



Findings from the Field

A Meta-Analytic Study of the Effectiveness of Healthy Marriage and Relationship Education Programs

Alan J. Hawkins and Kaylene J. Fellows

Disclaimer

Funding for this project was provided by the United States Department of Health and Human Services, Administration for Children and Families, Grant: 90-FH-0001. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the United States Department of Health and Human Services, Administration for Children and Families.

The first author can be reached for comments and questions at hawkinsa@byu.edu. Programs participating in this study were healthy marriage grantees funded by the U.S. Department of Health and Human Services, Administration for Children and Families, Office of Family Assistance. Participation in this study was voluntary.

This is a product of the National Healthy Marriage Resource Center, led by co-directors Mary Myrick, APR, and Jeanette Hercik, PhD, and project manager, Rich Batten, ThM, MEd, CFLE.

I. Executive Summary

Policymakers have become interested in exploring ways to help couples form and sustain healthy marriages and relationships as an effort to strengthen families. Most of these exploratory efforts have involved supporting voluntary marriage and relationship education (MRE) services, especially for lower income individuals and couples who are at greater risk for relationship difficulties. In 2006, the Office of Family Assistance (OFA), Administration for Children and Families, U.S. Department of Health and Human Services funded nearly 125 five-year demonstration programs in the form of grants for innovative efforts offering MRE services.

The healthy marriage grants funded by OFA were primarily provided to demonstrate what might be possible for effective programs to accomplish; therefore, most grantees collected basic field pre-post test outcome data on their program participants. These data were reported by grantees to OFA in semi-annual reports and were collected by the National Healthy Marriage Resource Center to inform this meta-analysis, which is a systemic review and aggregation of that data. For this meta-analysis, participating OFA marriage and relationship education programs voluntarily shared their evaluation data so that both the field and policymakers can gain a better understanding of what works. The research, summarized in this report, is designed to identify common findings across programs and provide an overview of what we know about the outcomes of these services.

The meta-analysis includes useable data from 31 OFA healthy marriage grantees operating 50 different programs. Some grantees offered multiple programs or services based on the different Allowable Activities. (Allowable Activities were defined by Congress in the Deficit Reduction Act of 2005 [DRA], which reauthorized Temporary Assistance for Needy Families [TANF] and provided federal funding for healthy marriage and responsible fatherhood grants. In the legislation, Congress prescribed eight Allowable Activities that these funds could be used for under the Healthy Marriage and Responsible Fatherhood Initiative, see Table 1.) In some cases one grantee may have offered distinct programs and tracked separate participants based on the Allowable Activity. For example, a grantee could serve unwed, expectant couples in a service distinct from what was offered to youth in high schools; therefore, two separate programs and evaluation efforts would come from one grantee.

The overall number of participants assessed in these 50 programs was $N = 49,888$. Almost all programs collected pretest and posttest data from program participants. A few grantees collected comparison data from a control group while only a handful of programs reported reliable follow-up data.

Across the programs, the immediate pre-posttest effect size was small to moderate, $d = .40$ ($p < .001$, $k = 46$). These figures suggest that at posttest, about 66% of program participants scored significantly above the pretest median score. A statistically significant, generally moderate effect size for each target population served or Allowable Activity (AA) was found.

Similarly, statistically significant, generally moderate effect sizes were found for all outcomes assessed, including relationship quality, communication skills, relationship confidence, relationship aggression,

unhealthy relationship knowledge, and co-parenting. There were no significant differences in effects for men and women. Moderate-dosage programs (9–20 hours) tended to have somewhat higher effects than lower dosage programs (8 hours). Programs having a larger proportion of participants with less than a high school education had stronger positive effects.

While this meta-analysis has important limitations, the results provide some early, encouraging evidence that OFA marriage and relationship education programs may be producing outcomes relevant to helping individuals and couples form and sustain healthy marriages and relationships. Moreover, policymakers who have followed these demonstration programs may have more reason to support them and to call for more rigorous efforts to assess their effectiveness.

II. Introduction



A generation of research has confirmed the value of healthy marriages and relationships to children and adults and the problems associated with unhealthy relationships and marital breakdown (Amato, 2005). The costs are more than personal, however; when couples are unable to form or sustain healthy marriages and relationships, society bears significant costs in the form of greater public assistance needs (Scafidi, 2008; Waite & Gallagher, 2000). Not surprisingly, then, state and federal policymakers have become interested in exploring ways to help couples form and sustain healthy marriages and relationships. Most of these exploratory efforts have involved supporting voluntary MRE services, especially for lower income individuals and couples who are at greater risk for relationship difficulties (Cherlin, 2009).

One of the major MRE efforts has been directed by OFA.¹ In 2006, OFA made grants available for innovative MRE services, and especially for programs targeted to lower income populations. Funding supported nearly 125 grantees who offered voluntary, preventative, educational services to different populations: youth, individuals, unmarried parents, engaged couples, and married couples across the United States. The programs offered were intended to help participants develop the skills and knowledge needed to form and sustain healthy marriages and relationships. Grantees provided services in one or more of eight Allowable Activity areas, presented in Table 1.

¹ The Office of Family Assistance is in the Administration for Children and Families (ACF), a part of the U.S. Department of Health & Human Services. The Deficit Reduction Act of 2005 (DRA) reauthorized the Temporary Assistance for Needy Families (TANF) program administered by ACF. The DRA reauthorization also included \$150 million to support programs designed to help couples form and sustain healthy marriages. On September 30, 2006, the Office of Family Assistance announced grant awards to 226 organizations to promote healthy marriage and responsible fatherhood as authorized by the DRA.

Table 1. Eight Allowable Activities, as Defined by Congress in the Deficit Reduction Act of 2005

Allowable Activity Number	Description
1	Public advertising campaigns on the value of marriage and the skills needed to increase marital stability and health
2	Education in high schools on the value of marriage, relationship skills, and budgeting
3	Marriage education, marriage skills, and relationship skills programs, that may include parenting skills, financial management, conflict resolution, and job and career advancement, for non-married pregnant women and non-married expectant fathers
4	Pre-marital education and marriage skills training for engaged couples and for couples or individuals interested in marriage
5	Marriage enhancement and marriage skills training programs for married couples
6	Divorce reduction programs that teach relationship skills
7	Marriage mentoring programs which use married couples as role models and mentors in at-risk communities
8	Programs to reduce the disincentives to marriage in means-tested aid programs, if offered in conjunction with any above activity

While the grants were provided primarily to demonstrate what might be possible for effective programs, all grantees collected program participation data. Many grantees also collected basic field pre-post outcome data (or pre-post data) on program participants.² These data have been reported by grantees to OFA in annual reports.

In this report, findings from a meta-analysis³ of the OFA MRE grantee programs are presented. The Administration for Children and Families (ACF) is also conducting large-scale program evaluation studies of a small number of healthy marriage and relationship education programs using rigorous randomized control trial designs.⁴ Other researchers have previously conducted valuable MRE evaluation studies.⁵ The current findings may supplement these other studies by providing some preliminary lessons about what has been accomplished by these programs and their potential for informing public policy.

2 Healthy marriage and relationship field practitioners in general, including these grantees, often did not have extensive experience with program evaluation. ACF worked with many of the grantees to provide technical assistance with evaluation tasks.

3 A meta-analysis is a study of studies on a particular topic. Outcome data is collected from different sources and standardized using systematic methods. The data is then combined into one large study (the meta-analysis) that yields an overall estimate.

4 See *Building Strong Families*, a program for unmarried expectant couples (<http://www.buildingstrongfamilies.org/>), and *Supporting Healthy Marriage*, a program for low-income married couples (http://www.acf.hhs.gov/programs/opre/strengthen/support_hlthymarr/index.html#resources).

5 For meta-analytic reviews of these programs, see Blanchard, Hawkins, Baldwin, & Fawcett (2009); Fawcett, Hawkins, Blanchard, & Carroll (2010); Hawkins, Blanchard, Baldwin, & Fawcett (2008); Hawkins & Fackrell (2010); Pinquart & Teubert (2010).

III. Method

Obtaining the Data

The OFA grantees were initially contacted in December 2010 through the National Healthy Marriage Resource Center and were asked to complete a brief, voluntary survey to see if their program outcome evaluation data might be appropriate for meta-analysis. About 50 grantees indicated that they had pre-post outcome data and would be willing to participate in the study. In the next few months, these grantees were contacted again, and asked for program evaluation data as well as other information about program and participant characteristics. Program evaluators were asked to fill out one survey for each target population or Allowable Activity that their program addressed. Thus, one grantee could submit data for multiple programs. For example, one grantee may have offered MRE for engaged couples, relationship education for high school youth, and targeted services for couples considering divorce. Grantees primarily collected data from program participants; very few collected data from nonparticipants in a control or comparison group.

The pre-post program data on participants was requested on the following general outcomes:

- Relationship quality
- Communication skills
- Relationship confidence
- Relationship aggression
- Knowledge of unhealthy relationships
- Co-parenting

Some of these outcomes were more pertinent to certain programs than others. For instance, knowledge of unhealthy relationships was a common outcome assessed in youth/high school relationship education programs while co-parenting outcomes were more commonly assessed in programs serving unwed parents (rather than youth or engaged couples).

Program evaluators were also asked to include any follow-up assessment data, if available. Some programs did not collect follow-up data, and programs that tried to do so often ended up with low rates of participation in the follow-up assessment. A variety of questions about the programs (e.g., curriculum used, program dosage), assessment measures, and participants (e.g., education level, level of relationship distress prior to the program) were also frequently asked. Some programs had collected outcome data, but their data could not be used in this meta-analysis study for the reasons outlined below.⁶

Reasons for Exclusion

Some evaluation data could not be included in this study for various reasons. Some programs, for example, provided data on participants' satisfaction with the programs, but did not assess outcomes such as reports of relationship quality or differences in communication skills. Outcomes such as participant satisfaction can be valuable for understanding service delivery, but are not the relationship outcomes pertinent to this particular meta-analysis.

⁶ In some circumstances, grantees did not fill out the requested survey but still provided their annual reports to OFA. These reports for usable data were scanned and on occasion, the outcome data across years from these reports were combined to compute appropriate statistics for the meta-analysis.

Some data were also in a form that was difficult to use in a meta-analysis. Some programs asked participants if they felt that they had improved in given areas because of the program, to which participants responded yes or no. Although valuable to program providers, these kinds of data are different from data where participants are asked before a program and then again after a program about specific feelings or behaviors.

A few programs also experienced a high level of attrition, due to participants dropping out of the program or not completing post-program outcome assessments; both types of attrition are common in field evaluations. Data was not included in this meta-analysis if more than 40% of participants who began the program and provided pretest data did not provide posttest data. This is because there is a significant risk that those who dropped out of the program, and therefore did not provide a posttest, could be systematically different from those who stayed in the program and did provide a posttest score. For example, the participants may have felt they were not benefitting from the program as much as those who stayed in. To limit this kind of potential bias, programs with high pre-post attrition rates were excluded.

Finally, some programs reported larger sample sizes at the posttest than at the pretest, likely because some program participants did not fill out a pretest (e.g., missed the first session) but did complete a posttest at the end of the program. This meant that the pretest and the posttest samples were not the same set of respondents, and thus the data were not a true indication of change for the participants. Therefore, programs that had more than 10% greater respondents at the posttest than at the pretest were excluded. In instances where a program had some outcomes with higher-than-acceptable attrition rates, but other outcomes within acceptable attrition levels, the program's outcomes with excessive attrition were not used but the program's outcomes within the acceptable attrition rates were used.

Issues That Did not Warrant Data Exclusion

There were also some instances in which data were incomplete in some way, but were still appropriate for use in this meta-analysis. The first instance of this was a lack of pre-post correlations. Computing precise effect sizes in meta-analysis requires the correlation (r) between pre- and post-test scores for each outcome, but some programs were not able to provide these correlations. To overcome this problem, the average pre-post r was computed for all of the studies that did report these correlations. The average value was then assigned to the studies that did not have a pre-post r ; this is a common practice in meta-analysis. This value was $r = .481$. To ensure that this procedure did not change the results of the study, an analysis compared studies with estimated pre-post correlations with those that calculated precise correlations, but there was not a statistically significant difference ($Q = 3.20, p = .07$).

Finally, a handful of programs collected outcome data in the form of retrospective pre-post data. This approach asks participants at the end of the program to assess their level on a given outcome before the program and then again after the program. Some researchers feel this is more accurate than an actual pre-test, since participants may be unrealistic about their skills prior to a program. Because this format still provides pretest and posttest data, it is easily adaptable to meta-analysis. Nevertheless, it is a distinctly different way of assessing change from the program than assessing participants before they begin the program, and then when they have completed the program. As a result, programs using retrospective pretests were analyzed separately to see if this type of measurement procedure produced results that were similar to or different from the more traditional pre-post measurement approach.

Effect Size

In a meta-analysis, the key statistic is the *effect size*. In this study, the effect size or d refers to the amount of change that occurred among program participants from pretest to posttest (or follow-up assessment).⁷ Basically, this figure measures the amount of change compared to the overall distribution of scores on an outcome. If $d = .50$, for example, then a group increased their scores on an outcome by one half of the standard deviation⁸ of all the scores. Researchers have found that an effect size of .50 is about average and is sometimes called a “moderate effect.” An effect size less than .30 is considered small; an effect size greater than .67 is considered large (Lipsey & Wilson, 2001).

IV. Results

Grantee Participation and Program Characteristics

Forty-four (44) of the grantees who responded to the initial survey indicated they had pre-post outcome data. Data from 31 of these grantees was used in this meta-analysis. (As explained above, all data that were submitted could not be included for various reasons.) This produced data on 50 different programs, as some grantees had data from more than one program or Allowable Activity. The overall number of participants assessed in these 50 programs was $N = 49,888$.

⁷ The effect size statistic employed is the standardized mean change score. All effect sizes were weighted by the inverse variance (squared standard error) and averaged to create the overall effect size. Random effects estimates, as opposed to fixed effects, were utilized. The random effects model allows for the possibility that differences in effect sizes from study to study are associated not only with participant-level sampling error but also with variations such as study and intervention methods (Lipsey & Wilson, 2001). In addition, the random effects model allows researchers to generalize beyond the studies included in the meta-analysis. Biostat's Comprehensive Meta Analysis II was used to perform these calculations.

⁸ A standard deviation is a measure of how scores on an outcome are distributed, whether they are very close together or very spread out. A large standard deviation means that scores are spread out; a small standard deviation means that scores are close together.

Table 2. Number of Participants and Programs by Program Target/Type of Participant

Program Target (# Programs)	Number of Participants
Youth (7)	23,488
Unwed Expectant Parents (8)	2,926
Engaged Individuals/Couples in Premarital Education (4)	5,602
Married Individuals/Couples (13)	11,643
Distressed Individuals/Couples in Divorce Reduction Program (1)	1,360
Premarital/Engaged and Married Couples in same program (10)	4,869
Total (43)	49,888

Note - Only 43 of the 50 programs broke down the data by target population.

Overall Program Impact



For the first analysis, all the outcomes that grantees reported in the survey were combined into one aggregate assessment of program impact. Program impact is an improvement in all the measured outcomes combined, or a positive change overall in program outcomes. The overall immediate pre-post effect size was small to moderate, $d = .40$ ($p < .001$, $k = 46$).⁹ Instead of conducting the posttest immediately following service delivery, some grantees followed up later (typically 1–2 months after program completion) with program participants to assess program impact.

The pre-follow-up effect size for eight programs was moderate, $d = .51$ ($p = .001$, $k = 8$). This indicates the programs that participated in the meta-analysis had statistically significant, moderate, positive effects on relationship outcomes measured. One interpretation of these change score effect sizes is that about 66% of program participants at posttest scored above the median¹⁰ of the pretest. (By definition, 50% of program participants scored above the median at pretest.) Similarly, about 70% of program participants at follow-up scored above the median of the pretest. Follow-up effect sizes should be interpreted cautiously, since the number of programs with follow-up data is quite small.

Many grantees reported outcomes separately by gender. This subset of data was analyzed to test whether men changed more, less, or the same as women, but a gender difference was not found ($Q^{11} = .56$, $p = .45$; for men, $d = .391$, $p < .001$, $k = 22$; for women, $d = .334$, $p < .001$, $k = 21$).

⁹ p refers to the statistical probability that a finding occurred by chance; a p value less than .05 suggests that the finding is very unlikely to be due to chance alone. k refers to the number of programs that were combined to produce the overall effect size.

¹⁰ The median is the score in the distribution that divides the scores into two equal halves, with 50% above that score and 50% below it. In a so-called normal orbell-shaped distribution, the median and the arithmetic average are the same.

¹¹ The Q statistic evaluates whether there is a significant difference between two groups, similar to the more commonly known t -test.

Retrospective Pre-Post Comparison

As was mentioned previously, a handful of grantees collected outcome data using a retrospective pretest, or by asking program participants at the end of a program to rate themselves or their relationship *before* they began the program and again now that they have *completed* the program. An analysis was run to determine whether programs that used retrospective pre-post assessments had different effect sizes from those of programs with traditional pre-post data, which is collected prior to and again after program participation. Retrospective pre-post reports produced larger effect sizes ($d = .64, p < .001, k = 6$) than did traditional pre-post reports ($d = .39, p < .001, k = 46$), although this difference was not statistically significant ($Q = 2.6, p = .07$).

Researchers are not in agreement about the merits and limitations of retrospective pre-post assessments of change. Some believe that they provide a better measure of self-reported change because people tend to overestimate their relationship skills before a program; that is, they learn going through the program that their skills are not as good as they thought. Part of the value of a program is to reveal to participants that they have areas for improvement. On the other hand, some researchers believe that program participants will overestimate change once they have completed the program, especially if they enjoyed the experience. This debate cannot be resolved in this study. However, it may be worth noting the possibility that the traditional pre-post assessments of change that were dominant in this study may have underestimated the extent of change in these programs. Further research is needed to rigorously test this possibility.¹²

Missing studies

Because not all OFA healthy marriage grantees' programs could be included in this study, there is a concern about whether the results of the study would be different if all programs were included. On analysis, however, little evidence was found that these "missing" studies would change the overall effect.¹³

Effects by Target Population Served (Allowable Activity)

Program effects were also explored separately by target population served (under the various Allowable Activities). The results by target population are shown in Table 3. Statistically significant effect sizes for each target population were found. The effects were generally moderate. The effect for programs serving premarital groups was smaller. This was expected; previous research has shown that it is hard to increase the already-high relationship satisfaction of engaged couples, although communication skills can be improved (Fawcett, Hawkins, Blanchard, & Carroll, 2010).

¹² To rigorously test this possibility, a study would need to include both traditional pre-post and retrospective pre-post measures. Ideally, it would also include a control group.

¹³ A funnel plot analysis showed a relatively symmetrical distribution. A trim and fill analysis indicated that zero studies were missing due to publication bias. Moreover, the fail-safe N for this study estimates it would take nearly 4,000 additional studies make the findings non-significant.

Table 3. Program Effects by Allowable Activity

Target Population Served (Allowable Activity)	Effect Size (<i>d</i>)	Number of Programs ¹⁴
Youth/HS students (AA 2)	.45**	7
Unwed expectant parents (AA 3)	.33***	8
Premarital couples (AA 4)	.25***	4
Premarital + Married couples ¹⁵ (AA 4, AA 5)	.39***	10
Married couples (AA 5)	.50***	13
Couples considering divorce (AA 6)	.50***	1

Note. ** $p < .01$. *** $p < .001$.

Effects by Relationship Outcome

The effects for specific relationship outcomes (for immediate-posttest effect and follow-up effects) were also calculated. The results are shown in Table 4. Significant, moderate, positive, pre-to-post change effects were found for participants' reports of relationship quality, communication skills, and knowledge of unhealthy relationship characteristics (the latter outcome was reported most in youth/high school programs). Small but statistically significant pre-to-post change effects were also found for relationship confidence (often asked of unmarried couples), relationship aggression, and co-parenting outcomes (often asked of unwed parents). The follow-up effects for some measures were a little stronger than the immediate posttest effects, but the number of programs reporting reliable follow-up effects is small, so caution is recommended in interpreting this result.

Table 4. Program Effects by Relationship Outcomes

Relationship Outcome	Timing	Effect Size (<i>d</i>)	Significance (<i>p</i>)	Number of Programs (<i>k</i>)
Relationship quality	Pre-post	.40	***	34
Relationship quality	Pre-follow-up	.51		8
Communications skills	Pre-post	.41	***	44
Communication skills	Pre-follow-up	.64	***	8
Relationship confidence	Pre-post	.21	***	25
Relationship confidence	Pre-follow-up	—		—
Relationship aggression	Pre-post	.26	***	19
Relationship aggression	Pre-follow-up	—		—
Unhealthy relationship knowledge	Pre-post	.44	***	15
Unhealthy relationship knowledge	Pre-follow-up	—		—
Co-parenting	Pre-post	.18	***	13
Co-parenting	Pre-follow-up	—		—

Note. *** = $p < .001$

¹⁴ This only includes the programs contributing to the overall effect size.

¹⁵ Some grantees combined these two areas together into one program, so these were included as a separate category.

Factors Related to Stronger or Weaker Effects

Three factors that may have strengthened or weakened program effects were explored. These were program dosage, participant education level, and level of participants' relationship distress.

Program dosage

The programs were divided into three categories based on the number of instructional hours provided (dosage). Program dosages ranged from 8 to 30 hours (of instructional contact) with an average dosage of 12.3 hours. Programs that lasted eight hours (the minimum required by OFA) were categorized as low dosage; 9–20 hour programs were medium dosage; and programs that lasted longer than 20 hours were high dosage.

The data was analyzed to see if higher dosage programs produced stronger effects. Medium-dosage programs ($N = 27$) had the strongest effects ($d = .42, p < .001, k = 27$), followed by low-dosage programs ($N = 12, d = .35, p < .001, k = 10$), with the small number ($N = 4$) of high-dosage programs having relatively small, but still significant effects ($d = .22, p < .001, k = 4$). The differences between dosage levels, however, were not statistically significant ($Q = 4.82, p = .09$), and there are a very small number of high dosage programs. Thus, dosage does not appear to be much of a factor in program impact, although medium-dosage may be optimal.

Participant Education Level

The programs were divided into two groups based on education; about half of programs reported on this demographic. In some programs, less than 80% of the sample had graduated from high school (considered low education). In contrast, in other programs more than 80% of the sample had graduated from high school (considered the high education groups). In the average program with education demographic data on participants, 21% of participants had not completed high school. (The youth programs were not included in this analysis, as none of the participants had graduated from high school.)

Results show the overall effect size differed based on the education level of the participants ($Q = 155.3, p < .001$). Programs serving less educated participants ($d = .45, p < .001, k = 16$) produced stronger effects than programs serving more educated participants ($d = .19, p < .001, k = 8$).

Participant Relationship Distress

Program impact was also examined relative to the level of self-reported relationship distress by participants prior to the start of the programs. (Twenty-four programs gathered data on relationship distress among participants.) A program was considered to have higher levels of participants in relationship distress when 40% or more of the participants were distressed and to have lower levels of relationships distress when less than 40% of the participants were distressed. Interestingly, 42% of programs that reported participant distress levels had 40% or more distressed couples. On average, 19% of couples in the programs were distressed, in programs that were able to report this information. (Programs that targeted youth were not included in this analysis.) There was no significant difference

in effects between these groups ($Q = .60, p = .44$; for low distress, $d = .32, p = .001, k = 13$; for high distress, $d = .39, p = .001, k = 7$). Thus, the proportion of distressed couples in a program may not have a significant impact on program effects.

V. Conclusion



Overall, for these 31 OFA Healthy Marriage grantees conducting 50 programs, serving nearly 50,000 participants, there were moderate positive program effects. These effects were obtained for programs across the Allowable Activity areas and for target populations (individual youth; and unwed expectant, premarital, marital, and couples participating in divorce reduction services). In addition, positive program outcome effects were obtained for reported improvements in relationship quality, communication skills, relationship confidence, relationship aggression, healthy relationship knowledge, and co-parenting. There was no evidence of gender differences in outcomes, for the subsample of programs reporting data by gender. Moderate-dosage programs tended to have somewhat higher effects than lower dosage programs. Programs with larger proportions of participants who did not have a high school education had stronger positive effects.¹⁶

Readers should note, however, that this study has important limitations that restrain its conclusions. First, only about one third of 125 grantees participated in the voluntary study. Even though supplementary analyses revealed no evidence that the missing programs would have nullified or biased these results, it would have been better if a larger proportion of grantees had been able to participate in the study.

Second, nearly all the programs in this study were evaluated without a comparison or a control group. In other meta-analytic studies of MRE programs (Blanchard, Hawkins, Baldwin, & Fawcett, 2009; Fawcett et al., 2010; Hawkins, Stanley, Blanchard, & Albright, in press), sometimes one-group/pre-post studies appear to overestimate the effects of programs when compared to control group studies, but at other times their results may look very similar to control-group studies. Thus, without control or comparison groups, we cannot rule out alternative explanations for the changes in outcomes other than participation in the program.

One way to view the results of this meta-analysis is that the effects reported here may be an upper limit of the true effects. That is, stronger effects might not be seen if the studies had included a control group of no-treatment participants.

¹⁶ Kerpelman et al. (2010) also found that for their youth relationship education program, results were stronger for groups with lower socioeconomic status.

Third, almost all of the programs were evaluated with conventional pre-post assessments. A handful of OFA healthy marriage grantees evaluated their programs with retrospective pre-post designs, and this meta-analysis shows these programs produced stronger effects compared to the conventional pre-post design studies. As mentioned earlier, some believe that retrospective designs are more sensitive to change because people tend to overestimate their skills before learning more about them. Nevertheless, studies with more sophisticated designs are needed before conclusions can be drawn about the best way to measure participant change from these programs.

Fourth, this study relied primarily on posttest effects from self-reports taken right at the end of the program or shortly thereafter; longer-term follow-up assessments were few in number and many of these were unusable because only a small proportion of the original program participants completed a follow-up survey. There are several issues to note here. First, program effects can fade, so longer-term follow-up assessments are a more rigorous evaluation of the effectiveness of programs' abilities to create meaningful change. In other meta-analytic work the authors have done evaluating MRE programs, there wasn't much evidence of program effects fading, but very few studies follow participants for more than six months (Blanchard et al., 2009; Hawkins, Blanchard, Baldwin, & Fawcett, 2008). On the other hand, sometimes short-term assessments can underestimate program effects because it takes a while for new skills to be implemented and then shift behaviors, thoughts, and feelings about the relationship as a result. More long-term follow-up studies are needed to boost confidence in the true effects of these programs.

Finally, all the outcomes reported in this study come from self-report measures. While this is an efficient way to learn about relationship outcomes, some research suggests that direct observations (of communication skills, for example) by trained researchers could yield stronger effects than self-reports by participants (Blanchard et al., 2009; Fawcett et al., 2010). Thus, it is possible that the effects would be somewhat stronger if evaluators could directly observe participants' communication or problem-solving behaviors.

With these limitations in mind, these results should be interpreted with caution and with confirming data. The Administration for Children and Families is conducting other, more rigorous studies of MRE programs. For instance, the *Building Strong Families* study and the *Supporting Healthy Marriage* study are following thousands of low-income, unmarried and married couples for several years across the United States. These participants volunteered and were randomly assigned to either a substantial MRE and couple-support program or a no-treatment control group. How these results may differ from these preliminary data remains to be seen.

There are always many "what-if" scenarios and caveats to research studies. Nevertheless, the results of this meta-analysis are an early positive sign that MRE works. Even with the study limitations, the results provide some early, encouraging evidence that OFA healthy marriage and relationship grantee programs

are finding success in helping individuals and couples form and sustain healthy relationships and marriages. This should be satisfying to those grantees who have worked hard over the past five years to implement these programs. Moreover, policymakers who have followed these demonstration programs have reason to support them and to call for more rigorous efforts to assess their effectiveness.

References

- Amato, P. R. (2005). The impact of family formation change on the cognitive, social and emotional well-being of the next generation. *Future of Children, 15*, 75–96.
- 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.
- Cherlin, A. J. (2009). *The marriage-go-round*. New York: Knopf.
- 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.
- Hawkins, A. J., & Fackrell, T. A. (2010). Does relationship and marriage education for lower-income couples work? A meta-analytic study of emerging research. *Journal of Couple & Relationship Therapy, 9*, 181–191.
- 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.
- Hawkins, A. J., Stanley, S. M., Blanchard, V. L., & Albright, M. (in press). Exploring programmatic moderators of the effectiveness of marriage and relationship education: A meta-analytic study. *Behavior Therapy*.
- Kerpelman, J. L., Pittman, J. F., Adler-Baeder, F., Stringer, K. J., Erygit, S., Cadely, H. S., & Harrell-Levy, M. K. (2010). What adolescents bring to and learn from relationship education classes: Does social address matter? *Journal of Couple and Relationship Therapy, 9*, 95–112.
- Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks, CA: Sage.
- Pinquart, M., & Teubert, D. (2010). A meta-analytic study of couple interventions during the transition to parenthood. *Family Relations, 59*, 221–231.
- Scafidi, B. (2008). *The taxpayer costs of divorce and unwed childbearing: First-ever estimates for the nation and all fifty states*. New York: Institute for American Values.

Waite, L. J., & Gallagher, M. (2000). *The case for marriage*. New York: Doubleday.

Wood, R. G., McConnell, S., Moore, Q., Clarkwest, A. & Hsueh, A. (2010). *The Building Strong Families project: Strengthening unmarried parents' relationships: The early impacts of Building Strong Families*. Princeton, NJ: Mathematica Policy Research.