other. It is important to note that the benchmark study was conducted in 2003 to 2004 during relatively good economic times in California, whereas this study was conducted in 2010 to 2012 during the slow recovery from the recession.

In sum, father involvement increased significantly in this replication of SFI – somewhat less than it had in the benchmark study – whereas parenting stress declined in this study as much as it had in the earlier benchmark findings. Satisfaction with the couple relationship remained

stable in this study as it had earlier, whereas the decline in violent problem solving for current parents was significantly greater than for parents in the benchmark couples groups. The stability in children's problem behavior in current and benchmark families contrasted with the increase in children's problem behavior for the benchmark controls. Finally, statistically significant increases in annual household income were experienced by couples in this study, although these increases were not different from those experienced by participants in the benchmark controls and couples groups.

For Whom Was the Intervention Effective?

Using separate GLM Time \times Potential Moderator \times Sex analyses, we examined the possible effects of two moderator variables on changes from baseline to post 2 to test whether the more inclusive recruitment in the current study affected the results: ethnic group membership (White, Mexican American, or African American families) and child age (0–7 vs. 8–11). None of the 11 measures showed significant interactions of Time \times Couple ethnicity. That is, patterns of change and stability in this study in African American families did not differ significantly from those of the White and Mexican American families. Only one of 11 measures showed a statistically significant interaction of Time \times Child Age (0–7 years vs. 8–11 years; $F(4, 854) = 5.51$, $p < .001$): according to the current parents of younger children, aggressive behaviors increased significantly, whereas current parents of older children reported a significant decline in aggression. Given that these findings reflect 22 GLM analyses, we conclude that the expanded inclusion criteria of ethnicity and age had little effect on the baseline to post 2 changes in the participants. The third criterion for expansion of the sample – the inclusion of couples in which one partner was not the biological parent of the child – resulted in the participation of only nine of 236 couples, not enough to test whether their inclusion made a difference to the results.

Within this study, we examined associations between baseline scores for couples entering the study and positive changes 18 months later. Fathers and mothers with lowest adaptation scores at the beginning of the study showed the most gains over time: reduction in harsh ideas about parenting, $r_{\text{fathers}}(234) = .39$, $p < .001$,

$r_{\text{mothers}}(234) = .31, p < .001$; reduction in parenting stress, $r_{\text{fathers}}(234) = .42, p < .001$, $r_{\text{mothers}}(234) = .50, p < .001$; increase in fathers' involvement, $r_{\text{fathers}}(234) = -.47, p < .001$, $r_{\text{mothers}}(234) = -.47, p < .001$; less decline in couple relationship satisfaction, $r_{\text{fathers}}(234) = -.09, ns$, $r_{\text{mothers}}(234) = -.38, p < .001$; and more decline in violent problem solving, $r_{\text{fathers}}(234) = .46, p < .001$; $r_{\text{mothers}}(234) = .57, p < .001$.

DISCUSSION

In a comparison of the current replication of an RCT benchmark study, we posed three specific questions about the SFI couples group intervention. All were answered in the affirmative: (a) Would the benchmark intervention results hold up in a new sample? (b) Could the results be replicated in a more diverse sample that included slightly older children and African American families? (c) Could we identify characteristics of participants associated with positive pre- to postintervention change? We had initially asked whether the replication was effective with nonbiological parents, but that subset of participants was too small to examine.

Replicating the Pattern of Results

What would constitute evidence that the current iteration of a couples group intervention was effective? We attempted to answer this question by comparing the findings in this study, which offered the same couples group curriculum to a more inclusive population, with the pattern of findings in our previous RCT. Despite this study's addition of an African American sample and an increase in the average age of the family's youngest child, couples entering this study were remarkably similar to the benchmark samples on the 11 measures. Thus, it is not likely that any differences in patterns of change between the current couples and the benchmark couples are confounded by differences in sample characteristics.

In this replication of the SFI couples group intervention, participant couples in all three ethnic groups, who were at risk because of their low incomes, showed positive changes in questionnaire measures of parent-child relationship quality, couple relationship quality, children's problem behaviors, and family income. One of 11 measures in the current replication of the SFI couples group intervention failed to replicate the

positive results obtained in the earlier benchmark RCT (an increase in men's psychological involvement as "father"). Of the remaining 10 measures completed by the current couples, one revealed a significant change that was not statistically different from the benchmark results (authoritarian parenting ideas), and one showed a positive change that was smaller than that reported by benchmark couples (father involvement). The central finding was that six of the measures showed positive baseline to post 2 changes that were equal to those of the benchmark intervention participants (declines in parenting stress, stability in couple relationship satisfaction, children's aggression, hyperactivity, social isolation, and psychological symptoms), and two showed significantly more positive changes than those of couples in the benchmark intervention (decline in couples' violent problem solving and their children's aggression).

Without the benchmark comparison data, it would have been difficult to interpret the results of this study. What are we to make of the fact that couple relationship satisfaction and three measures of child problem behavior remained stable over time? The benchmark RCT provides information that without intervention, couple relationship satisfaction declines and child problem behaviors increase. The results of the benchmark comparison also serve to temper the conclusions that we would have arrived at if we were simply to consider the current single-sample changes over time. Most importantly, our finding that household income increased significantly must be interpreted in light of the fact that it also increased in the benchmark controls.

It is also the case that this replication trial reflects on interpretations of the earlier benchmark study. First and foremost, it demonstrates that repeating the SFI intervention with a new and slightly more diverse sample produces results that are similar to those obtained in the earlier study, and therefore reinforces confidence in the effectiveness of SFI; replication is a central feature of intervention evaluation. Second, by failing to replicate the findings concerning men's psychological sense of themselves as fathers (The Pie), this study suggests that the earlier findings on this specific measure might represent chance fluctuation.

In comparison with the benchmark results, the present findings seem to suggest a weaker effect in the area of father-child relationships than in couple relationships or child behavior. The

results also suggest stronger effects on measures of current couples violent problem solving and children's aggressive behavior, outcomes that may be connected. Both of these conclusions require further replication, which are now in progress, to determine whether the findings represent random variation around a generally positive set of effects, or whether iterations of the intervention create systematic effects as they play out in different family domains over time.

Although the benchmark low-dose single-session meeting cannot serve as a randomized control condition for the present participants, it does represent a comparison sample with similar characteristics and staff in four of the five settings. Even without a randomized control condition, the finding concerning the maintenance of couple relationship quality over 18 months in this study is noteworthy. The vast majority of more than 50 transition-to-parenthood studies with no intervention find that couple relationship satisfaction declines over time (Twenge et al., 2003). The fact that parents in this study did not decline in couple satisfaction, whereas the satisfaction of the benchmark controls, without the advantage of a couples group, declined, adds weight to the inference that this pattern reflects the impact of the SFI intervention.

One measure, parents' household income, showed significant increases in all three conditions. Although these changes cannot be attributed to participation in the ongoing couples groups, they may be related to participation in the study, in which every family had a case manager to help with family challenges and facilitate referrals over 18 months. An ongoing focus on supporting family relationship challenges may have helped parents adopt a more collaborative, supportive, and organized approach to finding and keeping jobs to meet their family's needs.

In the benchmark study, with random assignment to experimental and control group and different change trajectories, we could infer that participation in the couples groups played a causal role in the positive outcomes. Here, we simply conclude that the pattern of pre- to postintervention comparisons in the current couples group parents is similar in pattern and quantitative measures of change to those in the benchmark RCT. If the current SFI couples group intervention was not affecting the results, it would be difficult to explain the consistent pattern of positive findings for the hundreds of families in both studies.

Who Benefits Most?

We noted earlier that for several of the measures, the adaptation score at baseline for the current couples group participants was significantly higher for those who completed postintervention assessments than for those who later dropped out. Two sets of findings argue against the hypothesis that the SFI intervention produces positive outcomes only for more well-adjusted participants. First, despite the fact that couples who dropped out tended to be in more distressed families, the range of scores and the shape of the distribution of scores was very similar in both the completer and dropout samples on all 11 measures. Second, within the completer sample, on all parent measures except fathers' couple relationship satisfaction, those with lower adaptation scores changed most in a positive direction. The fact that the mean scores for these measures were not close to either the lowest or highest possible scores (Table 1) indicates that these findings were not attributable to the fact that well-adapted participants had little room to improve. To the extent that we can attribute these results to participation in the intervention, it appears that the SFI intervention was especially effective for those at highest risk.

Limitations

If this study were the first attempt to evaluate the impact of the SFI interventions, the lack of random assignment to a no-treatment or low-dose control would constitute a serious limitation. Based on the results of the RCT study, we argued that random assignment to a no-treatment or low-dose control was not warranted on ethical grounds. Nevertheless, by evaluating similarities in patterns of change in two studies using the same interventions, this study supports the conclusion that the original positive results were not only repeatable but in two cases showed even stronger effects.

We found that there were no differences in change over time in this study as a function of ethnicity, but we do not know whether this finding results from the relatively lower number of African American and White couples in the study. We also note that four of the current intervention sites were in rural areas with large Mexican American populations, and one was in an urban site that contained most of our African American participants. The question of

whether the SFI intervention approach is equally appropriate for urban families of different ethnic origins is now being examined in intervention trials being conducted in the same five California counties, in other California counties, in Alberta, Canada, and in London, England.

In the measurement design for this study, we demonstrated positive changes in the quality of the couple relationship as reported by each parent, but we were not able to isolate effects on the couple's intimate relationship from effects on coparenting quality. It seems to us that the couples groups provided an opportunity to work directly on a positive coparenting relationship between mothers and fathers, but proof of this hypothesis awaits further research (see also Feinberg, Jones, Kan, & Goslin, 2010; McHale, 2007; Pruett & Barker, 2009).

The design of this study does not provide us with data concerning the mechanisms by which the curriculum of the intervention produces positive effects. Our hypotheses, supported by data from similar interventions with middle-class couples (C. P. Cowan & Cowan, 2000; P. A. Cowan et al., 2005), is that providing a safe environment in which parenting couples find support in exploring the connections among the key domains of family life while working on their own relationship challenges helps them discover ways to reduce the risks and increase the protective factors that affect their own and their children's adaptation. We expected that strengthening the relationship between the parents would lead to increased father involvement, more collaborative and effective coparenting, and lower parenting stress for mothers and fathers, with resulting benefits for children. We will test this hypothesis in future studies using path models that add observational data to the parents' reports.

CONCLUSIONS

Moving from a study of evidence-based practice in a benchmark study to practice-based evidence from this study, the data yielded support for our inference that the SFI couples approach has positive benefits for fathers' hands-on involvement in the care of their children, on the relationship between fathers and mothers, on both parents' parenting styles and parenting stress, and on their young children's behavior with them.

In regard to family intervention policy, the fact that focusing on couple relationships produces positive change in father involvement, couple relationship quality, and children's functioning suggests that it is time to reconsider the current tendency to offer programs for couple relationships, father involvement, and parenting in separate government and family service agency silos. Taken together, the results of the current and benchmark studies support the notion that combining the successful elements of programs focused on couple relationships, fatherhood, and parenting can have demonstrable benefits for mothers, fathers, and their children.

REFERENCES

- Abidin, R. R. (1997). Parenting Stress Index: A measure of the parent-child system. In C. P. Zalaquett & R. J. Wood (Eds.), *Evaluating stress: A book of resources* (pp. 277–291). Lanham, MD: Scarecrow Education.
- Adler-Bader, F., Callegas, A., Skuban, E., Keiley, M., Ketring, S. K., & Smith, T. (2013). Linking changes in couple functioning and parenting among couple relationship education participants. *Family Relations, 62*, 284–297.
- Caldwell, C. H., De Loney, E. H., Mincy, R. B., Klempin, S., Brooks, C. L., & Rafferty, J. (2011). Strengthening bonds between nonresident African American fathers and sons as a way to reduce or prevent youth risky behaviors. In C. Haen (Ed.), *Engaging boys in treatment: Creative approaches to the therapy process* (pp. 265–291). New York: Routledge/Taylor & Francis.
- Carlson, M. J., Pilkauskas, N. V., McLanahan, S. S., & Brooks-Gunn, J. (2011). Couples as partners and parents over children's early years. *Journal of Marriage and Family, 73*, 317–334. doi:10.1111/j.1741-3737.2010.00809.x
- Cowan, C. P., & Cowan, P. A. (1990a). *Couple Communication Questionnaire*. Berkeley, CA: Institute of Human Development.
- Cowan, C. P., & Cowan, P. A. (1990b). The pie. In J. F. Touliatos, B. F. Perlmutter, & M. A. Straus (Eds.), *Handbook of family measurement techniques* (pp. 278–279). Newbury Park, CA: Sage.
- Cowan, C. P., & Cowan, P. A. (1990c). Who does what? In J. F. Touliatos, B. F. Perlmutter, & M. A. Straus (Eds.), *Handbook of family measurement techniques* (pp. 447–448). Newbury Park, CA: Sage.
- Cowan, C. P., & Cowan, P. A. (2000). *When partners become parents: The big life change for couples*. Mahwah, NJ: Lawrence Erlbaum.

- Cowan, P. A., Cowan, C. P., Ablow, J. C., Johnson, V. K., & Measelle, J. R. (2005). *The family context of parenting in children's adaptation to elementary school*. Mahwah, NJ: Lawrence Erlbaum.
- Cowan, P. A., Cowan, C. P., & Heming, G. (1995). *Manual for the Child Adaptive Behavior Inventory (CABI)*. Unpublished manual, University of California, Institute of Human Development, Berkeley.
- Cowan, P. A., Cowan, C. P., Pruett, M. K., Pruett, K., & Wong, J. J. (2009). Promoting fathers' engagement with children: Preventive interventions for low-income families. *Journal of Marriage and the Family*, *71*, 663–679.
- Fagan, J. (2008). Randomized study of a prebirth coparenting intervention with adolescent and young fathers. *Family Relations*, *57*, 309–323.
- Feinberg, M. E., Jones, D. E., Kan, M. L., & Goslin, M. C. (2010). Effects of family foundations on parents and children: 3.5 years after baseline. *Journal of Family Psychology*, *24*, 532–542. doi: 10.1037/a0020837
- Freedman, B. (1987). Equipose and the ethics of clinical research. *New England Journal of Medicine*, *317*, 141–145.
- Gottman, J. M., & Katz, L. F. (1989). Effects of marital discord on young children's peer interaction and health. *Developmental Psychology*, *25*, 373–381.
- Heming, G., Cowan, P. A., & Cowan, C. P. (1990). Ideas about parenting. In J. Touliatos, B. F. Perlmutter, & M. A. Straus (Eds.), *Handbook of family measurement techniques* (pp. 362–263). Newbury Park, CA: Sage.
- Heyman, R. E., Sayers, S. L., & Bellack, A. S. (1994). Global marital satisfaction versus marital adjustment: An empirical comparison of three measures. *Journal of Family Psychology*, *8*, 432–446.
- Hinshaw, S. P. (2002). Preadolescent girls with attention-deficit/hyperactivity disorder: I. Background characteristics, comorbidity, cognitive and social functioning, and parenting practices. *Journal of Consulting & Clinical Psychology*, *70*, 1086–1098.
- Hunsley, J., & Lee, C. M. (2007). Research-informed benchmarks for psychological treatments: Efficacy studies, effectiveness studies, and beyond. *Professional Psychology*, *38*, 21–33.
- Knox, V., Cowan, P. A., Cowan, C. P., & Bildner, E. (2011). Policies that strengthen fatherhood and family relationships: What do we know and what do we need to know? *Annals of the American Academy of Political and Social Science*, *635*, 216–239. doi:10.1177/0002716210394769
- Markman, H., Stanley, S., & Blumberg, S. (2010). *Fighting for your marriage*. San Francisco, CA: Jossey-Bass.
- McHale, J. P., & Rotman, T. (2007). Is seeing believing? Expectant parents' outlooks on coparenting and later coparenting solidarity. *Infant Behavior and Development*, *30*, 63–81.
- Norton, R. (1983). Measuring marital quality: A critical look at the dependent variable. *Journal of Marriage and the Family*, *45*, 141–151.
- Parke, R. D. (2002). Fathers and families. In M. H. Bornstein (Ed.), *Handbook of parenting: Vol. 3: Being and becoming a parent* (2nd ed.; pp. 27–73). Mahwah, NJ: Lawrence Erlbaum.
- Pruett, M. K., & Barker, R. K. (2009). Effectively intervening with divorcing parents and children: What works and how it works. In M. Schulz, M. Pruett, & R. D. Parke (Eds.), *Strengthening couple relationships for optimal child development* (pp. 181–196). Washington, DC: APA Books.
- Pruett, M. K., & Johnston, J. R. (2004). Therapeutic mediation with high-conflict parents: Effective models and strategies. In J. Folberg, A. L. Milne, & P. Salem (Eds.), *Divorce and family mediation: Models, techniques, and applications* (pp. 92–111). New York: Guilford.
- Rienks, S. L., Wadsworth, M. E., Markman, H. J., Einhorn, L., & Etter, E. M. (2011, April). Father involvement in urban low-income fathers: Baseline associations and changes resulting from preventive intervention. *Family Relations*, *60*, 191–204.
- Tamis-LeMonda, C. S., & Cabrera, N. (Eds.). (2002). *Handbook of father involvement: Multidisciplinary perspectives*. Mahwah, NJ: Lawrence Erlbaum.
- Twenge, J. M., Campbell, W. K., & Foster, C. A. (2003). Parenthood and marital satisfaction: A meta-analytic review. *Journal of Marriage and Family*, *65*, 574–583.
- Wood, R. G., McConnell, S., Quinn, M., Clarkwest, A., & Hsueh, J. (2010). *Strengthening unmarried parents' relationships: The early impacts of building strong families*. Washington DC: Mathematic Policy Research.
- Yalom, I. D. (1995). *The theory and practice of group psychotherapy* (4th ed.). New York: Basic Books.