Controversies in Couple Relationship Education (CRE): Overlooked Evidence and Implications for Research and Policy

Philip A. Cowan and Carolyn Pape Cowan
University of California, Berkeley

Government interest in strengthening families in the United States and Great Britain has contributed to a rapid growth in couple relationship education (CRE) interventions, with a recent increase in programs for low-income families. We describe the policy contexts that initially led to increased support for CRE in both countries but now threaten its continuation. We summarize meta-analytic research and discussions of single studies by authors who draw opposing inferences about CRE effectiveness, often from the same studies. We discuss three sets of findings not featured in previous reviews, all of which focus on the potential benefits of CRE for the well-being of children. First, without intervention, average couple relationship satisfaction declines, with negative consequences for children. Second, including both parents in father involvement and parenting interventions results in value-added contributions to family functioning. Third, we describe 9 CRE intervention trials that include child outcomes, 8 of which support the assumption that CRE benefits children. These studies represent only a first step in determining what happens to children when their parents experience CRE. From both research and policy perspectives, there are too many positive findings to give credence to the claim that CRE programs should be discontinued and funding directed elsewhere. But there are too many negative findings to support the conclusion that CRE has been an unqualified success and that existing programs should be offered more widely. The negative findings and criticisms have much to teach us about potential modifications to CRE programs that will lead to more effective CRE approaches.

Keywords: couple relationship education, impact on couples, fathers, and children, moderator variables in meta-analyses, policy implications

Since the 1950s, and increasingly over the last two decades, concerns about the state of family life in the United States and the United Kingdom have led to the creation and proliferation of programs designed to strengthen couple relationships and increase fathers’ positive involvement with their children (P. A. Cowan, Cowan, & Knox, 2010; Panter-Brick et al., in press). We begin with an account of the U.S. and U.K. policy contexts that supported funding for intervention programs for couples, especially for those who are parents. The focus of this article is not on couples therapy but on the various prevention programs summarized under the general heading of Couple Relationship Education (CRE). Over the past decade, the question of whether CRE programs produce the desired effects has been the subject of many conflicting research articles, media stories, and Internet blogs, which make policy decisions challenging at best. A careful reading of the arguments suggested to us that the field was suffering from selective accounts of published data, often from the same studies, cited to support opposing conclusions concerning the benefits of CRE. Those who support CRE point to data showing small but significant effects on participants, at least in the short run, whereas those who criticize CRE focus on the small effects and the fact that the benefits often disappear over time. Policymakers who are willing to turn to systematic studies for evidence before making decisions about funding family programs are understandably confused. Public dialogue about the usefulness of CRE consists of a muddle of statements and counterstatements. We believe that this dialogue can be enhanced by paying more attention to the details—research designs and measurement approaches, specific outcomes in specific target populations, and the size of the intervention effects associated with each.

Although assessments of CRE studies rely increasingly on the results of meta-analyses, we suggest that combining a wide variety of CRE studies to obtain a few estimates of effect size glosses over or distorts critical information about the various factors that lead some programs to be effective and others to show few benefits for the participants. We conclude that we are not at the stage in which definitive generalizations can be drawn about whether CRE programs “work” until we can answer specific questions about how variations in central program features are related to outcomes for mothers, fathers, and children. We make suggestions for new studies that will inform policy decisions about strengthening relationships among family members in ways that foster parents’ and children’s development.

The Policy Context of CRE Programs

In the last half of the 20th century and continuing to the present day, concerns have been raised about changes in the structure and
quality of family life by social analysts, social scientists, family service providers, family clinicians, and politicians at all levels of government. Although there is controversy about the interpretation of social trends such as increasing divorce, single parenthood, father absence, cohabitation, and declining rates of childbirthing (Coontz, 2005; Parke, 2013), the shift away from so-called traditional family structures has often been interpreted as evidence that contemporary families are in a state of decline (Popenoe, 1993) in ways that place children at increased risk for developmental problems. In 1996, when the U.S. Temporary Assistance for Needy Families (TANF) program was established with passage of the Personal Responsibility and Work Opportunity Reconciliation Act, one of its four goals was “to encourage the formation and maintenance of two-parent families” (p. 8).

Influenced by the high levels of concern about families in general, and impressed by the U.S. Fragile Families study findings from a 20-city study of low-income women giving birth (Harknett, Hardman, Garfinkel & McLanahan, 2001), which revealed that a majority of fathers were still in romantic relationships with these women around the transition to parenthood, the first Bush Administration planned, and Congress authorized, awards of $150 million per year for 5 years for healthy marriage and responsible fatherhood programs within the Deficit Reduction Act of 2005, a reauthorization of the TANF program. The Administration for Children and Families of the U.S. Department of Health and Human Services distributed Healthy Marriage and Responsible Fatherhood Grants to state, local, and community-based service providers to address both marital instability and father absence. The Obama administration continued support for these activities with an additional round of grants distributed in 2010 and another round planned in 2014 or 2015.

The demographic family changes documented in the United States have also been occurring in families in the United Kingdom over the same period of time, with an even greater movement away from marriage, a large increase in the incidence of single women becoming mothers, and increases in both cohabitation and divorce (Statistics, 2007). At first, the policy response by the U.K. government was similar to that of the U.S. In 1996, government revisions of family law mandated the giving of grants for marriage support and the publication of the Marriage Support Services Directory, but the trajectory leading toward the support of couple relationship stability was not as linear as it has been in the United States. A major policy shift instituted by the British Labour government resulted in decreased support for services for couples (Boucher, 2008), and Every Parent Matters, a government policy document issued by the British Department for Education and Skills (2007), set out a new series of plans and programs for families that focused more on parenting than on the relationship between the parents, with the goal of increasing the involvement of fathers in children’s lives, and providing parenting skills training for parents. More recently, U.K. policy directions shifted once again as the Conservative–Liberal coalition government returned to a focus on promoting marriage, strengthening couple relationships, and encouraging fathers’ family involvement.

Based on similar demographic changes in the structure of families, the United States and the United Kingdom have currently arrived at similar policy decisions by somewhat different routes to fund programs that will increase family relationship quality and stability by strengthening couple relationships and increasing fathers’ family involvement. The main hope in each country was to reduce the incidence of single parenthood and relationship breakup in order reduce poverty and create more supportive environments for children.

We acknowledge at the outset that as creators or co-creators of three programs that embody elements of both couple strengthening and father-involvement interventions (C. P. Cowan & Cowan (2000); P. A. Cowan, Cowan, Ablow, Johnson & Measelle (2005); P. A. Cowan, Cowan, Pruett, Pruett & Wong (2009), we have some stake in the conclusions reached about program outcomes.1 In this review, we have made every effort to recognize conflicting opinions and to present a differentiated picture of an emerging field in which there is enormous heterogeneity of programs and results that defy simple aggregation in meta-analyses. We find evidence for both optimism and skepticism about the impact of CRE programs, and ambiguities that lead to a host of as-yet unresolved questions. We conclude that the evidence produced so far, although not always demonstrating clear and long-lasting intervention effects, reveals promising trends. Relatively new research, not previously included in discussions of CRE effectiveness, points to the value-added contributions to family life of couple-focused interventions. Also left out of current CRE controversies are studies showing that children benefit when their parents participate in interventions to strengthen their relationships as couples or coparents. This idea has been used as a justification for providing these programs, but has rarely been tested in systematic research. We describe eight studies with nine intervention trials in which all but one show positive effects on one or more measures of the children’s adaptation. We argue that these findings help to justify further efforts to design research that will provide a more nuanced picture of what kinds of CRE interventions work for which target populations.

The Current State of Controversies About CRE

Couple relationship strengthening programs emerged in the 1960s along with many other preventive interventions, stimulated by a zeitgeist of political and social science optimism that early interventions could prevent individuals and families from later distress and dysfunction. Initially, most programs were targeted to middle-class couples early in their marital careers and not already in marital or relationship distress (e.g., engaged, recently married, making the transition to parenthood). More recently, there has been a concern with providing services for couples who are at risk by virtue of low incomes but not otherwise identified as experiencing serious relationship difficulties.

CRE interventions typically involve group meetings, ranging from a single informational presentation, to weekend workshops, to ongoing weekly meetings spread over 3 or 4 months. Most CRE programs have focused on improving couple communication, although a very few also focus on parents’ individual well-being, coping with stress, parenting, and/or intergenerational issues. With the exception of a discussion of the results of adding fathers to therapeutic interventions for mothers and children, we have not

---

1 Our stake is intellectual, not financial. We have not been involved in any way in merchandising the interventions that we have created or benefiting financially from the sales of material or curricula created by other CRE programs.
included studies of _therapy_ for troubled couples or _therapy with parents_ of a child in treatment for diagnosed behavior problems (e.g., attention-deficit hyperactivity disorder, aggression), because the risk status of the participants, the intervention formats, and the outcomes are not comparable with those of CRE, which focus mainly on prevention and typically use a group rather than an individual or couple-by-couple format. Nor did we include evaluations of _parenting classes_, because they range widely in content, length, and leaders’ qualifications; are attended primarily by mothers; and rarely deal with couple relationship issues.

There have been thousands of couples groups conducted over the last 50 years, with the bulk of them springing up in the last two decades. The vast majority produced no published reports of program effects, either pro or con, including almost all of the programs funded in 2005 and 2010 by the Administration for Children and Families (for two important exceptions, see Hsueh et al., 2012; Wood, McConnell, Quinn, Clarkwest, & Hsueh, 2010). From meta-analytic studies, we estimate that there may be as many as 150 systematic studies of CRE, with about one third to one half being doctoral dissertations. The number of studies with random control trial (RCT) designs is quite small, and most meta-analyses include no more than 25 studies with follow-ups beyond an immediate posttest, and fewer than 10 with follow-ups beyond 7 months postintervention.

Almost all studies cited in articles on CRE research were conducted in the United States. There are isolated exceptions. CRE programs based on the Prevention and Relationship Enhancement Program (PREP) created and evaluated in the United States (Markman, Stanley, & Blumberg, 2010), have also been evaluated in Germany (Hahlweg & Richter, 2010), Denmark (Van Widenfelt, Hosman, Schaap, & van der Staak, 1996), and Sweden (Engsheden, Fabian, & Sarkadi, 2013). Programs with more emphasis on emotional communication, conducted in France by Bodenmann and colleagues (Bodenmann, Pihet, Cina, Widmer, & Shantinath, 2006) and in Australia by Halford and colleagues (Halford, Petch, & Creedly, 2010), have also been described by their authors as successful. The situation in the United Kingdom is puzzling. As a result of the government’s recent allocation of £30 million ($50,400,000) for relationship support, many organizations with long-standing interests in couples have expanded their programs (e.g., Tavistock Centre for Couple Relationships, Relate, One Plus One, Marriage Care), but until very recently, with the exception of marriage preparation classes, most of their efforts have been devoted to couples counseling, therapy, or education on a couple-by-couple basis for already troubled partners; all of these have been reviewed as having quite successful outcomes (Spielhofer et al., 2014). While the Spielhofer et al. (2014) and other reviews of couple interventions in the United Kingdom (Chang & Barrett, 2008; Coleman & Glenn, 2009) often cite American CRE studies, professionally led _groups_ for couples with children have generally not been offered in England until recently.

**Reviews of CRE Programs**

The most comprehensive qualitative summary of CRE programs that we have seen (Markman & Rhoades, 2012) includes 31 studies, almost all of which are interventions with middle-class families, and 21 of which were RCTs. Another review of 15 couples-focused programs (P. A. Cowan et al., 2010) included 13 RCTs, four with low-income participants. Halford and Bodenmann (2013) reviewed 17 RCT studies with follow-ups of at least one year. These and other literature reviews (e.g., Halford, 2011, who includes a number of European intervention trials) concluded that relationship education programs show small but statistically significant effects on marital quality, measured mostly by participants’ self-reports, and in a few cases, by observer measures of couple interaction quality. Halford and Bodenmann (2013) went beyond the assertion that the preponderance of studies showed significant intervention effects to assert that couples with elevated modifiable risk factors benefited substantially, whereas benefits for low-risk couples were not as clearly established.

Descriptive research reviews are valuable at the beginning of a new field, or new directions in a field, to acquaint readers with current research and intervention directions, and to provide a map of what has been done and what still needs to be done. However, descriptions of research studies contain a number of problems. First, there is no way of knowing how selective the author has been in reporting studies that support his or her point of view. Second, there is the “file drawer problem,” in which studies that fail to find statistical significance are less likely to be submitted and, if submitted, less likely to be published. Third, until very recently, significance at the _p < .05_ level has been accepted as the standard for determining whether an observed effect is likely to be attributable to participation in the intervention or is simply a chance finding, but until the increasing popularity of meta-analysis, there was no way to provide an overall summary of how strong an effect was attained by the average CRE intervention. Our review here attempts to summarize what has been learned from seven meta-analyses of couple relationship programs published during the past 11 years, five by Alan Hawkins and his colleagues.

Meta-analyses begin with an exhaustive literature search of adequately designed studies, including both published and unpublished findings, followed by procedures that produce a statistic that reflects the average effect size (typically expressed as Cohen’s _d_). The inclusion of unpublished studies is intended to guard against the “file drawer problem.” We have some concerns about how this bias has been interpreted, which we discuss later in this article.

Articles that evaluate the merits of CRE usually refer to one or two meta-analytic studies, either to assert support for the effectiveness of couples-group interventions (e.g., Markman & Rhoades, 2012), or to cast a skeptical eye on the endeavor by arguing that the effect sizes are small and tend to disappear over time (e.g., P. Cohen, 2014; Johnson, 2014). A more differentiated discussion can be stimulated by examining this body of work in detail and by considering the pattern of effect size results across the studies, as we do in Table 1 and the discussion that follows. In addition to presenting effect sizes for the impact of CRE interventions on couple relationship satisfaction and communication, when available, we include data concerning potential moderating variables that make a difference to the size of the obtained effect—

---


3 There was one meta-analysis published before 2000 (Butler & Wampler, 1999), but it contained 16 RCT studies of one intervention approach—the Couple Communication Program. Only four of the studies had been published. Effect sizes were given but not tested for statistical significance.
<table>
<thead>
<tr>
<th>Target population</th>
<th>Source of data</th>
<th>Dosage</th>
<th>Publication</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Immediate follow-up</td>
<td>Posttest follow-up</td>
<td>Low</td>
</tr>
<tr>
<td></td>
<td>Sat</td>
<td>Com</td>
<td>Sat</td>
</tr>
<tr>
<td>Hawkins et al. (2008)</td>
<td>Inclusive</td>
<td>.36*** (46)</td>
<td>.44*** (37)</td>
</tr>
<tr>
<td>Immediate OR follow-up</td>
<td></td>
<td>.24 (17)</td>
<td>.54** (13)</td>
</tr>
<tr>
<td>Blanchard et al. (2009)</td>
<td>Inclusive</td>
<td></td>
<td>.30*** (69)</td>
</tr>
<tr>
<td>Hawkins et al. (2012)</td>
<td>Inclusive</td>
<td></td>
<td>.22 (17)</td>
</tr>
<tr>
<td>Carroll &amp; Doherty (2003)</td>
<td></td>
<td></td>
<td>.99*** (7)</td>
</tr>
<tr>
<td>Fawcett et al. (2010)</td>
<td>Premarital couples</td>
<td>.09* (23)</td>
<td>.28** (19)</td>
</tr>
<tr>
<td>Pinquart &amp; Teubert (2010)</td>
<td>Transition to parenthood</td>
<td></td>
<td>.25 (3)</td>
</tr>
<tr>
<td>Hawkins &amp; Fackrell (2010)</td>
<td>Low-income couples</td>
<td></td>
<td>.29** (10)</td>
</tr>
</tbody>
</table>

Note. Sat = Relationship satisfaction; Com = Couple communication; Numbers in parens = number of studies included in the meta-analysis.
*p < .05. **p < .01. ***p < .001.
target population, timing of follow-up, source of data, dosage, and whether the intervention results had been published at the time the meta-analysis had been conducted. Although meta-analyses of CRE programs usually include data from quasi-experimental and single-sample pre–post studies, we include only the results of studies with random assignment to experimental and control conditions—with one exception, which is Hawkins and Fackrell (2010). Of the 15 studies in Hawkins and Fackrell’s meta-analysis, only three were RCTs, and so for this meta-analysis only, we include the 12 pre–post design studies in our table.

Table 1 presents the effect size (Cohen’s $d$), significance level, and numbers of studies included in the effect size calculation ($k$) of each of seven meta-analyses. The top three meta-analyses (in Table 1) aggregate a large number of studies of different participant populations. The next four include fewer studies and report effect sizes of CRE interventions with targeted populations (pre-marital couples, couples becoming parents, or low-income couples). Blank cells signify that there was no available information, and merged cells represent cases in which couple satisfaction and communication scores have been combined in the analysis. It is conventional to regard a $d$ of .2 as small, .5 as medium, and .8 as large (J. Cohen, 1988).

Maintenance of intervention effects over time. The three largest meta-analyses by Hawkins and colleagues (Blanchard, Hawkins, Baldwin, & Fawcett, 2009; Hawkins, Blanchard, Baldwin, & Fawcett, 2008; Hawkins, Stanley, Blanchard, & Albright, 2012) attempted to include a wide variety of CRE interventions, but because of what was available in the literature at that time, the participants were mostly young middle-class couples in the early stages of forming their families. In the earliest meta-analysis, Hawkins et al. (2008) showed that when studies were included that had either an immediate or longer term follow-up, there were statistically significant effects on immediate posttest relationship satisfaction ($d = .31$) and communication ($d = .45$) in favor of the intervention participants. A subset of these studies had both immediate and follow-up posttests, that is, they reported results on the same participants followed over time. In contrast with the immediate posttests, the effect sizes of the longer term follow-up assessments (satisfaction, $d = .28$; communication, $d = .37$) were no longer statistically significant. Similarly, at the immediate posttest, Pinquart and Teubert’s meta-analysis found small but statistically significant effects of the intervention on couple relationship satisfaction ($d = .09$) and communication ($d = .28$), but the longer term posttests failed to show statistically significant results. These findings suggest that the early intervention effects disappear over time.

The results of two meta-analyses that also contrasted immediate and longer term posttests challenge the conclusion that CRE effects are not maintained over time. With a sample of premarital couples enrolled in CRE, Carroll and Doherty (2003) found a large effect of CRE on combined measures of satisfaction and communication immediately after the intervention ($d = .99$), with a smaller but still significant moderate-sized effect on follow-ups later on ($d = .64$). Blanchard et al. (2009) noted that the Hawkins et al. (2008) meta-analysis had simply divided postintervention assessments into immediate and follow-up, without specifying the timing of the follow-ups, which was often very soon after the immediate posttests. They identified five studies with follow-ups more than 7 months after the completion of the intervention, and found that the longer term effect size of .59 was statistically significant and in the moderate range. These two meta-analyses, then, could be cited in support of the conclusion that although intervention effects may be smaller at longer term than at immediate posttests, they are maintained for some not-well-specified period of time.

The remaining three meta-analyses (Fawcett, Hawkins, Blanchard, & Carroll, 2010; Hawkins & Fackrell, 2010; Hawkins et al., 2012) did not compare immediate and longer term posttests, but reported results only for posttests at some point beyond the end of the interventions. All three found statistically significant effect sizes for communication ($d_s = .40, .45, .41$), and two also found significant effects for relationship satisfaction ($d_s = .30$ and .29). Going beyond measures of couple satisfaction and communication, Pinquart and Teubert (2010) reported large longer term follow-up effects on relationship stability ($d = .81$) and parental mental health ($d = .81$).

In sum, the data from seven meta-analyses of RCT studies of couple relationship interventions suggest that immediate posttests generally reveal small but statistically significant intervention effects on measures of couple satisfaction and communication, although Carroll and Doherty’s (2003) immediate posttest effect is large. In meta-analyses in which satisfaction and communication effect sizes were reported separately, the intervention effects on communication were larger than those for satisfaction, but none of the investigators tested this difference for statistical significance between the two outcomes. The question of whether CRE intervention effects are maintained over time (i.e., whether the effect size for longer follow-ups is statistically significant) has not been resolved. Rather than searching for a “yes” or “no” answer, we suggest that CRE researchers and program planners explore modifications that might prolong the immediate intervention impact (e.g., adding a focus on coparenting, booster sessions, additional interventions to reduce external stressors and increase social supports).

Moderator effects. The $d$ statistic represents an average of effect sizes across studies. Rather than assuming that the variation around the mean is attributable to “error,” all of the investigators who report on CRE meta-analyses search for systematic factors that might identify characteristics of interventions that work more or less well (i.e., moderator effects). The results of moderator analyses have not been given the attention they deserve. In addition to the timing of follow-ups, which we have discussed, potential moderators in the seven meta-analyses include the source of data (self-report vs. observation), dosage (number of sessions or hours of group meetings), qualifications or experience of the interveners, and whether the study included in the analysis has been published.

Of the three meta-analyses that examine differences in effect size attributable to the source of data (Blanchard et al., 2009; Fawcett et al., 2010; Pinquart & Teubert, 2010), the first two find that parent reports show statistically insignificant effects, and the third (Fawcett et al., 2010) shows a small, statistically significant effect ($d = .31$). By contrast, all three show large effect sizes on observed communication skills ($d = .83, 1.02, .80$). In addition, we have Pinquart and Teubert’s (2010) finding, cited earlier, that there were large effects on variables not usually measured in CRE studies (relationship stability and parental mental health). Note that the results from these moderator analyses do
not distinguish between immediate and longer term follow-ups, and so do not address the issue of maintenance of intervention effects over time.

One might expect that more intense interventions (more hours of group meetings) would produce stronger effects. Two of the meta-analyses that examined this question in primarily middle-class samples with heterogeneous target populations (Hawkins et al., 2008, 2012) supported this expectation. Both reported nonsignificant effects on couple satisfaction and communication at longer term follow-ups for low-dose interventions (up to 7 or 8 hr), with significant moderate to high effects for satisfaction (d = .52 and .38) and communication (d = .70 and .41) in moderate-dosage interventions (up to 20 hr). Pinquart and Teubert’s meta-analysis of couples in transition to parenthood found no significant effect sizes for low-dose interventions, and a significant effect only on communication for high-dose interventions. Finally, the Hawkins and Fackrell (2010) meta-analysis of CRE programs for low-income families found nonsignificant effects when the sample of three RCT studies was divided into one low-dose and two moderate-dose interventions (too small a sample to draw conclusions), and statistically significant effects for the six pre–post test design studies for both low- and high-dose interventions (d = .26 and d = .34). Here, as elsewhere, the findings are mixed. Across meta-analyses, then, there is some indication that dosage may have an effect on outcomes, but as we indicate in our suggestions for future research, there are design issues in determining dosage effects that have yet to be addressed.

Three meta-analyses compared effect sizes for published and unpublished studies. Two (Fawcett et al., 2010; Hawkins et al., 2008) found that the effect sizes for unpublished studies were negligible (d = .02 or .03), whereas the effect sizes for published studies were moderate or large (d = .70 and .99). Pinquart and Teubert’s (2010) meta-analysis of interventions for couples in transition to parenthood found no significant effects on measures of couple satisfaction for either published or unpublished studies, but did find that published studies reported a significant impact of CRE on couple communication (d = .36). These results suggest that the overall estimates of CRE effect size that we have cited would have been larger if only published studies had been included in the meta-analyses.

More thought needs to be given to the file drawer problem associated with the difference between published and unpublished studies. The typical interpretation of a publication “bias” is that the body of published studies presents an exaggerated picture of intervention effectiveness because unpublished, unsuccessful intervention studies stored in file drawers are hidden from public view. Including unpublished studies in the meta-analyses is the rationale for reducing the bias toward intervention effectiveness in published studies. However, it is likely that there are also biases associated with unpublished intervention studies not simply attributable to their lack of statistically significant findings. For example, a high proportion of unpublished CRE studies are reported in dissertations. Others may have been submitted and rejected. It is possible that these studies are not administered as competently, and do not use as sophisticated designs and measures as published studies do. If this hypothesis is correct, the inclusion of unpublished studies in meta-analyses would minimize estimates of the effectiveness of adequately mounted CRE programs. One finding from the Fawcett et al. (2010) meta-analysis supports this interpretation: In a search for moderator effects, they found that although both self-report and observational measures showed statistically significant impacts, the effects on observational measures were appreciably larger, and that the difference between published and unpublished studies in their meta-analysis might be explained by the fact that none of the unpublished studies (mostly dissertations) used observational measures. It is possible, then, that many of the studies residing in file drawers deserve to remain where they are and not be used to dilute overall estimates of CRE effects. At the least, all meta-analyses of CRE should report results separately for published and unpublished studies, and report on indicators of whether unpublished studies are equivalent in quality to those that have been published.

One important moderator effect was tested only in a single meta-analysis and was not listed in Table 1 for reasons of space. Pinquart and Teubert’s (2010) meta-analysis of CRE interventions for couples in transition to parenthood examined differences in effect size for couples groups led by paraprofessionals and groups led by leaders with professional clinical qualifications. Significant effect sizes for both satisfaction (d = .23, k = 11) and communication (d = .56, k = 9) were found only when the leaders had professional clinical training.

In sum, the pattern of moderator effect sizes for CRE interventions is quite different from the overall findings of small effects when the question is whether the interventions create measurable differences between experimental and control groups. According to the results summarized in Table 1, an array of effect sizes ranging from low through medium to high is obtained with (a) longer term follow-ups, (b) studies using observational measures, (c) moderate rather than low-dose interventions, and (d) professionally trained group leaders. Several cautions come with this general statement. First, the effect sizes for each moderator are averages. Some of the interventions in the category with the more positive characteristics do not produce desirable results. Second, moderator analyses are performed one variable at a time. We need studies that combine potential moderator variables to find out whether there are additive or even synergistic effects of the moderator effects we have described here and whether they combine with other potential moderators that have yet to be examined systematically.

Another possible caution in interpreting the fact that the same moderator effects can be found in more than one meta-analysis is that there is considerable overlap in the reference lists included in the articles listed in Table 1. Table 2 describes the overlap in terms of the proportion of references included in the later publications (moving from left to right, or top to bottom) that had also been cited in the earlier studies. The largest overlap (67%) is between Hawkins et al. (2008) and Blanchard et al. (2009), essentially the same research group publishing meta-analyses in adjacent years. Although they included some new studies and dropped some from the earlier article, the Blanchard et al. article did not simply repeat the analysis of overall effect sizes for immediate and follow-up posttests. Instead, they intended to make new points about (a) the importance of how long the follow-ups were conducted after the intervention was completed, and (b) the fact that observations of the couple revealed greater intervention effect sizes than partners’ self-reports about communication quality. The third meta-analysis (Hawkins et al., 2012) did not provide a reference list of included studies. The four remaining meta-analyses focused on specific
target populations and ranged from moderate overlap to none at all. While some of the replicated moderator results come from meta-analysis publications with overlapping studies, others appear in articles with different target populations and no studies in common. We conclude that there is no clear evidence that the pattern of findings in Table 1 is primarily attributable to the overlaps among studies included in the seven meta-analyses, but future reviews and meta-analyses should pay attention to the proportion of studies in their analyses that have already been included in earlier reports.

Focus on Single Studies: Johnson’s Critique of CRE

Pointed questions about whether CRE programs are effective, especially for low-income families, have been raised recently by Johnson (2012, 2014), who asserted that the government should not be funding marriage and relationship programs for low-income families. His criticism, based primarily on his interpretation of three large-scale, well-designed government-funded programs, elicited equally critical responses in defense of CRE (Hawkins, 2014; Hawkins et al., 2013). In the exchanges between Johnson and Hawkins, we can clearly see the polarization in the field.

Building Strong Families (BSF). The first large CRE program, BSF (Wood, McConnell, Quinn, Clarkwest, & Hsueh, 2010), included three different couple relationship intervention approaches conducted across eight U.S. sites, with 5,102 low-income unmarried couples randomly assigned to intervention and control conditions. Outside of their low-income status, couples were not identified as having relationship difficulties. Note that this intervention and the next one we describe, Supporting Healthy Marriage, for low-income married couples, used support workers in addition to the group experience to help partners with problems associated with poverty. Johnson (2012, 2014) emphasized the fact that BSF showed no statistically significant effects across the eight sites. Hawkins and colleagues (2012) countered that Johnson failed to acknowledge that the couples-group interventions did show significant positive effects for African American participants (more than half the sample) at the 15-month follow-up. Compared with controls, African American couples increased their level of support and affection, and their use of constructive and avoidance of destructive conflict management techniques. At the 15-month assessment, African American participants were also significantly less likely to report infidelity and intimate partner violence (no effect sizes given).

Hawkins and colleagues (2012) also pointed to the positive early findings in Oklahoma, one of the eight BSF sites. In comparison with couples in the control group, group participants in the Oklahoma site reported significantly more satisfying couple relationships ($d = .20$), greater use of constructive conflict behavior ($d = .17$), less use of destructive conflict behavior ($d = .09$), and higher coparenting quality ($d = .12$). Mothers reported fewer depressive symptoms ($d = .22$) and were more likely to report that fathers provided substantial financial support ($d = .27$). These details explain why Hawkins and others were heartened by the initial results, despite the lack of statistically significant overall effects at the eight sites. Johnson (2014) responded that Hawkins et al. omitted the fact that in another of the eight sites in BSF (Baltimore), couples-group participants fared significantly worse than the controls. He also noted that in a subsequent follow-up at 3
years postintervention (Wood, Moore, Clarkwest, & Killewald, 2014), there were no longer significant effects on couple relationship quality for African American participants or for couples at the Oklahoma site.

Turning to another issue, Hawkins et al. (2012) noted that BSF used an intent-to-treat analysis that included all couples assigned to intervention and control groups, despite the fact that actual attendance was very low, with only 55% of the couples attending even one group meeting. Johnson (2012, 2014) countered that intent-to-treat analyses are standard in the field and the only way to avoid selection bias that would inflate intervention effects, because couples who refuse to participate or who drop out early remove potentially low-functioning participants from the intervention condition. We agree with Johnson on merits of intent-to-treat analyses, but if 45% of couples who come in for an initial interview fail to attend the program offered at that site, there may be implementation flaws that contributed to the fact that couples who did attend received no measurable benefits.

Only two tiny rays of hope for CRE emerged at the final 3-year follow-up of BSF families (Wood et al., 2014). Although the impact on couple relationship satisfaction in the Oklahoma site did not hold up over time, 3 years after the families entered the study, the children of couples-group participants were significantly more likely to have parents still together. Furthermore, as we show in the section on child effects, across the eight sites, there was an overall slight, but statistically significant, positive impact of CRE participation on parents’ reports of children’s problem behavior. There is no question that the BSF findings were a disappointment to the investigators and to those who advocate the widespread dissemination of couple relationship interventions for low-income families. Johnson’s (2012, 2014) view is that the fading of even the few positive effects over time justifies throwing in the towel on this kind of intervention, at least from the point of view of large-scale government support. We believe that the failure of this massive study to produce meaningful long-term change in the participating couples is, understandably, responsible for much of the current skepticism about government funding of CRE interventions.

Supporting Healthy Marriage (SHM). The other two examples cited by Johnson in his criticism of CRE (2014) are not as on-point as his presentation implies. The SHM program (Hsueh et al., 2012) was a new RCT with 6,298 low-income married couples at eight sites assigned randomly to one of four couples-group programs plus a family support worker, or a no-treatment control condition. Johnson (2014) dismissed the results of this trial because it had only several “very small positive intervention effects” (p. 302). Hawkins (2014) responded that although the intervention effects were small, they were seen not only in self-reports at 12-and 30-month follow-ups but also in observational data collected at the first follow-up.

At the 12-month follow-up, compared with those in the control group, program group members showed statistically significant effects on higher levels of marital happiness (d = .13), lower levels of marital distress (no effect size given), greater warmth and support, (d = .09), more positive communication skills (dmen = .08; dwomen = .11), and fewer negative behaviors and emotions in their interactions with their spouses (dmen = .08; dwomen = .12). They also reported less psychological abuse (no differences in physical abuse). The effect sizes for self-report measures at the 30-month follow-ups were equal to or slightly greater than those obtained 18 months earlier.

Furthermore, observational measures of couple interaction (done only at the 12-month follow-up) indicated that for program couples, on average, both partners showed more positive communication skills (dmen = .10; dwomen = .09), and wives showed significantly less anger and hostility (d = .10), than partners in the control group. Taken together, the self-report and observational measures of intervention effects suggest that the SHM program changed not only the way participants viewed their marriages but also the extent to which they were able to implement the skills taught by the SHM curricula in ways that were observable by others. Describing these effects as small does not entitle a critic to dismiss them entirely.

Finally, in contrast with 55% of couples in the BSF study who attended at least one meeting, 83% of the couples in the SHM study attended at least one meeting, with an overall attendance rate of just over 70%. This raises the question of whether the SHM results were more positive than BSF, because the intervention was addressed to married rather than unmarried couples, or whether the differences resulted from lessons learned about implementation by the later program designers from the earlier investigators. Wood et al. (2014) speculate that the difference in populations could explain the results; cohabiting partners may be less committed to each other, less trustful of each other as they enter the study, and less willing and able to do the hard work of applying what is learned in group meetings to their relationship. We lean toward an explanation based on incremental progress in the field of CRE implementation. Only future studies can enlighten us about this issue.

Community Healthy Marriage Initiative. In our view, the third program, the Community Healthy Marriage Initiative, a large quasi-experimental program (Bir et al., 2012) cited by Johnson as representing a failure of CRE, did not actually provide quantitative assessments of CRE participation. Media messages and an array of relationship skills classes and service referrals were delivered to three targeted cities with large low-income populations, and results were compared with those in three matched control cities (therefore, this was not an RCT). Surveys were administered to a representative sample (more than 77,000 individuals) to assess attitudes about marriage and family life, household and marital status, quality of couple relationships, income and earnings, and child well-being. There were no differences in the modest rates of participation in relationship classes (6% to 8%) between targeted and comparison cities, and no differences in any of the self-report survey measures. Hawkins (2014) noted that the study had design flaws and discussed it no further. We can add that the only direct assessment of CRE participants in the Bir et al. study was conducted in qualitative interviews with 750 individuals (not couples) who attended a class, 80% of whom reported that the classes improved their relationships with their partners, and 97% of whom reported that they would recommend the classes to others. While we do not believe that consumer satisfaction measures constitute proof of CRE effectiveness, the limited data here are not consistent with a “no effects on the participants” verdict. Our conclusion is

4 The authors were two of a large group of academic consultants on this project.
that this was a study of the impact on a community of a number of programs designed to encourage healthy marriages, not an RCT to evaluate the impact of CRE programs.

In terms of the controversy between Johnson and Hawkins, we believe that Johnson was justified in raising the point about government spending in light of the failure of BSF to show overall significant effects of CRE for unmarried couples. On the other hand, we support Hawkins’s point that CRE should not be generally dismissed because of the failure of BSF with unmarried couples, or based on the fact that the effects of SHM on married couples were “small.” As we have noted, data from other single studies and meta-analyses with moderators show effect sizes ranging from small to medium, and occasionally large. The overall pattern of findings indicates to us that it is time to stop arguing about whether CRE programs are generally effective, and to find out more about the characteristics of programs that have shown positive effect on the couples who participate.

Critiques of CRE based on other small studies. The Johnson–Hawkins et al. controversy is not the only one in which questions have been raised about the efficacy of CRE. Findings from three other intervention studies have been used to support claims that CRE programs have no discernible effect (Doss, Rhoades, Stanley, & Markman, 2009; Markman, Rhoades, Stanley, Ragan, & Whithon, 2010; Rogge, Bradbury, Hahlweg, Engl, & Thurmaier, 2006). The form of the three studies is very similar. Doss et al. (2009) and Markman et al. (2010) followed up an intervention study by Stanley et al. (2001), whereas the Rogge et al. (2006) study is based on a German trial of PREP by Hahlweg and colleagues (Hahlweg, Markman, Thurmaier, Engl, & Eckert, 1998). In the early follow-ups, both studies found statistically significant intervention effects. In the later follow-up articles, the authors predicted postintervention outcomes based on preintervention data, and included intervention participation as a control variable. All three found that this statistical control did not alter the strength of the predictions. That is, pretest data predicted posttest data equally well for participants in the CRE and control groups. From these findings, the conclusion has been drawn (e.g., Bradbury & Lavner, 2012) that the intervention has no effects, a conclusion based on faulty logic. The finding of no intervention effects on pre–post correlations is completely independent of whether there are intervention effects on participants.

Four studies raise additional questions about whether couples benefit from CRE. Laurenceau, Stanley, Olmos-Gallo, Baucom, & Markman (2004) compared a university-based PREP program with a religious organization PREP program and a no-treatment control, and followed couples for 14 months after they entered the study. Growth curve analyses revealed no effects of participation in any of the programs on self-reported marital satisfaction; this has been cited as a failure of CRE, but there were some effects of the religious organization PREP compared with no-treatment controls on wives’ observed communication (d = .60) and positive behavior (d = .55).

Trillingsgaard, Baucom, Heyman, & Elklit (2012) offered Danish couples a PREP intervention for a total of 17.5 hr, an information-based parent program (INFO) for a total of 11 hr, and compared them 24 months later with couples receiving no intervention. All three groups declined in marital satisfaction and there were no differences in self-reported communication measures—a blow to CRE advocates. This was not a random assignment study, however, as the no-intervention controls were neither invited to participate in this study nor informed about the possibility of intervention.

Rogge, Cobb, Lawrence, Johnson, and Bradbury (2013) issued a more serious challenge to CRE. Engaged and newlywed couples (n = 174) were randomly assigned to a PREP intervention teaching communication skills and conflict management, or a CARE program teaching acceptance, support, and empathy—each conducted in four meetings for a total of 15 hr. Both were compared with couples receiving a one-session relationship awareness (RA) intervention (instructions to watch and discuss four movies with intimate relationship plots within the next month) and couples receiving no intervention. Follow-ups conducted semiannually over 3 years assessed relationship stability and satisfaction. There were no differences among couples in the PREP, CARE, and RA groups on rates of dissolution or on a global index of relationship satisfaction. The surprising finding was that RA couples—who received information about attending to and maintaining their relationships but no direct skills training—had similar outcomes to couples who received skills training. The fact that in the randomized part of the study (assignment to the three intervention conditions) a 1-hr group meeting plus watching and discussing four movies had the same impact on relationship satisfaction as a 15-hr PREP or CARE intervention is disquieting for supporters of CRE.

Finkel and colleagues (Finkel, Slotter, Luchies, Walton, & Gross, 2013) raise a similar issue about shorter interventions for couples, this time without using a couples-group intervention. One hundred twenty couples were enrolled in a study of marital quality and marital conflict for 2 years, during which they were assessed every 3 months, including a task to write about a recent argument with their partner and to rate their level of anger. No interventions were conducted during the first year. In the second year, half of the couples were randomly assigned to a reappraisal condition, in which they were asked at Months 12, 16, and 20 to write about a recent conflict, describe it from the perspective of a third person, and rate how angry they were. Growth curve analyses indicated that all couples (on average) declined in marital quality during the first year, but that the decline was halted for the intervention participants during the second year. Both Rogge et al. (2013) and Finkel et al. remind us that CRE is not the only way to prevent the normative decline in relationship satisfaction. Further studies are needed to study whether the different approaches to couples intervention have different effects for different kinds of couples.

We have been discussing a set of descriptive reviews, meta-analyses, and critiques of findings from single studies that have been interpreted differently by CRE supporters and critics. We believe that it is premature at this point to come to a definitive conclusion about CRE’s general effectiveness for couples. Although some studies do show a strong impact of couples-group interventions on couple relationship quality, the overall average in meta-analyses that include unpublished studies is consistently small at immediate posttest and in question at longer term follow-ups. However, it is also too soon to declare that CRE is moribund, and that research and service funds should be allocated in other directions. Enthusiasm for CRE has certainly been muted by the failure of the large BSF intervention to produce measurable short-term and long-term effects. The fact that some single studies do not show significant effects, and a few others achieve positive outcomes for couples with interventions much less complex than
CRE, does raise serious questions for advocates of couples-group programs. In response, we have noted that some studies cited as proof that CRE is not effective turn out to have flaws that reduce the power of the criticism. Furthermore, tests of moderator effects in meta-analytic studies indicate that some CRE programs do have substantial effects. The main problem in evaluating the usefulness of CRE is not that we lack research studies, but that for almost every claim that CRE has positive or negative effects on couples, there has been a counterclaim or a reason why the study or analysis under discussion does not support the author’s position on CRE. That is, the jury is out on the merits of CRE, in part because supporters and critics are telling very different stories constructed from the same events. To the extent that the differences arise from ideological positions on the need for CRE programs, there may be no way of coming to a mutually agreed-upon conclusion.

**Studies Currently Missing From the CRE Controversy: A Focus on Child Outcomes**

Before concluding that the controversies concerning CRE are unresolvable, we believe it is necessary to examine a number of important studies that have not been considered in the debates about CRE effectiveness. The disagreements about the effectiveness of CRE we have described center on whether couples-group participants fare better in relationship quality, and, occasionally, in parenting skills, than nonparticipants. Missing from the discussion so far is research concerning three important questions: What happens to couples and their children without intervention? Does a couples approach add value to interventions for mothers or fathers? Do children benefit when couples, parents, or parenting figures participate together in a CRE intervention?

**What Happens to Couples Over Time Without Intervention?**

The implications of a well-accepted finding have often been ignored in recent accounts of CRE interventions. In more than 50 studies in a number of industrialized countries, marital or couple relationship satisfaction declines over time, even more quickly after couples become parents (Twenge, Campbell, & Foster, 2003). This fact has been cited repeatedly as a justification for creating and disseminating CRE programs. What has not been brought into the discussion is that fact that correlational data suggest that this decline increases risks for children’s cognitive, social, and emotional development (e.g., Cummings & Davies, 2010), and therefore that interventions for couples could have a salutary effect on their children. Of course, satisfaction does not erode for all couples. Recent research suggests that couples initially highly satisfied decline least and those in distress decline most (Lorber, Erlanger, Heyman & O’Leary, 2014). Because of the effort and cost expended in the mounting of CRE programs, it might be useful to think about identifying couples in distress and providing interventions for those most in need.

**Including Both Parents in Parenting Interventions**

**Adding fathers to interventions for mothers.** The field of CRE can learn from an emerging theme in studies of parenting interventions, in which mothers are the usual “designated parent.” Behavioral therapies for mothers of aggressive children have been remarkably successful (Miller & Prinz, 1990), at least in the short run. Three decades ago, confronted by substantial numbers of families in which children reverted to baseline or did not change at all, several parenting therapy programs in the United States and Australia identified a potential explanation of the failures; in these families, unresolved marital conflict was high. After recruiting fathers and adding a new focus on coparenting and marital issues, several studies found that a combined marital and parenting emphasis was more successful in reducing sons’ problem behaviors than a parenting skills approach with mothers alone (Brody & Forehand, 1985; Dadds, Sanders, Behrens, & James, 1987; Webster-Stratton, 1985). Meta-analyses of several different types of parent-training interventions by Lundahl and colleagues (Lundahl, Tollefson, Risser, & Lovejoy, 2008) and Bakermans-Kranenberg and colleagues (Bakermans-Kranenburg, van Ijzenoom, & Juffer, 2003) came to the same conclusion. In future studies, it would be useful to examine whether the increased involvement of fathers with their children is responsible for the positive intervention effects.

Two RCT studies illustrate the contribution made by fathers’ participation in parenting programs. A Canadian study (Besnard, Capuano, Verlaan, Pulin, & Vittor, 2009) that randomly assigned participants to couples participation, solo mother participation, or a control group found that the quality of mothers’ parenting improved when both parents were involved. The second, in the United States (Rienks, Wadsworth, Markman, Einhorn, & Etter, 2011), is the most relevant to our concerns with CRE because the intervention involved group meetings with a CRE approach. The Fatherhood Relationship and Marriage Education program, an adaptation of PREP called “Within our Reach” and designed to meet the needs of low-income couples (Stanley et al., 2006), also included modules on parenting and coping with economic stress. One hundred thirty-seven primarily low-income couples, diverse in ethnic background, were recruited and randomly assigned to (a) groups with both parents, (b) groups attended by only one of the parents, or (c) a no-treatment control condition (average age: fathers = 36, mothers = 31, focal child = 9). Workshops were led by trained leader pairs (usually male–female) selected for their professionalism and sensitivity regarding cultural and economic issues.

Of interest here is that father involvement remained stable in the control group, increased in couples-group participants, and declined when only mothers attended the groups and attempted to convey the information to their partners at home. The effect size of the father-involvement difference between couples and mothers was $d = .34$. The authors speculate that the mothers-only condition may have placed the mothers in a gatekeeping role and led to defensiveness in the fathers. The failure of the mothers-only condition to promote father involvement is concerning, the authors point out, because it is similar to most parenting interventions attended only by mothers.

**Adding mothers to interventions for fathers.** Here we turn the story around to examine what happens when mothers participate along with their partners in an intervention previously targeted toward men. Interventions to enhance father involvement began in the 1980s, but remarkably few received systematic evaluation. A diligent search by Avellar et al. (2011) found 150 studies of “responsible fatherhood programs” since 1990—90 that included low-income fathers, but only 15 that used well-accepted
research design and measurement procedures. A meta-analysis (Holmes, Galovan, Yoshida, & Hawkins, 2010) of 16 father-involvement interventions for resident, low-income, married, or cohabiting fathers showed an overall effect size of $d = .26$, suggesting that father-involvement interventions, like CRE interventions, produce small but statistically significant effects.

Regardless of whether relationship issues are introduced into father-involvement programs, almost all that we are aware of work with men, in group or individual sessions, with male group leaders or individual counselors. The Supporting Father Involvement (SFI; P. A. Cowan, Cowan, Pruett, Pruett, & Wong, 2009) program is the only study in addition to Rienks et al. (2011) that used an RCT research design to evaluate a couples approach to father involvement, but SFI contrasted a couples group and a fathers-only group with the same leaders and curricula. The SFI approach was 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, Pikkauskas, McLanahan, & Brooks-Gunn, 2011), a finding that holds for married, cohabiting, separated, and divorced coparents (Pruett & Johnston, 2004).

During Phase I of the SFI study, parents who expressed interest in the program were invited to take part in one of the following on a randomly assigned basis, all with the same staff: a one-time informational group meeting (3 hr; $n = 98$), a group for fathers that met for 16 weeks (32 hr; $n = 96$), or a group for couples that met for 16 weeks (32 hr; $n = 95$). All three variations of the SFI program were delivered by clinically trained male–female pairs of group leaders, and both fathers and mothers completed all assessments. Every family was also offered the support of a case manager or family worker to help with referrals to other services, as needed, during their time in the project. The majority of the families in Phase I of the SFI study were Mexican American, with most of the remaining families White. Their youngest children ranged in age from birth to 7 years, with a median age of 2.5 years. Two thirds of the families had household incomes below twice the federal poverty line.

Fathers and mothers who participated in the one-time meeting showed no positive and some negative changes; for example, at the 18-month follow-up assessment, father involvement did not change, and satisfaction as a couple declined significantly. Participants in both the 16-week fathers and couples groups showed significant increases in fathers’ involvement in the care of the children. In addition, couples-group participants reported significant reductions in parenting stress, and no decline in satisfaction as a couple, unlike fathers-group and control-group participants. Effect sizes comparing the couples-group participants with those in the single-meeting control group include increases in couple relationship satisfaction ($d = .36$) and father involvement ($d = .34$), and reductions in parenting stress ($d = .31$). A unique feature of this study is that it also showed benefits for the children of participants—a finding we discuss in the next section.

It seems clear that there is a value-added contribution of including both parents when interventions hope to affect children. Compared with parenting interventions for mothers, programs that intervene with both parents show greater positive effects on family relationships and children’s behavior. According to the only existing study to contrast work with couples and fathers alone (SFI), couples groups have a more positive impact on the relationship between the parents than fathers groups, although both enhance men’s participation in children’s lives and appear to prevent a rise in young children’s problem behaviors. These two findings speak to the desirability of testing CRE interventions not only against no-treatment controls but also against father-involvement interventions and programs targeted to one parent at a time, the situation for most parenting classes.

Do Children Benefit From Their Parents Participation in CRE Interventions?

Almost all of the current dialogue about CRE effects has focused on the couples themselves, despite the fact that the primary justification for funding these interventions rests on the argument that benefits for parents would be also benefit their children. Based on direct searches of the literature, research reviews, and meta-analyses, we were surprised to find that of the more than 150 adequately designed CRE studies, we found only nine studies with RCT designs and measures of how the participants’ children fared. In our view, the question of whether children whose parents participate in CRE interventions are better off than those who do not is, or should be, central to policymakers and service providers considering whether to allocate resources to couples-group interventions.

Our expectation that CRE interventions have positive effects on children comes in part from family theory, and partly from empirical studies that document correlations between couple functioning or positive father involvement and children’s cognitive, social, and emotional adaptation. Family systems models (Walsh, 2012) all suggest that positive changes in individual family members, or in the relationships among them, set processes in motion that result in improved adaptation for both adults and children. A large number of studies document correlations between high levels of unresolved conflict between the parents and their children’s academic problems (see meta-analysis of multiple ages by Kitzmann, Gayford, Holt, & Kenny, 2003) and emotional distress or behavior problems in infants and toddlers (McHale & Fivaz-Depeursinge, 1999), preschoolers and school-age children (Cummins & Davies, 2010; Grych & Fincham, 2001), and teenagers (Emery, 1988). The parents’ conflict need not be loud and overt; cold, distanced withdrawal in either or both parents also places children at higher risk for problematic behavior. While much of this research began in the United States, studies of U.K. families provide similar findings (Grych, Harold, & Miles, 2003; Shelton & Harold, 2008).

In the past decade, McHale and Lindahl (2011), among others (Belsky & Jaffee, 2006), made a distinction between two aspects of couple relationships: the interaction between partners around their intimate relationship, and the ways partners work as a team, or fail to do so, as they interact with the child. More collaborative, less competitive coparenting is associated with positive outcomes for children. Taken together, these findings would lead us to expect that interventions to foster more effective couple and coparenting relationships would have important consequences for their offspring, but before we accept the assumption of causal linkage, this expectation must be tested empirically in the context of RCT intervention studies to show that the impact of the intervention on couple, coparenting, or parent–child relationships is actually associated with beneficial outcomes for the child.
Of the nine CRE trials that included assessments of child outcomes, three attempted to enlist partners making the transition to parenthood for the first time, and the rest began with parents of children in the preschool and early elementary school years. Meta-analyses have been done with even fewer than nine studies, but we have chosen not to combine this highly heterogeneous set of studies into one or two overall summary statistics that describe the average impact of parents’ participation on the child. It is true that six of the nine studies contain measures of internalizing or externalizing behavior, but Table 3 shows variability across studies on a number of different dimensions. Five of the intervention trials included primarily middle- and high-income couples, whereas four focused primarily on low-income couples. Intervention intensity (number of hours in the different CRE interventions) varies from 3 to 48, and the age of the child at follow-up from 3 months to 15 years, with no clear indication that measures of these constructs at such widely different ages have similar meaning. Analyses of intervention effects used different methods (ANOVA, regressions), making it difficult to combine in a single effect size. Across studies, data came from parent ratings, teacher ratings, or experimenter observations. Tests of moderator effects of income level, intervention intensity, child age, or source of data with two or three studies in each category would not make sense, especially because some of the moderator variables are confounded. At this point, we have chosen to describe the strengths and weaknesses of each study, with the goal of stimulating discussion and further research on whether and how parents’ CRE participation affects their children.

**Becoming a family (C. P. Cowan & Cowan, 2000).** In this study, with random assignment to no-treatment and intervention conditions, clinically trained faculty and graduate students in clinical psychology formed male–female coleader teams who met weekly with small groups of working-class to middle-class couples from midpregnancy to 3 months after the birth of their first child (48 hr over a period of 6 months). Couples were assessed at baseline, 9 months, 1.5 years, 3.5 years, and 5.5 years postpartum once the children had made their transition to kindergarten. The intervention with the parents had a long-term significant effect over the course of the study on couples’ marital satisfaction (Schulz, Cowan, & Cowan, 2006), preventing the decline over time found in most nonintervention studies (Twenge et al., 2003). Although both self-reports and observations of marital quality were correlated with parent reports, laboratory observations, and kindergarten teacher ratings of the children’s school adaptation, there were no significant differences between intervention- and control-group children 3 and 5 years after the intervention ended.

**Bringing Baby Home (BBH; Shapiro & Gottman, 2005).** This program was originally presented as a weekend workshop for expectant couples (16 hr), with psychoeducational classes led by clinical psychologists that included lectures, demonstrations, videotapes, role plays, and communication exercises. Topics included couple communication, promoting father involvement, sensitive parenting, coparenting, and infant development. A subsequent RCT (Shapiro, Nahm, Gottman, & Content, 2011) was conducted with 142 expectant middle-income couples assigned to a no-treatment control condition (n = 46), a weekend workshop (n = 45), or a workshop plus a 12-week “support group” (n = 51) that met 12 times over a 6-month period after the birth of a child (24 hr). The workshops and support groups were led by childbirth preparation teachers, some of who had master’s or nursing degrees. The Shapiro et al. article reported on assessments occurring at baseline and again when the infants were 3 months old, a time when the support groups had just begun. Observations were made of mother, father, and infant in the Lausanne Trilogue Play paradigm (Fivaz-Depeursinge & Corboz-Warnery, 1999), in which father and mother take turns playing with the infant while the other parent remains passive, and then both participate together in parent–infant play. The BBH program was successful in promoting less competition between parents during family play. There were no significant differences between control and intervention families (workshop only and workshop plus support group combined) on infant facial affect, vocalization, or gaze, but this result could be attributed to the fact that the support group was in its early phases when the babies’ outcomes were assessed. An in-progress report of a follow-up when the children were 30 months (Gottman, Gottman, & Shapiro, 2010) asserted that there were dramatically increased effects from the 12-session Cowan-type couples support group . . . added to the workshop . . . [including] less negative ratings of child behavior on the Child Adaptive Behavior Inventory [P. A. Cowan, Cowan, & Henning, 1995] and better language development in the toddlers [no effect sizes reported]. In the experimental groups, children ask more questions, engage in more pretend play, and have better and more complex spoken vocabulary. (p. 174)

We are not aware of published detailed results of this follow-up.

**Family Foundations (Feinberg, Kan, & Goslin, 2009).** This program has a unique emphasis based on data from earlier correlational studies showing that the coparenting relationship is not identical to the couple relationship, and that it adds significant predictive power to path models that chart the connections between couple relationships, parenting quality, and children’s outcomes (Feinberg, 2003). One hundred sixty-nine middle-income couples expecting a first child, most of whom were married and White, were randomly assigned to intervention and no-treatment conditions. The intervention couples attended a total of eight classes (16 hr), pre- and postbirth, with a special emphasis on managing disagreements about parenting through the development of communication skills, problem-solving, and conflict management techniques. The program also included modules on partners sharing expectations for each other, and on parenting an infant. Based on parent-report data at the 3.5-year postpartum follow-up (Feinberg et al., 2009), parents of boys showed a positive effect of group participation on relationship satisfaction (d = .43), but there were no significant effects for parents of girls. Other results were more consistently positive: The intervention participants described significantly more positive coparenting (d = .18) and more effective parenting than the parents in the control group (less lax, less likely to inflict physical punishment; d from .30 to .36). According to the parents, children of CRE participants showed fewer externalizing, internalizing, and attention problems, espe-

---

5 Calculated from the published data.
<table>
<thead>
<tr>
<th>Study</th>
<th>Income</th>
<th>N</th>
<th>Hrs. of CRE</th>
<th>Child age at pre</th>
<th>Child age at follow-up</th>
<th>Impact on relationship quality</th>
<th>Impact on parenting</th>
<th>Child outcomes</th>
<th>Effect sizes</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Becoming a family</strong> (Cowan &amp; Cowan, 2000)</td>
<td>Mid/hi</td>
<td>96</td>
<td>48</td>
<td>In utero</td>
<td>5 yrs.</td>
<td>Parent report</td>
<td>Observer</td>
<td>Teacher report</td>
<td>n.s.</td>
</tr>
<tr>
<td><strong>Bringing baby home</strong> (Gottman et al., 2010)</td>
<td>Mid/hi</td>
<td>142</td>
<td>34</td>
<td>Newborn</td>
<td>3 mos.</td>
<td>Parent report &amp; observer</td>
<td>Parent report &amp; observer</td>
<td></td>
<td>n.s.</td>
</tr>
<tr>
<td><strong>Family foundations</strong> (Feinberg et al., 2009)</td>
<td>Mid/hi</td>
<td>169</td>
<td>16</td>
<td>Pre and post birth</td>
<td>3.5 yrs.</td>
<td>d = .43</td>
<td>d = .30–.36</td>
<td>Ext. 12%</td>
<td>Mdn d = .74</td>
</tr>
<tr>
<td><strong>Emotional security</strong> (Cummings et al., 2008)</td>
<td>Mid/hi</td>
<td>90</td>
<td>3</td>
<td>4–8 yrs.</td>
<td>6–10 yrs.</td>
<td>Sig.</td>
<td>Sig.</td>
<td>Child adjustment</td>
<td>Sig.</td>
</tr>
<tr>
<td><strong>Schoolchildren and their families</strong> (Cowan et al., 2005)</td>
<td>Mid/hi</td>
<td>100</td>
<td>32</td>
<td>5 yrs.</td>
<td>7 yrs.</td>
<td>Sig.</td>
<td>Sig.</td>
<td>Teacher and test</td>
<td>Path model</td>
</tr>
<tr>
<td><strong>Building strong families</strong> (Wood et al., 2010)</td>
<td>Lo</td>
<td>5,102</td>
<td>32</td>
<td>0–15 yrs.</td>
<td>2–17 yrs.</td>
<td>n.s.</td>
<td>n.s.</td>
<td>Socioemotional devel.</td>
<td>d = .08</td>
</tr>
<tr>
<td><strong>Supporting Healthy Marriage</strong> (Hueh et al., 2012)</td>
<td>Lo</td>
<td>6,298</td>
<td>32</td>
<td>0–2 yrs.</td>
<td>2.5–4.5 yrs.</td>
<td>Parent report observer</td>
<td>Observe</td>
<td>Self-reg.</td>
<td>d = .07</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Income</th>
<th>N</th>
<th>Hrs. of CRE</th>
<th>Child age at pre</th>
<th>Child age at follow-up</th>
<th>Impact on relationship quality</th>
<th>Impact on parenting</th>
<th>Child outcomes</th>
<th>Effect sizes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mid/hi</td>
<td>96</td>
<td>48</td>
<td>In utero</td>
<td>5 yrs.</td>
<td>Parent report</td>
<td>Observer</td>
<td>Teacher report</td>
<td>n.s.</td>
</tr>
<tr>
<td>Mid/hi</td>
<td>142</td>
<td>34</td>
<td>Newborn</td>
<td>3 mos.</td>
<td>Parent report &amp; observer</td>
<td>Parent report &amp; observer</td>
<td></td>
<td>n.s.</td>
</tr>
<tr>
<td>Mid/hi</td>
<td>169</td>
<td>16</td>
<td>Pre and post birth</td>
<td>3.5 yrs.</td>
<td>d = .43</td>
<td>d = .30–.36</td>
<td>Ext. 12%</td>
<td>Mdn d = .74</td>
</tr>
<tr>
<td>Mid/hi</td>
<td>90</td>
<td>3</td>
<td>4–8 yrs.</td>
<td>6–10 yrs.</td>
<td>Sig.</td>
<td>Sig.</td>
<td>Child adjustment</td>
<td>Sig.</td>
</tr>
<tr>
<td>Mid/hi</td>
<td>100</td>
<td>32</td>
<td>5 yrs.</td>
<td>7 yrs.</td>
<td>Sig.</td>
<td>Sig.</td>
<td>Teacher and test</td>
<td>Path model</td>
</tr>
<tr>
<td>Lo</td>
<td>5,102</td>
<td>32</td>
<td>0–15 yrs.</td>
<td>2–17 yrs.</td>
<td>n.s.</td>
<td>n.s.</td>
<td>Socioemotional devel.</td>
<td>d = .08</td>
</tr>
<tr>
<td>Lo</td>
<td>6,298</td>
<td>32</td>
<td>0–2 yrs.</td>
<td>2.5–4.5 yrs.</td>
<td>Parent report observer</td>
<td>Observe</td>
<td>Self-reg.</td>
<td>d = .07</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>N</th>
<th>Hrs. of CRE</th>
<th>Child age at pre</th>
<th>Child age at follow-up</th>
<th>Impact on relationship quality</th>
<th>Impact on parenting</th>
<th>Child outcomes</th>
<th>Effect sizes</th>
</tr>
</thead>
<tbody>
<tr>
<td>96</td>
<td>48</td>
<td>In utero</td>
<td>5 yrs.</td>
<td>Parent report</td>
<td>Observer</td>
<td>Teacher report</td>
<td>n.s.</td>
</tr>
<tr>
<td>142</td>
<td>34</td>
<td>Newborn</td>
<td>3 mos.</td>
<td>Parent report &amp; observer</td>
<td>Parent report &amp; observer</td>
<td></td>
<td>n.s.</td>
</tr>
<tr>
<td>169</td>
<td>16</td>
<td>Pre and post birth</td>
<td>3.5 yrs.</td>
<td>d = .43</td>
<td>d = .30–.36</td>
<td>Ext. 12%</td>
<td>Mdn d = .74</td>
</tr>
<tr>
<td>90</td>
<td>3</td>
<td>4–8 yrs.</td>
<td>6–10 yrs.</td>
<td>Sig.</td>
<td>Sig.</td>
<td>Child adjustment</td>
<td>Sig.</td>
</tr>
<tr>
<td>100</td>
<td>32</td>
<td>5 yrs.</td>
<td>7 yrs.</td>
<td>Sig.</td>
<td>Sig.</td>
<td>Teacher and test</td>
<td>Path model</td>
</tr>
<tr>
<td>5,102</td>
<td>32</td>
<td>0–15 yrs.</td>
<td>2–17 yrs.</td>
<td>n.s.</td>
<td>n.s.</td>
<td>Socioemotional devel.</td>
<td>d = .08</td>
</tr>
<tr>
<td>6,298</td>
<td>32</td>
<td>0–2 yrs.</td>
<td>2.5–4.5 yrs.</td>
<td>Parent report observer</td>
<td>Observe</td>
<td>Self-reg.</td>
<td>d = .07</td>
</tr>
</tbody>
</table>
cially for boys, and higher levels of social competence for both sexes (median of six measures, $d = .74$).

**Emotional Security Theory Program for Community Families** (Cummings, Faircloth, Mitchell, Cummings, & Schermerhorn, 2008; Cummings & Schatz, 2012). Based on emotional security theory (Cummings & Davies, 2010) as a construct linking marital conflict to family relationships and children’s outcomes, Cummings and colleagues developed and tested a four-session (45 min per session for a total of 3 hr) group intervention for nonclinical couples who were parents of 4- to 8-year-old children. Couples were randomly assigned to one of three conditions—(a) a parent-only group ($n = 24$), (b) a parent–child group ($n = 33$), and (c) a self-study condition ($n = 33$)—with pre- and posttests and 6-month and 1-year follow-ups. Intervention staff members were given a brief preparation but were not clinically trained. The four sessions focused on improving parents’ ways of expressing disagreement, with special attention to helping couples distinguish between productive and nonproductive conflict, the effects of marital conflict on children, and the importance of maintaining the quality of emotional bonds among all family members during stressful family situations. A 2-year follow-up (Faircloth, Schermerhorn, Mitchell, Cummings, & Cummings, 2011) with substantial attrition (now 43% of the original couples) found positive, large effects of the intervention, compared with a randomized control, on measures of knowledge about the consequences of marital conflict for children ($d = 1.35$) and constructive conflict behaviors ($d = 1.06$), but not on couple problem solving. The authors report that there was no difference between couples who participated in the follow-up and couples who did not on any of the pretest measures.

The data for main effects of the intervention on child outcomes were not reported. Multilevel growth curve models found that increases in the constructiveness of marital conflict behavior during the marital conflict resolution task were linked with greater marital satisfaction, better parenting practices, and improved child adjustment (Child Behavior Check List, Achenbach & Edelbrock, 1983).

**Schoolchildren and Their Families Project** (P. A. Cowan, Cowan, Ablow, Johnson, & Measelle, 2005). One hundred middle-income couples were recruited to participate in a randomized clinical trial of a program for couples with a first child about to make the transition to elementary school. One third of the parents were assigned to a low-dose control condition, in which they had the option of consulting once each year with the clinically trained male–female staff team (psychologists and social workers) who interviewed them when they entered the study. The remaining two thirds participated in couples groups of four of five couples meeting for 16 weeks (32 hr) with the same staff teams. The group meetings began with an open-ended check-in followed by an agenda for each week that covered one of the topics of a five-domain risk-protective model of family functioning (individual well-being; couple communication, including coparenting; parenting; what to carry over and what to avoid from parents’ families of origin; and life stresses and social supports outside the nuclear family). Couples were randomly assigned to one of two variations of the couples-group condition, with the same pairs of group leaders conducting each variation. In one set of groups, in the open-ended part of each meeting, leaders helped participants focus on parenting and their relationship with the child, and in the other groups, on their relationship as coparents. Two years after the couples groups ended when the children were in first grade, observations of parent–child interaction revealed that participants in the intervention that emphasized parenting were significantly warmer and more structured with their child than parents in the control condition; their children showed fewer internalizing behaviors in their first-grade classroom and reported a greater sense of well-being in a puppet interview (C. P. Cowan, Cowan, & Heming, 2005b). Parents in the groups that emphasized the couple and coparenting relationship used significantly more effective parenting strategies than the control parents 2 years later, and in contrast with controls, they also showed no increase in their level of couple conflict. Reports from first-grade teachers revealed that their children were less aggressive and showed significantly higher levels of academic achievement than children of the control participants when tested individually by a member of the research team. Over and above the quality of marital and parenting relationships, interventions with the parents accounted for 12% of the variance in children’s academic achievement scores in kindergarten, 21% of the variance in children’s perceptions of their kindergarten adjustment, 31% of the variance in reductions in externalizing behavior between kindergarten and first grade, and 12% of the reductions in their internalizing behaviors over the same period of time. In an exploration of mediating effects, the study found that changes in marital and parent–child interaction in the year from prekindergarten to kindergarten were linked with the child outcomes assessed at the end of first grade. Finally, in an unusually long-term follow-up assessment 10 years later, as the children made the transition to ninth grade in high school, intervention effects on mothers, fathers, and children were still apparent (P. A. Cowan, Cowan, & Barry, 2011).

**BSF.** In their final report of the 3-year follow-up of BSF for low-income unmarried couples at eight sites, Wood et al. (2014) presented new data on the effect of the intervention on three measures of child well-being. The first two, family stability and economic well-being, are measures of potential stressors or protective factors that affect children. The third is socioemotional development, for which they found a very small but statistically significant impact of the intervention on parents’ reports of children’s behavior problems ($d = .08$). It is puzzling that this effect, however tiny, occurred in the absence of effects on the parents. One speculation provided by the authors of the BSF report is that the behavior problem finding was significant only in the four sites providing home visiting services that included instructions in parenting skills.

**SHM** (Hsueh et al., 2012; Lundquist et al., 2014). We described this RCT study of CRE at eight sites with more than 6,000 low-income married couples earlier, along with outcomes for self-report and observational measures of couple relationship quality. Here we focus on what happened to the children. All four of the couples-group programs in SHM focused on couple communication and also included modules concerning individual distress and coping with external stressors. Although only one dealt directly with parenting issues, many of the couple communication issues in all of the programs involved disagreements about coparenting. The groups were led by male–female teams with a mix of backgrounds, including parent educators, doctoral graduate students, nurses, social workers, and marriage and family therapists.
Staff qualifications at some of the eight sites were not described in detail.

The results for child outcomes were reported in two sections—primary and secondary measures. Primary measures were four composite measures (including observations and parent reports) that the investigators predicted would show intervention effects (i.e., children’s self-regulation, internalizing, and externalizing behavior). Analysis of the total sample at the 30-month follow-up, when children ranged from 4 to 17 years, revealed two small effects on children’s behavior (self-regulation, \( d = .03 \), and internalizing, \( d = .04 \)), which were not statistically significant when corrections for performing multiple tests were applied. However, when the whole sample was divided by children’s age categories, significant positive intervention effects were found for the youngest sample (2- to 4-year-olds) on three of the four measures (self-regulation, \( d = .07 \); internalizing, \( d = .10 \); externalizing, \( d = .08 \)), even after correcting for multiple tests. There were no intervention effects on these measures in the 5- to 8.5-year-olds or the 8.5- to 17-year-olds.

Secondary measures were conceptualized as more exploratory. The investigators reported small but statistically significant effects for 2- to 17-year-old children whose parents participated in the couples groups: The children were more socially competent (\( d = .07 \)); were less distressed (\( d = .06 \)) and dysregulated (\( d = .05 \)); became less involved when their parents were fighting (\( d = .06 \)); and were less likely to attribute blame to themselves when their parents fought (\( d = .13 \)). Given that five of nine secondary measures showed intervention effects, the results could not be attributed to the number of statistical tests.

**SFI.** In addition to effects on couple relationship satisfaction and father involvement, the SFI study of low-income couples (P. A. Cowan et al., 2009) tested for intervention effects on parent-reported child behavior on four factor-constructed dimensions: aggression, hyperactivity, shy/withdrawn behavior, and anxious/depressed behavior. In the single-session control condition, parents reported a significant increase in problem behaviors on all four child dimensions; there were no significant increases for children of couples in which fathers participated in groups, or for couples who participated in groups. Two of the effect sizes contrasting control-group and couples-group participants were statistically significant: child hyperactivity (\( d = .22 \)) and child shy/withdrawn, (\( d = 1.88 \)).

A partial replication study of the SFI couples-group intervention was conducted in Phase II (P. A. Cowan, Cowan, Pruitt, Pruitt, & Gillette, 2014) with a new set of 239 low-income Mexican American, White, and African American couples. Because the low-dose controls in Phase I showed negative changes over time and the fathers groups were not as effective, Phase II analyses were conducted as a single-sample pre–post assessment of participants in couples groups. In the absence of a randomized control, the results of Phase I were used as a benchmark comparison to evaluate the Phase II findings. Statistically significant effect sizes in favor of the intervention participants were found for father involvement (\( d = .38 \)), couple relationship satisfaction (\( d = .37 \)), and child aggression, (\( d = .34 \)).

**Summary.** In sum, eight of nine intervention trials that included measures of child effects showed statistically significant effects of CRE interventions on children of the participants on at least some of the child outcome measures—only the BAF intervention did not; lack of statistical power (\( n = 66 \)) and parents’ fairly high level of functioning at entry to the study could account for the failure to find child effects associated with parents’ participation in the intervention 5 years earlier. In addition, there was little in the BAF materials related to parenting a newborn that were relevant to the parenting issues of toddlers and preschoolers. It is also the case that, like every other set of intervention studies, CRE programs had different effects on different child outcomes, even when, as in the case of SFI, the two intervention trials were almost identical.

The data we have reported in the text and listed in Table 3 do not simply represent a box score. The results include an array of small positive effects, and several medium or large effects, across many measures from different sources in both higher and lower income families. Although these nine CRE trials are not sufficient to draw a general conclusion about the impact of couples-group participation on children, they provide some support for what had been an untested assumption—that parents’ participation in these groups can have a positive effect on their children. They also provide suggestions for studies of program variables that might be associated with more robust child outcome effects.

**Future Directions for CRE Research**

Our review of research on CRE interventions reveals that evaluation studies of programs to strengthen couple relationships in middle-income and low-income families provide evidence for both supporters and critics. Almost every statement from either perspective is subject to some form of rebuttal, and sometimes a rebuttal to the rebuttal. Descriptive reviews generally include studies favoring CRE, but often do not do justice to negative findings. Overall effect sizes from meta-analyses are small; some show that effects disappear over time. Countering these results are moderator analyses that show some medium and large effects of CRE. Like the large BSF trial with unmarried couples that was disappointing to CRE supporters, the large SHM trial with married couples has been dismissed as having small effects by critics, but, in fact, the SHM results show consistent 12- and 30-month follow-up findings of effects on both parent reports and observer measures. Some single studies have been cited as casting doubt on CRE but, like studies supporting CRE, have flaws that reduce the force of their conclusions.

Our review also included data from studies that have not been considered much in discussions of CRE, especially to answer the question of whether parents’ participation in CRE benefits their children. Without intervention, average marital satisfaction declines, and this decline affects children negatively. Adding fathers to interventions designed for mothers, and adding mothers to interventions designed for fathers (i.e., taking a couples approach to intervention), increases positive effects on both couples and children. Finally, we described nine RCT intervention trials of CRE that examined child outcomes, with eight reporting statistically significant effects.

In our view, there are too many negative findings and cautions associated with CRE findings to conclude the CRE is an unqualified success and that existing programs should be offered more widely to large sections of the population. There are also too many positive findings and promising indications of factors associated with stronger results to recommend that funding for CRE programs
be discontinued and that we should look to other ways of helping families. The meta-analyses we have surveyed suggest that some CRE programs are more effective than others—moderate rather than low-dose interventions that use professionally trained group leaders and include observational measures. We need to disentangle the separate and combined contributions of these dimensions, and other dimensions yet to be tested, in order to create a more differentiated picture of what CRE interventions can realistically be expected to accomplish for families, and what their limitations may be.

Study Design

We focused on randomized assignment studies in the earlier part of this article, but there are many more reports in the literature of CRE programs that use single-sample pre–post designs. Although any attempt to provide systematic evaluations of intervention programs is welcome, and contributions to refining interventions can be made without using RCTs, a critical problem occurs when single-sample pre–post studies show no change over time. The single-sample design fails to identify circumstances in which the absence of change represents a positive finding, for example, when marital satisfaction remains stable for intervention participants, but would likely have shown a decline over time if a comparison or control group had been included in the study.

We noted that CRE intervention studies have relied on relatively short-term follow-ups. Especially because early preschool interventions found “sleeper effects” long after the interventions concluded (Schweinhart et al., 2004), it is important for researchers to return to participants at least 18 months to 2 years postintervention, or later, to determine whether couples have integrated what has been learned in these interventions into their daily lives.

Program Design and Measurement

Up to the present, there has been almost no effort to compare the effectiveness of different kinds of programs. Only very occasionally have investigators gone beyond intervention versus control comparisons to evaluate differences between programs with different characteristics. Benefits to participants may vary with (a) the structure of the program, (b) the content of the curriculum, and (c) the process involved in the way the intervention is delivered. So far, analyses of intervention dosage have been conducted by comparing studies with fewer and greater numbers of sessions, or by correlating attendance rates and outcomes. Analyses of attendance rates do not provide good measures of dosage because self-selection is usually a factor in whether couples or fathers continue or drop out, and so those who attend more sessions (receive higher treatment dosage) may be those who are more motivated or find the program useful. We need information about the effects of an intervention when dosage is varied systematically, using random assignment to shorter and longer versions of the program.

Programs with different types of curriculum content and different approaches to teaching and learning are hidden within the summary statistics yielded by meta-analyses. At present, the most highly researched CRE intervention is the PREP program (Markman et al., 2010). PREP focuses on teaching moderate to large groups of couples effective communication skills. Other programs, which we could call “communication plus,” include additional aspects of family life including three-generational issues (Gordon, DeMaria, Haggerty, & Hayes), parenting (C. P. Cowan & Cowan, 2000; Gottman et al., 2010), and outside-the-family life stress (P. A. Cowan et al., 2009; Rienks et al., 2011).

Often confounded with curriculum content is the program’s approach to intervention. PREP programs take a psychoeducational approach, with material presented by instructors and followed by practice sessions for each couple with coaching from the instructor or assistants. Other programs, such as SFI, have a defined curriculum but place more emphasis on participants’ particular issues, and on group process and interaction among the participants. As of now, we know that some programs with very different approaches (e.g., PREP, Family Foundations, SFI) have helped to improve or at least maintain couple relationship quality, but we do not know how these programs compare with each other when tested in the same participant populations. Only by adding and subtracting curriculum modules and measuring outcomes is it possible to determine whether specific aspects of the intervention curriculum produce positive effects (e.g., couple communication training, father-involvement discussions, parenting skills).

Differences in curriculum and intervention approach are often confounded with two additional variables. First, programs that train specific skills can accommodate a larger number of attendees, whereas programs that emphasize personal issues and group discussion must be smaller (ideally about four to six couples). We are aware of no CRE studies that attempt to link group size to outcomes. Second, although there is one meta-analysis showing that professionally trained group leaders are more effective than para-professionals, much more investigation of this key topic is needed. One important issue is whether the effectiveness of higher levels of training varies with program structure. From observation, we believe that the more the program relies on teaching skills with lessons derived from a highly specific manual and communication problems chosen by the instructor(s), the less training is required for those who deliver the program. The more the program relies on group leaders who raise topics, present exercises, and foster discussion of spontaneous personal examples between partners and among group members, the more training and clinical skills the leader must have in order to deal with the complex personal issues that couples bring and to contain the level of individual and couple distress that often emerges. We are aware that hiring leaders with less extensive training is less costly to those responsible for mounting the program, which means that answers to questions about training and staff experience have important practical and policy implications.

Almost all evaluations of CRE interventions focus on couple relationship satisfaction and communication skills. A few assess other domains, such as parents’ adjustment, fathers’ involvement, and parenting style. And, as we have described, a very few examine effects on the children. Elsewhere (Cowan et al., 2005), we have shown that couple relationship quality, parents’ individual adjustment, and parent–child relationship quality are three of five major dimensions of risks and buffers affecting family adaptation. The other two are factors associated with relationship patterns in one’s family of origin, and the balance between life stressors and social supports outside the family. Very few studies target these factors in the CRE curriculum or assess these dimensions as potential mediating or moderating variables.
The source of information about the family also varies from study to study. A wide variety of self-report questionnaire measures are included in evaluation studies. Some evidence suggests that observational measures might provide clearer evidence of intervention effects (e.g., Fawcett et al., 2010). The question is not only whether observational measures are better than self-reports in looking for program effects but also whether observational measures show changes earlier than self-reports. It may be, for example, that it is more difficult to change overall perceptions and attitudes about a relationship than it is to make small changes in behavior when the partners disagree. It would help immeasurably if the field would commit to developing some widely accepted self-report and observational measures to be used across studies so that direct comparisons of other study variations can be made more clearly.

**Target Populations**

**Income, ethnicity/race, and culture.** Initially, almost all evaluations of CRE were conducted on middle-class, primarily White samples; more recently, low-income participants have been included from White, African American, and Latino samples, but until the BSF, SHM, and SFI evaluations, systematic evaluations of couple-focused interventions on non-White participants were in short supply. Even less evident in CRE research is information about other ethnic groups such as various subgroups of Asian, Middle Eastern, etc. Whether intervention outcomes differ across income levels and ethnic groups has yet to be determined.

So far, the bulk of the studies of CRE have been located in the United States, although we have noted a few located in other countries. We know that across countries and cultures, there are wide differences in conceptions of marital roles and gendered parenting roles, but we do not know whether some programs may be more effective than others in reaching different cultural groups, or anything about the extent to which CRE programs must be tailored to fit with cultural expectations and norms for couples.

**Family life intervention points.** Another important source of variation across studies has been the point in the family life cycle at which the intervention was conducted. For CRE studies, the greatest programmatic emphasis has been on premarital preventive interventions, with a second emphasis on couples having a baby, usually a first baby. Confounded with family life stage is the age of the parents and the child. While very few CRE programs have been offered to teens (see Florsheim, 2014, for an important exception), the vast majority occur early in the family cycle with young parents. We need to know much more about whether there are critical periods or intervention points at which these programs produce their strongest effects—for parents and for their children.

The prevention focus of CRE has been shifting in recent efforts to address populations at risk by virtue of their poverty status. Other important extensions of CRE could profitably be explored, including coparenting interventions for divorcing couples or for families involved in the child welfare or juvenile justice systems, and a few of these programs are emerging. We assume that as we climb the ladder of risk, leaders of CRE programs would be more highly trained, and programs would be conducted in combination with other services addressed to meeting the needs of these multi-risk families.

We should note that despite the intent of programs to target homogeneous populations, the participants in any couples or father-involvement program are likely to be heterogeneous on a number of dimensions (age, race/ethnicity/culture, income, mental health, quality of coparenting relationship, etc.). Future studies could profitably address the question of whether there is an optimal level of homogeneity or heterogeneity among CRE-group participants to obtain positive effects.

**Dynamic interactions among variables that influence outcomes.** Adding to the complexity of our recommendations for future research is the reasonable assumption that some of these variables may act in combination to facilitate or interfere with intervention outcome. To choose one obvious possibility, longer interventions, with longer follow-ups, more highly trained staff, and increased opportunities for family support may be required to provide effective CRE interventions for parents at higher risk by virtue of poverty, clinically diagnosable mental health problems, or fragile relationship status (e.g., never married, never lived together). It is not possible to include all factors affecting CRE outcomes in one study. Our hope is that this more inclusive map of the variables most likely to affect the selected target population will encourage investigators to select a combination of variables that will provide more information than we have now about what works for whom.

**Testing Theories About How the Interventions Work**

Only a few studies that we are aware of have gone beyond the determination of whether intervention participants fare better than participants in control groups to examine potential pathways of influence. For example, using path analyses and regressions, C. P. Cowan et al. (2005) showed that intervention effects on child outcomes were mediated by intervention-induced changes in observed couple communication and parenting effectiveness. Similarly, reduction in marital conflict was associated with improved parenting in the emotional security-based intervention described by Cummings et al. (2005). As in the rest of our review, the results do not always follow the expected pattern. For example, in a German study, Rogge et al. (2013) found that changes in variables targeted by the intervention curriculum either did not occur or were unrelated to outcomes. Specifically, wives in the German PREP intervention focused on reducing couple conflict did not decline in conflict as much as wives in an empathy-based program. Another of several examples cited by Bradbury and Lavner (2012) is a French study by Bodenmann and colleagues (Bodenmann, Bradbury, & Phet, 2008) in which increases in female negative communication were associated with positive relationship outcomes, counter to the assumption in the PREP program that reducing negative communication is one of the keys to successful CRE programs. Bradbury and Lavner advance an interesting hypothesis: It may be that investigators are not measuring some of the important variables that produce CRE intervention effects. Rather than a focus on prescribing behaviors, they argue, it may be necessary to identify principles that underlie relationship-sustaining communication, and to examine whether changes in how couples adopt these principles underlie positive CRE intervention effects.
Intervention studies also have unique power to contribute to the testing of hypotheses about causal relationships (P. A. Cowan & Cowan, 2002). If we can show that CRE interventions affect specific aspects of family relationships, and that intervention-induced improvement is associated with positive changes in children’s adaptation, we will have obtained strong evidence for our theories of how family factors affect children’s development and well-being.

Policy Implications

Cost-Benefit Analyses: Translating Findings Into Outcomes of Interest to Policymakers

An unresolved issue at the interface of research and policy is the question of potential costs and benefits of CRE interventions. Policymakers want data showing that the benefits ascribed to a proposed program will be greater than the costs, but so far, only the SHM program provided an estimate of costs (between $9,000 and $11,000 per family); we are not aware of attempts to determine the financial benefits of reducing couple conflict or decreasing behavior problems in young children, although beginning attempts are being made in the United Kingdom. The report by Spielhofer et al. (2014) asserts that for every £1 spent for marriage preparation, the returns to society are £11.5 ($18.5 for every $1 spent), and that the returns are similar to those for couples counseling. The main problem in this endeavor for current CRE programs is that, with the exception of facts like marital status, presence of father in the home, or contribution to child support, the outcomes are typically reported as continuous measures (couple relationship satisfaction, positive parenting, child behavior problems) that have yet to be assigned monetary values. What cost–benefit researchers and politicians want to know is whether CRE interventions produce change in the percentage of individuals and families involved in clinical or social categories: diagnosed depression, divorce/separation, and children and adolescents needing therapy or involved in child protective or juvenile justice systems. A further difficulty in monetizing child outcomes is that policymakers want to know about socially important outcomes when children are older, whereas most of these intervention programs, especially those with a preventive focus, are with parents of children from newborn to the preschool and elementary school years.

Service providers with an eye on costs also need more information about program parameters such as those we listed earlier (number of meetings, level of staff training, curriculum content or approach). The question of costs and benefits is especially relevant to the discussion of CRE programs because, at best, the effect sizes of many of these interventions are small. Until we know whether the benefits to be gained by these programs result in substantial savings (e.g., whether reductions in young children’s aggression ultimately pay off in terms of less need for therapy or involvement in the juvenile justice system), it will be difficult to evaluate the meaning of the effect sizes that have been found.

One specific outcome of interest to policymakers has been a reduction in the incidence of divorce. Given the political context of “marriage promotion” and increases in federal funding for CRE in the last decade, the prevention of divorce may be a primary goal of politicians who vote to fund these programs. And yet neither the BSF nor the SHM programs (except at one site), nor the SFI project, found that CRE interventions keep low-income couples from separating and divorcing. While there have been occasional reports that the rate of separation and divorce was reduced after participation in a CRE program (e.g., Stanley et al., in press), most studies (a) have short-term follow-ups in which it would be unlikely to find changes in family stability, (b) fail to find such effects, or (c) do not report relationship stability data. The question to consider is whether the findings represent a criticism of CRE programs or a misunderstanding of program goals. The policy intention of some supporters of CRE is to “promote marriage” and/or “maintain two-parent families,” whereas the intervention curricula are devoted to improving couple relationship quality—for the benefit of the parents and the children. From the former perspective, a divorce or separation of a participant couple would count as a program failure. From the latter perspective, if the couple were able to maintain a collaborative coparenting relationship after making the decision to separate or divorce, the program would be regarded as a success, and from all we know, both parents and children should benefit. It is possible that parental separation or divorce might be protective of the child’s well-being if it lessened the conflict the child was exposed to, or that an improved, collaborative coparenting relationship might benefit children if it lessened the strain in either or both of the parent–child relationships, regardless of whether their parents are married, cohabiting, separated, or divorced (Pruett & Barker, 2009). None of the studies that we are aware of go beyond the fact of parental status to examine the quality of the coparenting relationship after separation or divorce in experimental and control participants.

How Do CRE Interventions Apply to Same-Sex Marriages?

In the funds made available for CRE programs, the moral emphasis on marriage promotion in the United States at a time when same-sex marriage was not legal led to the exclusion of same-sex couples from federally funded CRE intervention programs. Especially in view of the changing legal climate in the United States, but also as a personal value, we believe that this exclusion is not warranted, and that same-sex couples should not only be eligible but also sought out for inclusion in CRE programs. Gay and lesbian couples struggle with challenging issues as partners and as parents, just as male–female couples do, and by any logic, helping to strengthen their relationships should foster more attentive parenting and their own and their children’s well-being.

There is a question about whether current versions of CRE would require alterations to fit specific needs of lesbian and gay parents. Our impression from the literature on lesbian couples (Blake, Casey, Jadva, & Golombok, 2012; Goldberg & Perry-Jenkins, 2007; Gottman et al., 2003) is that role differences and disagreements about issues lead to very similar communication dynamics as those experienced by heterosexual couples. Regardless of sexual orientation, all couples face relationship issues arising from attempts to deal with different family of origin patterns, personal mental health challenges, differences in parenting ideas, work–family balance, and other modern family stresses. The speculations we have raised can only be addressed empirically, with systematic evaluations of intervention trials with both male and female same-sex couples.
Marriage Promotion, CRE, and Alleviating Poverty

One objection to the funding of CRE programs is that the original premise of “promoting marriage” in order to raise families out of poverty is flawed. The fact that children of single mothers are more likely to live in poverty does not mean that marriage (any marriage, regardless of the relationship quality) will raise the economic level of the family. What is missing from the correlation between marital status and poverty is an answer to the question “What happens when single mothers marry?” According to Graefe and Lichter (2007), who analyzed data from 3,872 women who participated in the National Survey of Family Growth, poor single mothers who marry tend to have low marital satisfaction, divorce early, and be even worse off economically than single mothers who do not marry. That is, in this study, the transition from single motherhood to marriage in low-income families did not result in the hoped-for economic benefits.

In the controversies over government funding of CRE programs, a frequently raised issue is that there are other interventions, especially for low-income families, that would be more effective. The most frequent view (e.g., P. Cohen, 2014; Williams, 2014) is that poor families would be better served by programs that elevate their economic circumstances than by relationship-enhancement interventions. There is no question that poverty affects couple relationships and family functioning (Conger, Cui, & Lorenz, 2011). There is also no question that the normative decline in marital satisfaction over time also affects family relationships, with negative outcomes for children. What is in question here is whether alleviating poverty through income supplementation will increase the quality of couple and family relationships. We have found older studies in both the United States and the United Kingdom in which family income supplement programs under various names have had either a negative (Hannan, Tuma, & Groenfeld, 1977; Knox & Redcross, 2000) or neutral (Cain, Wissoker, Hannan, & Tuma, 1990) effect on maintaining marital stability. Despite the correlations between income level and marital satisfaction, we could find no information about whether income supplements improve the quality of relationship between the partners. It may be that job training and other forms of interventions to reduce poverty may have beneficial effects on the family, but there is no evidence (yet) that they improve the quality of relationships between partners, or between parents and children. It is curious that the policy argument has been framed in terms of a choice between relationship and economic interventions. Why not develop a program that combines both, and tests them in additive combinations: CRE alone, economic intervention alone, or relationship and economic intervention combined? Only in this way will policymakers be able to make empirically informed decisions about the value added by either approach.

Current Policy Issues in the United States and the United Kingdom

In our view, controversies in both the United States and the United Kingdom about government funding of CRE programs have been complicated by the confounding of four different issues: (a) an argument about the wish to promote marriage and reduce divorce, sometimes stated in moral-religious terms, sometimes in economic terms; (b) an empirical argument about whether increasing the stability of marriages will reduce family poverty; (c) a social policy argument about the need to strengthen couple relationships (largely noncontroversial); and (d) a concern, relatively unexplored until recently, about whether CRE interventions benefit couples and their children. Some, but certainly not all, of the opposition to government funding of programs to strengthen couple relationships have come from critics who would support the funding of programs to strengthen couple relationships, but oppose programs to promote marriage as one solution to the poverty associated with single-parent families.

Our recommendation to combine CRE interventions with economic and other kinds of interventions to benefit families runs into an immediate obstacle, because federal, state, and county government departments serving families are organized in independent silos, with support for mothers and children in one department, father-involvement programs in another, and jobs programs in yet another. The same is true of nongovernmental public and private agencies addressing needs of families. The important of reducing or eliminating silos was brought home to us in our summaries of CRE and father-involvement programs. Our perception is that father-involvement interventions became more successful when they included more focus on family relationships. Reciprocally, CRE interventions, by their very nature, include fathers and often produce enhanced involvement of fathers in the rearing of their children. The data from the SFI project (P. A. Cowan et al., 2009) indicate that a couples approach to father involvement had a broader impact on the family than group meetings that included only fathers, with the same facilitators and curriculum. Given that digital media now make it possible for individuals across the world to communicate with each other, might it be possible to establish communication links across corridors and buildings, so that multiple approaches to strengthening families can result in even more widespread benefits for parents and their children?

We began this article by describing similar current U.S. and U.K. government policy decisions about supporting preventive interventions for couples. The similarity of goals—strengthening couple relationships to increase father involvement and family stability in the service of providing more supportive environments for children—has not led to the same decisions about what kinds of programs to fund. In the United States, given a historical absence of government and institutional support for couple relationships (e.g., most insurance companies will not pay for couples therapy), beliefs about the need to strengthen marriage led to the adoption of a currently popular modality—groups for CRE. As we have seen, results from the U.S. large-scale funding of the BSF and SHM programs have recently been released, to mixed reviews. Reacting to the fact that most of the smaller programs funded by the original $100 million in Healthy Marriage Funds and the more recent allocation of $75 million have produced no systematic evidence of positive effects, the Administration for Children and Families has hired Mathematica, a policy research evaluation company, to lead Parents and Children Together, a systematic evaluation of some of the newer CRE programs, with a specific focus not only on potential benefits for couple relationship quality but
also for economic self-sufficiency and the well-being of children. A similar endeavor, the Fatherhood Research and Practice Network, has been funded by the Administration for Children and Families to offer small grants to encourage research on father involvement, with one of three aims being to (a) plan, initiate, and carry out a research agenda to build an evidence base for effective responsible fatherhood interventions and the development of appropriate measures; (b) develop capacity within the evaluator and practitioner communities to conduct and participate in high-quality evaluation studies; and (c) disseminate findings and best practices. One of three working groups of this new organization focuses specifically on defining new directions for the evaluation of coparenting interventions. It seems that although large-scale intervention programs are not going to receive government funds in the near future, smaller studies aimed at providing a better evidence base for couple relationship strengthening programs will continue to receive federal government support.

The U.K. government, with a history of ongoing support for professional couples counseling and therapy as part of the National Health Service, recently chose to create additional funding for services directed to one couple at a time, along with some support for marriage preparation couples groups. This has not prevented the government from looking to the couples-group interventions in the United States and beginning to consider CRE as a possible option, especially for low-income families. The Department for Education has recently allocated £2.9 million ($4.84 million) to the Tavistock Centre for Couple Relationships and to Family Action, for a trial of the SFI intervention for low-income and vulnerable families (P. A. Cowan et al., 2009), renamed “Parents as Partners.” Because the emphasis of government funding has shifted from couple support to parenting support, depending on whether the Conservative Party or Labour Party is in power, future directions in the United Kingdom will depend on the outcome of elections to be held in May 2015.

It seems that despite the current controversies in the research literature concerning the benefits of couple relationship strengthening interventions, support for trials of small, well-evaluated CRE programs will continue. Our reading of the literature leads us to the conclusion that these trials are warranted. A more extensive research program is needed before final decisions are made about whether a new generation of preventive interventions for coparenting partners can fulfill the original promise of CRE to strengthen family relationships in ways that foster both parents’ and children’s healthy development.

References

See http://www.acf.hhs.gov/programs/opre/research/project/the-fatherhood-research-and-practice-network


Received January 21, 2014
Revision received July 1, 2014
Accepted July 2, 2014