Received: 17 December 2020 | Revised: 11 May 2021 | Accepted: 26 September 2021

DOI: 10.1111/famp.12739

FAMILY PROCESS

ORIGINAL ARTICLE

How effective are ACF-funded couple relationship education programs? A meta-analytic study

Alan J. Hawkins | Sarah Hokanson | Eden Loveridge | Emily Milius | Misha Duncan | McCall Booth | Brittany Pollard

School of Family Life, Brigham Young University, Provo, Utah, USA

Correspondence
Alan J. Hawkins, School of Family Life, Brigham Young University, Provo, UT, USA.
Email: hawkinsa@byu.edu

Abstract

Since 2006, the U.S. Administration for Children and Families (ACF) has allocated $1.2 billion to a Healthy Marriage and Relationship Education (HMRE) policy initiative that provides grants to community organizations to support relationship education (RE) services for lower income couples and individuals. The policy aim was to help disadvantaged couples and individuals form and sustain healthy, stable relationships and marriages. A significant body of research on the effectiveness of these programs has now accumulated. This meta-analytic study reviews all evaluation research reports of adult couple relationship education (CRE) programs supported by the ACF policy initiative to examine their impact on an array of couple, family, and individual well-being outcomes. Overall, our review of 32 control-group studies found a range of small but significant effects for couple relationship quality (d = .114), relationship skills (d = .132), mental health (d = .074), and coparenting (d = .033), but non-significant effects for relationship stability, parenting, and child well-being. Supplemental analyses with 19 1-group/pre-post studies showed larger effects. Planned moderator analyses explored significant heterogeneity in most effects, however, revealing interesting implications for practice and research going forward.

KEYWORDS

couple relationship education, meta-analysis, relationship education programs, social policy

INTRODUCTION

In 2006, the U.S. Administration for Children and Families (ACF) launched a new social policy initiative to provide relationship education services to lower income individuals and couples to reduce family instability (Hawkins, 2019). Assistant Secretary Wade F. Horn wanted federal policy to attend directly to the reality that significant numbers of children were experiencing the risks associated with family instability, which contributes, on average, to a host of negative outcomes for children (and adults) and significant public costs (Sawhill, 2014). This controversial new policy initiative, now known as the Healthy Marriage and Relationship Education (HMRE) initiative, has provided a total of $1.2 billion to community organizations to support relationship education (RE) services aimed at helping lower income couples and individuals gain knowledge and skills to form and sustain healthy relationships and marriages.

Since 2006, an estimated 2.5 million people have participated in HMRE programs at a median cost of about $400 per participant (Hawkins, 2019). Demographically, a little more than a third of participants were White, nearly 30% were African American, and nearly 30% were Hispanic. More than half (56%) of participants had incomes below the federal poverty line, with another third (30%) below twice the poverty level. According to one summary, about half of participating couples were distressed and about 10% had experienced relationship violence in the past few years (Bradford et al., 2015).

The HMRE policy initiative sparked considerable controversy and debate among social science and policy scholars (Hawkins, 2019). Several scholars argued that effective policy to address family instability should focus on improving the social and economic ecologies of these relationships so that they could thrive rather than directly targeting lower income couples' relationship (soft) skills (Randles, 2017; Trail & Karney, 2012). Other scholars worried that RE programs that were first developed for more advantaged couples would be ineffective for lower income couples and the substantial stresses of their everyday lives (Johnson, 2012).

A distinguishing feature of the HMRE policy initiative is that ACF launched rigorous evaluations of the initiative from the start, consistent with the Obama administration's evidence-based policy initiative (Haskins & Margolis, 2015). As early rigorous evaluation studies of HMRE programs emerged that showed no or only small effects, some scholars urged ACF to abandon its new policy initiative (Reeves, 2014), but these critiques focused on early research in the HMRE evaluation lifespan. A more robust body of HMRE program evaluation research has accumulated now with more positive results.

This meta-analytic study summarizes the full body of evaluation research on ACF-funded HMRE programs targeted to lower income adult couples as one way to assess the potential of this new social policy initiative. Of course, a good number of meta-analytic studies have focused on the effectiveness of relationship education in general (Blanchard et al., 2009; Fawcett et al., 2010; Hawkins et al., 2008, 2012; Lucier-Greer & Adler-Baeder, 2012; McAllister et al., 2012; Pinquart & Teubert, 2010; Simpson et al., 2018; Spencer & Anderson, 2021), but none has focused exclusively on ACF-funded HMRE programs. Two meta-analytic studies (Arnold & Beelman, 2019; Hawkins & Erickson, 2015) focused on programs for lower income couples, but they included non-ACF-funded programs in their review (and the Arnold & Beelman study included fathering programs). Moreover, most of these meta-analytic reviews have not captured the large number of relevant studies that have been published in the last five years or the wider range of outcomes often assessed in recent studies.

METHOD

Inclusion/exclusion criteria and search procedures

For inclusion in this study, a report had to meet the following criteria: (a) Empirical evaluation of a relationship education program. (We included both control-group and 1-group/pre-post

studies but analyzed them separately). Some CRE programs included coparenting, parenting, or fathering content. To include studies of these kinds of hybrid programs, we required that the primary focus of the evaluated program—at least 50% of the curriculum—be on the couple relationship. (b) Quantitative evaluation. A substantial number of evaluation studies that reported only qualitative data from CRE programs were excluded. (But interested readers can read a summary of these 14 studies in the Appendix S2.) (c) Program funded by ACF as part of the HMRE initiative. A few CRE studies were funded by other federal government offices (e.g., NIH) or by state government offices (e.g., California Office of Child Abuse Protection). We excluded these studies in order to retain a clear focus on the ACF HMRE policy initiative. (d) Program targeted to adult couples in a relationship. A number of studies evaluated relationship education programs for youth and single adults, but because the curricula for these programs are quite different and the studies targeted different outcomes from adult CRE programs, we excluded them from this systematic review. (See Simpson et al. (2018), for a meta-analytic review of these programs.) Similarly, some CRE programs allowed individuals to attend without partners. When the proportion of lone-attenders was less than one third of the sample, we coded the study. (Some of these studies separated out the individual participants from the couple participants in analyses; we did not code data for the individual participants.) (e) Sufficient data to compute an effect size on one or more of the following generic outcomes: relationship quality, relationship stability, relationship skills, intimate partner violence/aggression (IPV), mental health, coparenting, parenting, and child well-being/behavior. A few studies did not report analyses from which we could compute simple effect sizes (e.g., only reported multilevel modeling parameter estimates of program effects). In these cases, we contacted the researchers to ask for basic data (*M*s, *SD*s, *N*s) to compute effect sizes. Most (but not all) researchers responded with the needed data.

Our search process included several steps. To identify potential studies, we first searched two electronic databases (PsychINFO; Family & Society Studies Worldwide). We experimented with a range of search terms and combinations of terms; the standard search term "relationship education program" proved to be the best starting point because it captured RE evaluation studies while limiting the number of (more plentiful) correlational studies of couples. We limited the search to studies published from January 2009 to May 2020 because 2009 is when the earliest evaluation study of an ACF-funded HMRE program was published. In addition, we looked at several review articles and recently published reports for citations of potential studies. Also, we examined references from studies that met our search criteria for potential studies that may have been missed, including so-called "grey literature" (e.g., technical reports and dissertations). Further, we browsed the websites of ACF-funded programs known to be actively publishing evaluation studies for lists of research reports and published studies. During the process, we also contacted active researchers in this field for clarifications about certain studies (e.g., overlapping samples) and asked them whether they had any studies (published or unpublished) not on our accepted-study list. This yielded several more recent (in press) studies.

Figure 1 graphically summarizes the search and inclusion/exclusion decision process. (A reference list of accepted studies is provided as an Appendix S1.) Ultimately, we coded 51 studies (32 control-group studies and 19 1-group/pre-post studies). These 51 studies yielded 817 effect sizes. Note that this final figure of 51 studies includes several adjustments made during the coding process. Some reports included multiple independent sites/samples or reported effects for multiple treatment groups, so they were coded separately. This added to the total number of studies. On the other hand, numerous reports included samples that overlapped with other reports (e.g., ACF grantee organizations publishing multiple articles using the same or similar sample). Including studies with overlapping samples can bias results (Bon & Rachinger, 2017). Rather than employing the generalized-weights approach described by Bon and Rachinger (2017), we did the work of identifying overlapping samples and excluding them from our final

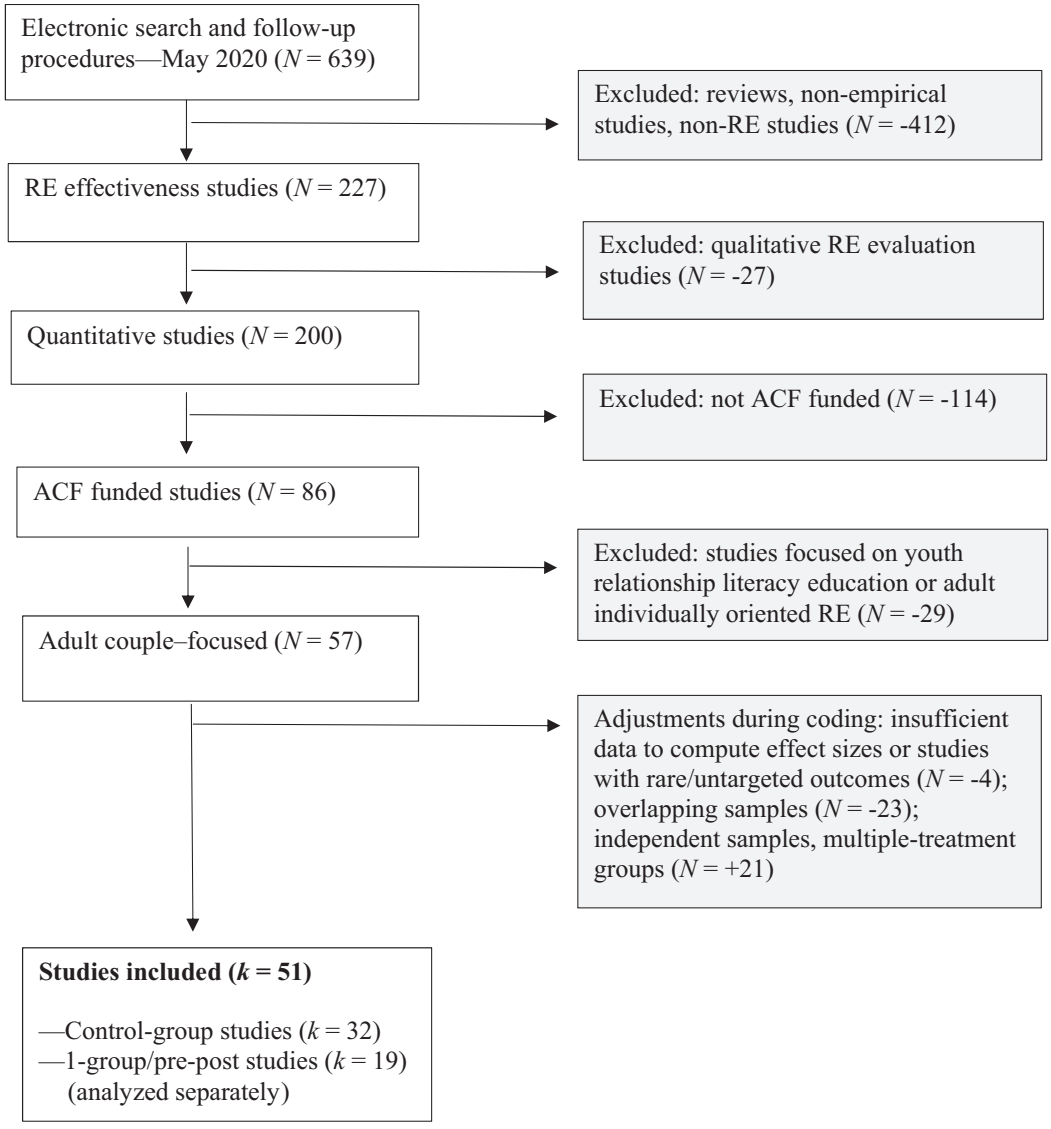

FIGURE 1 Flow chart of Inclusion/Exclusion Decisions

set of studies. Or, if they added outcomes not reported in previous studies, we combined overlapping-sample studies for coding purposes. Also, a few studies employed alternative-treatment comparison groups that provided some kind of relationship intervention. But because the comparison group experienced some kind of treatment, it was not appropriate to include these effect sizes with studies of no-treatment control groups. Instead, we coded the treatment group pre-post change (and the alternative-treatment group, if appropriate) as 1-group/pre-post studies. Finally, a couple of studies with wait-list control groups conducted longer-term follow-up assessments with the treatment groups after wait-list controls were terminated from the study. In these few cases, we coded the experimental-wait-list-control comparisons but also added the pre-to-follow-up within-group effect sizes to the 1-group/pre-post studies.

Based on this final set of studies, we conducted power analyses for this meta-analysis using Tiebel's meta-analysis power calculator (https://www.jtiebel.com/2018/08/26/how-to-calcu

late-statistical-power-of-a-meta-analysis/). A 2015 meta-analysis of the effects of CRE for lower income couples yielded an overall effect size of d = .07 for control-group studies and d = .35 for 1-group/pre-post studies (Hawkins & Erickson, 2015), although recent studies have shown larger effect sizes. For control-group study main effects, our power analyses suggested powers of .64–.86 to detect an effect size of .07 for our various outcome categories. Two outcome categories—parenting and child well-being—were underpowered (.64) to detect small effect sizes. Relationship quality, relationship skills, and IPV exceeded power of .80.

Computing effect sizes and plan for analyses

For control-group studies, we computed standardized mean difference effect sizes. For 1-group/pre-post studies, we computed standardized mean change effect sizes. (Note that when computing standardized mean change effect sizes, pre-post correlations of outcomes are required but often unreported. In these cases, we assigned a value of r = .50, as recommended by Rosenthal (1983). This assigned correlation is generally close to most actual unreported correlations and does not appear to bias the overall effect size (Nowak & Heinrichs, 2008).) We computed effect sizes based on the latest assessment available. For a few studies, this was the immediate post-intervention assessment, but most studies included follow-up assessments. Taking the latest assessment available allows for deterioration (or growth) of initial effects. However, in some 1-group/pre-post studies, follow-up attrition was high (>40%), so we coded just the immediate post-assessment. We coded so that positive effect sizes always reflected a positive intervention effect.

We employed Biostat's Comprehensive Meta-analysis III program to compute effect sizes. All effect sizes were weighted by the inverse variance; this gives greater weight to effect sizes with smaller standard errors (larger samples). We used random effects models to estimate overall effect sizes, which allow for the possibility that variation in the distribution of effect sizes is a result of not only sampling error but also differences in programs, intervention methods, and other factors (Borenstein et al., 2009). Overall effect sizes produced with random effects models are more generalizable to the large variety of CRE programs in the field.

Two teams of trained undergraduate coders did the initial coding of all study variables and effect sizes, but then these codes were checked by two trained graduate students and/or the first author. When coding differences emerged, coders returned to the report to further investigate and resolve them. We coded all study effect sizes for variables related to the following outcome categories: relationship quality (e.g., happiness, adjustment, warmth, and commitment), relationship stability (e.g., still together, divorce, divorce ideation, and future of relationship), relationship skills (e.g., positive/negative communication and effective problem solving), relationship violence (e.g., physical assault and psychological abuse), mental health (e.g., depression/anxiety), coparenting (e.g., quality and cooperation), parenting (e.g., parent–child relationship, parenting stress, and effective discipline), and child well-being (e.g., internalizing/externalizing behaviors).

Our primary focus was on rigorous control-group studies that are subject to fewer internal validity problems compared to 1-group/pre-post designs (Reichart, 2019). But for several reasons, we also examined pre-post studies. Early ACF support for CRE was characterized as demonstration grants, so rigorous evaluation was not emphasized. But a number of program administrators were active researchers and conducted their own intervention studies that included assessing participant change from pre- to post-treatment. Part of the rationale for including 1-group/pre-post studies, even with their greater susceptibility to internal validity threats, is that a large number of control-group studies have already established the positive impact of CRE programs, so the need for control groups in every study is reduced. Furthermore, these kinds of pre-experimental studies are useful to program administrators to

see whether any change is occurring, even if the magnitude of that change may be biased by internal validity threats. Accordingly, we conducted separate analyses for these studies.

In both sets of analyses, we employed two methods to check for potential missing studies or publication bias (studies with non-significant findings are less likely to be published). First, we conducted a careful and extensive search process, as detailed earlier. Second, we used the Trim and Fill funnel plot method (Duvall & Tweedie, 2000) that looks for a disproportionate number of small-*N* studies with effect sizes above the mean and then adjusts for any bias detected by imputing mirrored "missing" studies below the mean.

We expected substantial heterogeneity in the distribution of effect sizes due to the differences in interventions, target populations, methods, and other factors. Accordingly, although we report mean effect sizes, we focus more on heterogeneity. We report the range of effect sizes to draw attention to potential outlier studies that may indicate especially low or high effects. In addition, we report the prediction intervals (*PI*) for each of our seven outcome categories. The prediction interval specifies the two effect size values between which 95% of the true effect sizes would be expected to fall. When the *PI* indicated substantial heterogeneity, we pursued moderator analyses to try to explain why some studies had stronger effect sizes. However, we limited moderator analyses to eight planned, substantive variables that likely explained heterogeneity in the distribution of effect sizes and that had application for practice. These analyses were performed on an aggregation of outcomes that indicated substantial heterogeneity to maintain as much statistical power as possible and because we did not have specific hypotheses of how distinct moderators would operate differently on specific outcome categories. Of course, tests of moderation are observational rather than experimental; a significant difference in a moderator test should be tested directly in primary experimental studies before drawing causal inferences.

The planned moderator analyses included the following tests for differences in study methods as well as participant and program characteristics: (a) Assessment timing. We tested whether timing of last assessment influenced the effect size. (b) Relationship status. Some studies have found that married participants, who are generally more stable and more committed, benefit more from CRE (Hawkins & Erickson, 2015) but other studies have not found relationship status differences (Moore et al., 2018). (c) Relationship distress. Numerous studies have found that couples in more distress at the start of the intervention benefit more from CRE (Amato, 2014; Hawkins et al., 2017) perhaps because they have more room to improve and because they are more motivated. (d) Economic disadvantage. Similarly, many studies have found that economically disadvantaged couples benefit more from CRE (Hawkins et al., 2017) but one meta-analysis found that studies with samples of mostly near-poor participants did better than studies with samples of mostly poor participants (Hawkins & Erickson, 2015). (e) Intervention dosage. Some studies have found that programs with low dosage have smaller effects (Hawkins et al., 2008), but there may not be an advantage for the highest dosage interventions (Stanley et al., 2006). (f) Program funding. We tested for differences between the three large-scale, RCT studies funded by ACF's Office of Planning, Research, and Evaluation (BSF, SHM, and PACT) and the studies of programs funded by the ACF Office of Family Assistance. (g) Initiative phase. Because anecdotal observation of studies over the past decade indicates that program effects have been improving, we looked for evidence that studies of earlier CRE programs found smaller effects than studies of later programs, suggesting a learning curve in the HMRE policy initiative. (h) Analyses. Finally, we coded for whether studies reported intent-to-treat (ITT) analyses on all participants regardless of actual program participation or treatment-on-the-treated (ToT) analyses on those participants who received a fair dosage of the intervention. Both analytic approaches yield valuable information, but ITT analyses are more conservative and best estimate the effect of treatment at a population level.

Throughout this report, we employ the Lipsey and Wilson (1993) metric designating effect sizes less than .30 as small, .30–.66 as medium, and .67 or greater as large. These categories are

derived from a distribution of more than 300 meta-analytic studies of psychological, behavioral, and educational interventions.

RESULTS

Primary analyses for control-group studies

We conducted a preliminary analysis to test whether RCT studies differed significantly from a few (k = 3) quasi-experimental control-group studies in which we could not confirm that participants were randomly assigned. The difference was not significant (d_{exp} = .084, k = 27; d_{quasi} = .131, k = 3; Q = .16, p = .693). Accordingly, we combined these studies for analyses.

Effect sizes for the eight outcome categories for control-group studies are presented in the upper panel of Table 1. The aggregated program effect was d = .074 (k = 32, p < .001), but that single figure masks effect variation between the eight outcome categories and substantial heterogeneity within many of the outcome categories. Relationship skills (d = .132, k = 29, p < .001) and relationship quality (d = .114, k = 29, p < .001) had the strongest (but still small) effects; effects for IPV, parenting, and child well-being were not significant, but the effect for coparenting (d = .033, k = 21, p < .05) was significant though very small.

We explored how robust these effect size estimates were with trim-and-fill procedures that attempt to detect and adjust for missing-study bias. Of course, there are various reasons why smaller-N studies could have larger effect sizes, but if publication bias is one of them, then the adjusted effect sizes may be a better representation of the true effect size. However, we found evidence of potential upward bias with only one outcome (IPV: adjusted d = −.071, *ns*).

The prediction intervals of the outcome categories revealed substantial heterogeneity in five of the outcome categories (relationship quality, stability, skills, IPV, and mental health), so we pursued the set of planned moderator analyses on the aggregate of these five outcomes. (Note that all cell sizes in the moderator analyses had five or more studies contributing to the effect size, which diminished the risk of overinterpreting potentially unreliable group differences). These analyses are detailed in the upper panel of Table 2 and reveal some important variation. First, the three OPRE-funded studies (BSF, SHM, PACT), which contributed more than half of the control-group studies analyzed, had significantly smaller effect sizes than did studies that were supported with OFA grant funds (d_{OPRE} = .038, k = 19, p = .089; d_{OFA} = .265, k = 13, p < .001; Q = 27.3, p < .001). This OPRE-OFA difference is likely implicated in several other comparisons. Studies published in the earlier years of the ACF-HMRE initiative (2009–2014) had significantly smaller effects than those published after 2014 (d_{early} = .057, k = 24, p = .023; d_{late} = .192, k = 8, p < .001; Q = 7.2, p < .01). Two of the 18 OPRE-funded studies were early, and most of the OFA-funded studies were later, but the PACT OPRE study published in 2018 had somewhat stronger effects than the earlier BSF and SHM studies. (Note that when controlling for OPRE vs. OFA studies, early vs. late studies was no longer statistically significant). Also, studies with effect sizes computed with intent-to-treat analyses were significantly lower than studies with effect sizes computed with treatment-on-the-treated analyses (d_{ITT} = .072, k = 25, p < .01; d_{ToT} = .335, k = 6, p < .001; Q = 10.7, p < .001); OPRE-funded studies were all conducted using ITT analyses, while OFA studies were evenly divided between ITT and ToT analyses. Furthermore, the timing of follow-up assessments was noteworthy. Studies with follow-up assessments longer than 6 months (dominated by the OPRE-funded studies) had significantly smaller effect sizes compared to studies with short-term follow-up assessments (d_{long} = .044, k = 21, p = .051; d_{short} = .270, k = 11, p < .001; Q = 9.4, p < .001). (Note that when controlling for OPRE vs. OFA studies, follow-up timing was no longer statistically significant.) Finally, intervention dosage also was a significant moderator of effects: moderate-dosage programs (8–19 h) had stronger effects than high-dosage programs (20+ hours) (d_{mod} = .208, k = 11, p

TABLE 1 Effect sizes for outcome categories

Moderator	*k*	*d*[a]	Range (Low–High)	*CI* (Low–High)	*PI* (Low–High)	*p*	*Q*	*p*	Adjusted *d* (trimmed studies)
Control-group studies									
Relationship aggregate	32	.100	−.163–1.030	.049–.151	−.147–.347	.000	138.7	.000	.066 (0)
Relationship quality	29	.114	−.133–1.030	.059–.169	−.144–.432	.000	128.3	.000	.096 (0)
Relationship stability	21	.041	−.270–.530	−.046–.128	−.350–.388	.354	117.5	.000	.043 (0)
Relationship skills	29	.132	−.025–.780	.078–.185	−.117–.381	.000	101.5	.000	.108 (0)
IPV	22	.014	−.756–.327	.074–.102	−.373–.401	.759	210.2	.000	−.071 (7)
Mental health	25	.074	−.185–.254	.047–.107	.000–.148	.000	31.4	.142	.076 (0)
Parent/child aggregate	22	.019	−.106–.160	−.004–.043	−.018–.052	.110	16.3	.755	.014 (2)
Coparenting	21	.033	−.159–.223	−.002–.065	−.001–.073	.038	19.8	.469	.033 (0)
Parenting	20	.005	−.106–.107	−.025–.035	−.033–.043	.744	10.5	.939	.002 (1)
Child well-being	16	.027	−.050–.147	−.015–.068	.007–.047	.209	4.2	.997	.027 (0)
Program effect (all outcomes)	32	.074	−.107–1.030	.035–.113	−.089–.237	.000	82.2	.000	.054 (0)
1-group/pre-post studies									
Relationship aggregate	17	.474	−.072–1.397	.331–.617	−.682–1.630	.000	489.7	.000	.357 (0)
Relationship quality	16	.448	−.072–1.401	.315–.581	−.122–1.018	.000	744.2	.000	.351 (0)
Relationship stability	5	.602	.117–.891	.341–.863	−1.17–2.38	.000	213.8	.000	.460 (1)
Relationship skills	11	.700	.232–1.390	.467–.934	−.731–2.131	.000	1,202	.000	.343 (0)
IPV	3	.199	.170–.502	.105–.293	−.531–.929	.000	4.2	.459	.188 (1)
Mental health	6	.074	.233–.712	.264–.642	−.931–1.079	.000	49.1	.000	.384 (0)
Parent/child aggregate	5	.277	.143–.521	.082–.472	−1.23–1.78	.005	84.6	.000	.356 (0)
Coparenting	4	.297	.152–.521	.081–.513	−.741–1.335	.007	71.3	.000	.375 (0)
Parenting	2	.135	.130–.135	.070–.200	—	.000	.003	.954	—
Child well-being	1	.347	—	.211–.482	—	.000	0.0	1.00	—
Program effect (all outcomes)	19	.451	−.072–1.397	.319–.582	−.155–1.057	.000	800.5	.000	.408 (0)

[a]For control-group studies, standardized difference at last assessment; for 1-group/pre-post studies, standardized mean gain from pretest to last assessment.

TABLE 2 Results of planned moderator analyses

Moderator	*k*	*d*	*CI* (Low–High)	*p*	*Q* (diff.)	*p*
Control-group studies[a]						
Program funding						
ACF-OFA	13	.265	.181–.347	.000		
ACF-OPRE (BSF/SHM/PACT)	19	.038	−.006–.081	.089	27.3	.000
Initiative phase						
Earlier (2009–2014)	24	.057	.008–.106	.023		
Later (2015–2020)	8	.192	.107–.276	.000	7.2	.007
Analyses						
Intent-to-treat	25	.072	.026–.118	.002		
Treatment-on-the-treated	6	.335	.184–.485	.000	10.7	.001
Follow-up assessment timing						
</= 6 months	11	.339	.182–.295	.000		
>6 months	21	.044	.000–.087	.051	9.4	.000
Intervention dosage						
Medium (8–19 h)	11	.208	.121–.336	.000		
High (20+ h)	20	.053	.004–.101	.032	10.5	.01
Relationship status						
Mostly married	17	.119	.065–.173	.000		
Mostly unmarried	8	−.025	−.092–.042	.467		
Mixed	6	.226	.131–.321	.000	20.3	.000
Relationship distress						
<50%	5	.147	.020–.273	.023		
>/= 50%	20	.067	.013–.122	.016	1.3	.260
Economic disadvantage						
Poor	15	.098	.028–.160	.006		
Near poor	15	.092	.022–.150	.009	.07	.798
1-group/pre-post studies[b]						
Initiative phase						
Earlier (2009–2014)	12	.465	.289–.641	.000	0.0	.852
Later (2015–2020)	6	.493	.252–.734	.000		
Analyses						
Intent-to-treat	6	.662	.425–.898	.000	3.8	.052
Treatment-on-the-treated	10	.361	.171–.550	.000		
Follow-up assessment timing						
</= 6 months	12	.491	.322–.660	.000	0.1	.741
> 6 months	6	.440	.192–.688	.001		
Relationship status						
Mostly married	8	.422	.272–.511	.000	3.3	.192
Mostly unmarried	1	.733	.356–1.11	.000		
Mixed	6	.353	.194–.513	.000		

[a]Included outcomes: relationship quality, stability, skills, IPV, mental health.

[b]Included outcomes: relationship quality, stability, skills, coparenting, mental health.

< .001; d_{high} = .053, k = 20, p < .05; Q = 10.5, p < .01). The high-dosage programs were almost exclusively OPRE-funded studies, and the moderate-dosage programs were almost exclusively OFA-funded studies. (Note that when controlling for OPRE vs. OFA studies, dosage was no longer statistically significant).

Participant characteristics of economic disadvantage and relationship distress were not significant moderators of effect sizes. However, there was an interesting difference for participant relationship status. Studies with programs that included both married and unmarried couples had the largest effect sizes (d = .226, k = 6, p < .001), larger than for studies with programs of just unmarried couples (d = −.025, k = 8, *ns*; Q = 18.2, p < .001), but not statistically larger than programs with only married couples (d = .119, k = 17, p < .001; Q = 2.7, *ns*).

Secondary analyses for 1-group/pre-post studies

Effect sizes for the eight outcome categories for 1-group/pre-post studies are presented in the lower panel of Table 1. Overall, effect sizes for these 19 studies were substantially larger than for control-group studies. The aggregated program effect was d = .451 (k = 19, p < .001), but that single figure masks effect variation between the eight outcome categories. Relationship skills (d = .700, k = 11, p < .001), relationship stability (d = .602, k = 5, p < .001), and relationship quality (d = .448, k = 16, p < .001) had the strongest effects; the effects for mental health (d = .074, k = 6, p < .001) and IPV (d = .199, k = 3, p < .001) were small but significant. There were only five studies that measured effects for coparenting, parenting, or child well-being. When aggregating these three outcome categories, the effect size was small (d = .277, k = 5, p < .01). Notably, there is an important confound: only one 1-group/pre-post study employed a sample of unmarried couples. The other studies had either married or mixed participants.

The prediction intervals of the outcome categories revealed substantial heterogeneity in five of the outcome categories (relationship quality, stability, skills, mental health, and coparenting), so we pursued the set of planned moderator analyses on the aggregate of these five outcomes. However, half of the planned moderator analyses were hampered by small cell sizes (so we do not report them here) and the remaining moderator analyses (detailed in the lower panel of Table 2) did not reveal any statistically significant differences.

DISCUSSION

The results of this meta-analysis of ACF-funded HMRE programs can support concerns about as well as hopes for the social policy initiative. Critics will highlight that the most rigorous, large-scale (OPRE) studies with the longest follow-up assessments and those using the most conservative analyses (intent-to-treat) failed to produce effects on IPV, parenting, and child outcomes. And even the significant (but still small) effects on relationship quality (d = .114) and skills (d = .132) have not translated into significant effects on family stability (d = .041), perhaps the most important objective from a social policy perspective (although the most recent OPRE study [PACT: Moore et al., 2018] found reduced divorce rates for married couples). Also, studies with the longest follow-up assessments yielded very small effects (although this analysis is confounded with the OPRE-ACF moderator). Moreover, the effects found here are weaker than those found in meta-analytic studies with more advantaged samples (Blanchard et al., 2009; Fawcett et al., 2010; Hawkins et al., 2008; Pinquart & Teubert, 2010), suggesting that diverse, lower income couples may not be in a position to gain as much from these kinds of soft skills, relationship-strengthening programs as middle-class couples.

It is valuable to place these small effects in context, however. A number of factors lead to an expectation of relatively small effect sizes for these kinds of programs. For one, these

CRE programs are not tailored, one-on-one clinical interventions, which show stronger effects (Bradbury & Bodenmann, 2020; Roddy et al., 2020); all participants receive the same intervention regardless of their unique circumstances and needs. In addition, these are field studies, not laboratory trials with carefully controlled conditions; field studies of couple interventions usually yield smaller effect sizes than laboratory studies (Bradbury & Bodenmann, 2020). And program participation rates are often lower in field studies (e.g., Moore et al., 2012), so many participants do not receive the full intended treatment. Of course, some seek out these interventions to *maintain* an already healthy relationship rather than substantially improve it, which limits room for improvement in outcomes. At the other end of the spectrum, however, many couples who participate in these programs are ambivalent about or lack hope for the future of their relationship. Such feelings may suppress motivation to work on the relationship for some participants. Indeed, ending unhealthy relationships with poor prospects should be seen as a successful result of CRE, albeit one that will register statistically as lower scores on outcomes. Also, the stressful conditions of these program participants' lives may work against maintenance of learned skills over time (Bradbury & Bodenmann, 2020; Halpern-Meekin, 2019), especially for the most disadvantaged, and this is why some critics of the policy have predicted that the programs will not be successful (Randles, 2017). Given these realities facing ACF-funded CRE programs, we should expect average effect sizes to be relatively small.

Expectations notwithstanding, of course, critics may still argue that the small effects are justification for ACF to retire the 15-year HMRE policy initiative. Proponents, however, would push to place the effect sizes observed in this study in a broader social policy context by comparing these effects to those from other federal social policy initiatives, some funded at much greater expense over a longer period of time. For instance, a rigorous, major impact evaluation of the well-known Head Start program found that early positive impacts on cognitive, social, and health outcomes for Head Start children mostly dissipated by the third-grade follow-up (Puma et al., 2012). The few positive effects that were found were generally between .10 and .20, and several negative effects in the same range were observed. Similarly, a rigorous, major impact evaluation of home visiting programs to support low-income new mothers and improve child well-being found few significant impacts at the 18-month follow-up on maternal health, economic self-sufficiency, or child health and development, although there were small, positive impacts on preventing IPV and on some parenting outcomes, with effects generally between .08 and .11 (Michalopoulos et al., 2019).

In comparison, effect sizes for federally supported CRE programs are similar to these two policy initiatives. Also, a rigorous, major impact study of the parallel ACF responsible fatherhood initiative found that these educational programs and services had some small impacts on fathers' parenting quality, with effects usually between .10 and .20, but no effects on coparenting quality or fathers' mental health (Avellar et al., 2018). Similarly, a meta-analysis of responsible fatherhood programs (including ones not funded by ACF) for non-resident fathers found an overall effect size of .10 (Holmes et al., 2020). An argument to terminate the ACF-HMRE policy initiative would need to consider, in fairness, a similar fate for other family policy initiatives. Proponents could also argue that this social policy initiative is relatively young compared to such initiatives as early childhood education and more recent research appears to be producing stronger effects than the first wave of evaluation work.

Moderators of program effects

Moreover, when evaluating this (and other) policy initiatives, proponents may argue that a focus on average effects is not especially helpful when it masks variation (Bradbury & Bodenmann, 2020). It may be more important to attend to the variation inherent in these studies. For instance, we found evidence in our moderator analyses that CRE programs targeting

unmarried couples have been ineffective overall. Yet, we also found evidence that programs that put married and unmarried couples together in the intervention produced the strongest effects (d = .226). While it is unclear exactly why this is the case, program administrators still could consider capitalizing on this finding by putting unmarried and married couples together in groups. In addition, we found that programs with moderate dosage actually outperformed programs with higher dosage (20+ hours). (The dosage difference was confounded by the highest dosage studies all being in the OPRE study group, so this finding should be interpreted cautiously.) Also, online programs (that were some of the lowest dosage programs reviewed) produced some of the largest effect sizes in this study (d = .26, k = 4) and in studies of online programs not reviewed here (see Spencer & Anderson, 2021). Doss, Knopp et al. (2020) have speculated that online delivery of CRE attracts couples with greater motivation to improve their relationship (in the "action" stage of change) and gets the intervention to them in a more timely way. And there are emerging anecdotal data that online delivery substantially boosts participation numbers (Lutui & Hawkins, 2021). This suggests a path to more efficient delivery of these services that program administrators should consider. ACF programs by necessity learned a great deal about online delivery of programs during the COVID-19 pandemic (Lutui & Hawkins, 2021). ACF should consider funding more online delivery of CRE and perhaps revisit its recent requirement that grantees offer programs of at least 12 h.

On the other hand, we did not find significant variation in program effects for economic disadvantage and relationship distress. That is, while effects were small, they were similar for distressed and non-distressed couples and for poor and near-poor couples. These findings stand in contrast to a growing body of primary research that documents stronger effects for more disadvantaged and more distressed couples (Palmer & Hawkins, 2019). A possible explanation for this is that these programs have been successful in reaching the most disadvantaged families. Previous work in this area has noted that the poorest participants may struggle to benefit fully from these interventions compared to near-poor and middle-class participants (Doss, Knopp et al., 2020; Hawkins & Erickson, 2015). There are two potential implications of this observation. One is that program developers still have work to do to find more effective ways to strengthen couple relationships of the most disadvantaged couples. But another is that these programs may be most useful when targeted to near-poor families rather than the poorest, whose lives are too materially stressed to be able to benefit from a focus on soft skills.

The effect sizes observed in this meta-analysis for 1-group/pre-post studies were substantially larger than those for control-group studies for all outcome categories except mental health, underlining the importance of research designs that control for more potential confounds. However, these larger effect sizes may be an indicator of maximum effects we might expect in these kinds of studies rather than average effects. One-group/pre-post studies can be helpful in understanding intervention processes, but the lack of control groups limits how reliable they are as indicators of the magnitude of real change. There are design enhancements to 1-group/pre-post studies that can increase confidence in their outcomes, such as including multiple pretest observations to establish a stable baseline or adding a dependent variable to the outcome assessment that would not be expected to be affected by the intervention (Reichart, 2019). Unfortunately, these design adaptations have been rare in this field. Going forward, researchers should consider adaptions to strengthen their 1-group/pre-post designs.

Limitations and implications for future work

While there are important strengths to this study, we want to acknowledge limitations and caveats. We conducted the search for relevant studies near the end of the 2015–2020 grant cycle and a handful of grant holders told us that they would have further studies with fuller samples at the end of the year. We postponed analyses until recontacting these grant holders to get any

last-minute manuscripts. We received a few but there were still a couple unfinished studies and a few that were under review and researchers chose not to share them with us. Thus, we know that we were not able to include a small number of relevant, most recent studies (although they may have had overlapping samples). Also, we urge readers not to overgeneralize the results of this meta-analysis that was focused exclusively on CRE programs funded through a federal policy initiative. Most CRE programs are supported through other means and differ from the programs in this review in systematic ways. There are several other meta-analytic studies reviewing the effects of these other programs (Fawcett et al., 2010; Hawkins et al., 2008; Lucier-Greer & Adler-Baeder, 2012; Pinquart & Teubert, 2010) that have shown generally stronger effects.

Our meta-analysis raises important considerations for CRE program developers, administrators, and evaluators. First, although there is evidence that these programs are modestly improving couples' relational skills and sense of marital quality, how can programs be improved to produce stronger effects for relationship stability? The effect size for relationship stability for 1-group/pre-post studies ($d = .602$) suggests that moderate effects may be possible, but the control-group studies showed no effect. The recent PACT evaluation (Moore et al., 2018) encouragingly showed a small reduction in divorce rates, and the Oklahoma site of the BSF study also showed 20% improvement in stability at 3 years post-treatment, but these studies stand out as exceptions. (And in the larger body of evaluation work on non-ACF-funded CRE, only a handful of studies have looked at and found couple stability effects [Hahlweg & Richter, 2010; Stanley et al., 2014], though it is difficult to follow couples long enough for stability effects to emerge.) Family instability is implicated in a host of serious problems (Sawhill, 2014). Perhaps program developers could consider curriculum innovations to target relationship stability directly. Further, program developers need to improve effects for coparenting because many couples break up. Also, one of the challenges program developers face is that some participants are in unhealthy relationships and break-ups that happen sooner rather than later are probably a positive outcome. So evaluators will need to distinguish between break-ups in unhealthy vs. healthy relationships, which they have not done. Doing so will likely yield better stability effects.

Similarly, we did not find overall effects on IPV prevention in control-group studies, although a number of primary studies have found small, positive effects (Bradley et al., 2014; Moore et al., 2018). Better communication and problem-solving skills may not be translating straightforwardly into reduced risk for IPV. But because the public costs incurred by IPV are enormous (McLean & Bocinski, 2017), stronger effects for IPV prevention among the higher-risk couples that participate in CRE programs would quickly increase the benefit-to-cost ratio of this social policy, not to mention the important personal gains from reduced family violence. Programmers would do well to give more thought and experimentation to curriculum approaches that may reduce or prevent IPV.

Furthermore, we did not find significant program effects in the control-group studies for parenting and child well-being (and very small effects for coparenting). While a handful of studies of CRE programs have documented a positive pathway to child effects (Doss, Roddy et al., 2020; Feinberg et al., 2016; Pruett et al., 2019) and a recent meta-analysis of CRE programs' impact on these outcomes has shown some positive effects on coparenting and child well-being (Hawkins et al., 2020), programs funded by the ACF-HMRE policy initiative have not yet produced this desired overall outcome. A number of scholars have pleaded for breaking down the silos that typically divide couple, coparenting, fathering, and parenting education (Cowan & Cowan, 2019). A more holistic approach to intervention may produce better outcomes for children.

We see positive signs for these programs, however, in our finding that programs that began later in the ACF policy initiative had stronger outcomes than did the earliest programs. There were concerted efforts by ACF over the past 15 years to communicate lessons learned and best practices with other program administrators through various channels. These efforts may be paying off and suggest continued work to identify best practices through conceptual and analytic means and then share them generously and quickly.

REFERENCES

A list of references for studies included in the meta-analysis is available in online supplemental Appendix S1.

Amato, P. (2014). Does social and economic disadvantage moderate the effects of relationship education on unwed couples? An analysis of data from the 15-month Building Strong Families evaluation. *Family Relations*, *63*, 343–355. https://doi.org/10.1111/fare.12069

Arnold, L. S., & Beelman, A. (2019). The effects of RE in low-income couples: A meta-analysis of RCT studies. *Family Relations*, *68*, 22–38. https://doi.org/10.1111/fare.12325

Avellar, S., Covington, R., Moore, Q., Patnaik, A., & Wu, A. (2018). *Parents and Children Together: Effects of four responsible fatherhood programs for low-income fathers*. OPRE Report #2018-50. Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health & Human Services.

Blanchard, V. L., Hawkins, A. J., Baldwin, S. A., & Fawcett, E. B. (2009). Investigating the effects of marriage and relationship education on couples' communication skills: A meta-analytic study. *Journal of Family Psychology*, *23*, 203–214. https://doi.org/10.1037/a0015211

Bon, P. R. D., & Rachinger, H. (2017). *Estimate dependence in meta-analysis: A generalized-weights solution to sample overlap*. Retrieved from https://www.researchgate.net/publication/326019523_Estimate_Dependence_in_Meta-Analysis_A_Generalized-Weights_Solution_to_Sample_Overlap

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. John Wiley & Sons.

Bradbury, T. N., & Bodenmann, G. (2020). Interventions for couples. *Annual Review of Clinical Psychology*, *16*, 99–123. https://doi.org/10.1146/annurev-clinpsy-071519-020546

Bradford, A. B., Hawkins, A. J., & Acker, J. (2015). If we build it, they will come: Exploring policy and practice implications of public support for couple and relationship education for lower income and relationally distressed couples. *Family Process*, *54*, 639–654. https://doi.org/10.1111/famp.12151

Bradley, R. P. C., Drummey, K., Gottman, J. M., & Gottman, J. S. (2014). Treating couples who mutually exhibit violence or aggression: Reducing behaviors that show a susceptibility for violence. *Journal of Family Violence*, *29*, 549–558. https://doi.org/10.1007/s10896-014-9615-4

Cowan, C. P., & Cowan, P. A. (2019). Enhancing parenting effectiveness, fathers' involvement, couple relationship quality, and children's development: Breaking down the silos in family policy making and service delivery. *Journal of Family Theory & Review*, *11*, 92–111. https://doi.org/10.1111/jftr.12301

Doss, B. D., Knopp, K., Roddy, M. K., Rothman, K., Hatch, S. G., & Rhoades, G. (2020). Online programs improve relationship functioning for low-income couples: Results from a nationwide randomized controlled trial. *Journal of Consulting and Clinical Psychology*, *88*, 283–294. https://doi.org/10.1037/ccp0000479

Doss, B. D., Roddy, M. K., Llabre, M. M., Salivar, E. G., & Jensen-Doss, A. (2020). Improvements in coparenting conflict and child adjustment following an online program for relationship distress. *Journal of Family Psychology*, *34*, 68–78. https://doi.org/10.1037/fam0000582

Duvall, S., & Tweedie, R. L. (2000). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, *56*, 455–463. https://doi.org/10.1111/j.0006-341X.2000.00455.x

Fawcett, E. B., Hawkins, A. J., Blanchard, V. L., & Carroll, J. S. (2010). Do premarital education programs really work? A meta-analytic study. *Family Relations*, *59*, 232–239. https://doi.org/10.1111/j.1741-3729.2010.00598.x

Feinberg, M. E., Jones, D. E., Hostetler, M. L., Roettger, M. E., Paul, I. M., & Ehrenthal, D. B. (2016). Couple-focused prevention at the transition to parenthood, a randomized trial: Effects on co-parenting, parenting, family violence, and parent and child adjustment. *Prevention Science*, *17*(6), 751–764. https://doi.org/10.1007/s11121-016-0674-z

Hahlweg, K., & Richter, D. (2010). Prevention of marital instability and distress: Results of an 11-year longitudinal follow-up study. *Behaviour Research and Therapy*, *48*, 377–383. https://doi.org/10.1016/j.brat.2009.12.010

Halpern-Meekin, S. (2019). *Social poverty: Low-income parents and the struggle for family and community ties*. New York University.

Haskins, R., & Margolis, G. (2015). *Show me the evidence: Obama's fight for rigor and results in social policy*. Brookings Institution.

Hawkins, A. J. (2019, September 3). *Are federally supported relationship education programs for lower income individuals and couples working? A review of evaluation research*. American Enterprise Institute. Retrieved from https://www.aei.org/research-products/report/are-federally-supported-relationship-education-programs-for-lower-income-individuals-and-couples-working-a-review-of-evaluation-research/

Hawkins, A. J., Allen, S. E., & Yang, C. (2017). How does couple and relationship education affect relationship hope? An intervention-process study with lower income couples. *Family Relations*, *66*, 441–452. https://doi.org/10.1111/fare.12268

Hawkins, A. J., Blanchard, V. L., Baldwin, S. A., & Fawcett, E. B. (2008). Does marriage and relationship education work? A meta-analytic study. *Journal of Consulting and Clinical Psychology*, *76*, 723–734. https://doi.org/10.1037/a0012584

Hawkins, A. J., & Erickson, S. E. (2015). Is couple and relationship education effective for lower income participants? A meta-analytic study. *Journal of Family Psychology*, *29*, 59–68. https://doi.org/10.1037/fam0000045

Hawkins, A. J., Hill, M. S., Eliason, S. A., Simpson, D. M., & Hokanson, S. (2020). *Do couple relationship education programs affect coparenting, parenting, and child well-being outcomes? A meta-analytic study*. Manuscript submitted for review.

Hawkins, A. J., Stanley, S. M., Blanchard, V. L., & Albright, M. (2012). Exploring programmatic moderators of the effectiveness of marriage and relationship education: A meta-analytic study. *Behavior Therapy*, *43*, 77–87. https://doi.org/10.1016/j.beth.2010.12.006

Holmes, E. K., Egginton, B. M., Hawkins, A. J., Robbins, N. L., & Shafer, K. (2020). Do responsible fatherhood programs work? A comprehensive meta-analytic study. *Family Relations*, *69*, 967–982. https://doi.org/10.1111/fare.12435

Johnson, M. D. (2012). Healthy marriage initiatives: On the need for empiricism in policy implementation. *American Psychologist*, *67*, 296–308. https://doi.org/10.1037/a0027743

Lipsey, M. W., & Wilson, D. B. (1993). The efficacy of psychological, educational, and behavioral treatment: Confirmation from meta-analysis. *The American Psychologist*, *48*, 1181–1209. https://doi.org/10.1037/0003-066X.48.12.1181

Lucier-Greer, M., & Adler-Baeder, F. (2012). Does couple and relationship education work for individuals in stepfamilies? A meta-analytic study. *Family Relations*, *61*, 756–769. https://doi.org/10.1111/j.1741-3729.2012.00728.x

Lutui, S., & Hawkins, A. J. (2021, April 20). *What did the pandemic teach us about online relationship education and therapy?*. Institute for Family Studies Blog. Retrieved from https://ifstudies.org/blog/what-did-the-pandemic-teach-us-about-online-relationship-education-and-therapy

McAllister, S., Duncan, S. F., & Hawkins, A. J. (2012). Examining the early evidence for self-directed marriage and relationship education: A meta-analytic study. *Family Relations*, *61*, 742–755. https://doi.org/10.1111/j.1741-3729.2012.00736.x

McLean, G., & Bocinski, S. G. (2017, August 14). *The economic cost of intimate partner violence, sexual assault, and stalking*. IWPR#B367. Institute for Women's Policy Research. Retrieved from https://iwpr.org/publications/economic-cost-intimate-partner-violence-sexual-assault-stalking/

Michalopoulos, C., Faucetta, K., Hill, C. J., Portilla, X. A., Burrell, L., Lee, H., Duggan, A., & Knox, V. (2019). *Impacts on family outcomes of evidence-based early childhood home visiting: Results from the Mother and Infant Home Visiting Program Evaluation*. OPRE Report 2019-07. Washington, DC: Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health & Human Services.

Moore, Q., Avellar, S., Patnaik, A., Covington, R., & Wu, A. (2018). *Parents and Children Together: Effects of two healthy marriage programs for low-income couples*. OPRE Report Number 2018-58. Washington, DC: Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health & Human Services.

Moore, Q., Wood, R. G., Clarkwest, A., Killewald, A., & Monahan, S. (2012). *The long-term effects of Building Strong Families: A relationship skills education program for unmarried parents, technical supplement*. OPRE Report 2012-28C. Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

Nowak, C., & Heinrichs, N. (2008). A comprehensive meta-analysis of Triple-P Positive Parenting program using hierarchical linear modeling: Effectiveness and moderating variables. *Clinical Child and Family Psychological Review*, *11*(3), 114–144. https://doi.org/10.1007/s10567-008-0033-0

Palmer, H., & Hawkins, A. J. (2019, March 27). Who benefits most from relationship education? A review of recent relationship education research. *Relationship Educator*. Retrieved from https://relationshipeducator.org/blog/topical/who-benefits-most-from-relationship-education/

Pinquart, M., & Teubert, D. (2010). A meta-analytic study of couple interventions during the transition to parenthood. *Family Relations*, *59*, 221–231. https://doi.org/10.1111/j.1741-3729.2010.00597.x

Pruett, M. K., Cowan, P. A., Cowan, C. P., Gillette, P., & Pruett, K. D. (2019). Supporting father involvement: An intervention with community and child welfare–referred couples. *Family Relations*, *68*, 51–67. https://doi.org/10.1111/fare.12352

Puma, M., Bell, S., Cook, R., Heid, C., Broene, P., Jenkins, F., Mashburn, A., & Downer, J. (2012). *Third grade follow-up to the Head Start impact study final report, executive summary*. OPRE Report # 2012-45b.Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health & Human Services.

Randles, J. (2017). *Proposing prosperity? Marriage education policy and inequality in America*. Columbia University.

Reeves, R. V. (2014, February 13). *How to save marriage in America*. Brookings Institution. Retrieved from https://www.brookings.edu/articles/how-to-save-marriage-in-america/

Reichart, C. S. (2019). *Quasi-experimentation: A guide to design and analysis*. Guilford.

Roddy, M. K., Walsh, L. M., Rothman, K., Hatch, G. S., & Doss, B. D. (2020). Meta-analysis of couple therapy: Effects across outcomes, designs, timeframes, and other moderators. *Journal of Consulting and Clinical Psychology*, *7*, 583–596. https://doi.org/10.1037/ccp0000514

Rosenthal, R. (1983). *Meta-analytic procedures for social research*. Sage.

Sawhill, I. V. (2014). *Generation unbound: Drifting into sex and parenthood without marriage*. Brookings Institution.

Simpson, D. M., Leonhardt, N. D., & Hawkins, A. J. (2018). Learning about love: A meta-analytic study of individually oriented relationship education programs for adolescents and emerging adults. *Journal of Youth and Adolescence*, *47*, 477–489. https://doi.org/10.1007/s10964-017-0725-1

Spencer, C. M., & Anderson, J. R. (2021). Online relationship education programs improve individual and relationship functioning: A meta-analytic review. *Journal of Marital and Family Therapy*, https://doi.org/10.1111/jmft.12491

Stanley, S. M., Amato, P. R., Johnson, C. A., & Markman, H. J. (2006). Premarital education, marriage quality, and marital stability: Findings from a large, random household survey. *Journal of Family Psychology*, *20*, 117–126. https://doi.org/10.1037/0893-3200.20.1.117

Stanley, S., Rhoades, G., Loew, B., Allen, E., Carter, S., Osborne, L., Prentice, D., & Markman, H. (2014). A randomized controlled trial of relationship education in the U.S. Army: 2-year outcomes. *Family Relations*, *63*, 482–495. https://doi.org/10.1111/fare.12083

Trail, T. E., & Karney, B. R. (2012). What's (not) wrong with low-income marriages. *Journal of Marriage and Family*, *74*, 413–427. https://doi.org/10.1111/j.1741-3737.2012.00977.x

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Hawkins, A. J., Hokanson S., Loveridge E., Milius E., Duncan M., Booth M., & Pollard B. (2022). How effective are ACF-funded couple relationship education programs? A meta-analytic study. *Family Process*, 61, 970–985. https://doi.org/10.1111/famp.12739