IS MARRIAGE AND RELATIONSHIP EDUCATION EFFECTIVE? A COMPREHENSIVE META-ANALYSIS

Alan J. Hawkins, Victoria L. Blanchard, Elizabeth B. Fawcett, & Jaelynn Jenkins
School of Family Life, Brigham Young University
Poster presented at the Society for Prevention Research Annual Conference, May 30, 2007, Washington D.C.

ABSTRACT

Marriage and relationship education (MRE) recently has become a tool of public policy to help couples form and sustain healthy marriages. With increased public funds comes increased public scrutiny. In this meta-analytic study, we examined the effects of MRE on two common outcomes: relationship satisfaction/quality and relationship communication. A thorough search produced 133 codable reports that yielded 197 studies and nearly 600 effect sizes. We report effect sizes separately by study design groups. The effect sizes for relationship satisfaction/ quality were generally in the d = .30-.40 range, while the communication effect sizes were somewhat larger, generally in the d = .40-.70 range. We also employed a large number of methodological, sample, and intervention moderators to examine the substantial heterogeneity of these effect size distributions. Moderator analyses yielded some interesting findings relevant to the work of MRE practitioners. While MRE evaluation studies now provide ample evidence of their effectiveness, there are important deficits in this work especially for guiding public policy. We discuss several important deficits and make suggestions for further research.

BACKGROUND

Over the last decade, marriage and relationship education (MRE) has gone beyond programs offered by private professional and lay practitioners to become a tool of public policy initiatives. U.S. federal policy makers have promoted MRE as a way to help couples—especially more disadvantaged couples—form and sustain healthy marriages as an additional weapon to fight poverty. In 2006, federal legislation allocated $100 million a year for five years to support promising MRE programs and initiatives targeted primarily to at-risk couples. In addition, a handful of states (e.g., AL, FL, GA, MI, OH, OK, NC, NM, NY, TX, UT) also are allocating significant public funds to support MRE efforts in their states. With greater public support for MRE comes greater public scrutiny.

Scholars have conducted many evaluation studies of various MRE programs since the 1970s (Hawkins et al., 2004). But a comprehensive evaluation of the effectiveness of MRE is the province of meta-analysis. This study reports a meta-analysis of MRE evaluation research. Scholars have produced a handful of meta-analytic studies of MRE in the past (Butler & Wampler, 1999; Carroll & Doherty, 2003; Giblin, Sprinkle, & Sheehan, 1985; Hight, 2000; Reardon-Anderson et al., 2005). However, these studies each have significant limitations. This study is the most comprehensive and sophisticated meta-analysis of MRE programs to date.

Our purpose in this study is to address what is now an important public policy question: *Is there reliable evidence that marriage and relationship education can help couples form and sustain healthy marriages?* We break down this generic question into a set of more specific research questions and sub-questions.

RESEARCH QUESTIONS

Q1: Relationship Satisfaction/Quality. What is the overall effect size of MRE on relationship satisfaction/quality at immediate post-assessment? And what is the overall effect size of MRE at later follow-up assessments? Is there significant change in the overall effect size from post-assessment to follow-up assessment, and does timing of the follow-up assessment moderate the effect size?

Q2: Relationship Communication. What is the overall effect size of MRE on relationship communication at immediate post-assessment? And what is the overall effect size at later follow-up assessments? Is there significant change in the overall effect size from post-assessment to follow-up assessment, and does timing of the follow-up assessment moderate the effect size?

Q3: Moderator Effects. What methodological, sample, and intervention moderator variables help explain the heterogeneity of effect size distributions for relationship quality/satisfaction and communication?

METHODS

Search Procedure

Through various means, a thorough search for relevant published and unpublished reports yielded:

133 codable reports → 197 coded studies → nearly 600 effect sizes

Selection and Inclusion Criteria

In contrast to previous meta-analytic studies of MRE, our general strategy was to include all relevant evaluation research but code for and analyze important methodological features that could bias effect size estimates.

- *Psychoeducational intervention*. We included studies that assessed the effects of a psychoeducational intervention that included improving marital or couple relationships or communication skills as a goal. Therapeutic interventions were excluded in order to provide a clear picture of the effects of psychoeducational vs. clinical intervention.
- *Outcomes.* We included studies that measured reports of relationship satisfaction or quality and/or some kind of relationship communication, the most common outcome reported in the literature. Data had to reported in a form that could generate an effect size. Extensive efforts were made to "rehabilitate" uncodable studies, with modest success.
- *Timeframe.* We included studies from 1975, when serious research on MRE picked up momentum, through 2006, when significant federal funding for MRE was first targeted.
- *Methodological design*. We considered evaluation studies regardless of methodological design. Because real-world intervention places constraints on the ability to adhere to the strict requirements of experimental design, quasi-experimental and one-group/pre-post research designs were included in analyses. However, we report the effect sizes for these research designs separately.
- *Publication status*. Both published and unpublished studies were sought for inclusion in this meta-analysis so that the issue of publication bias could be directly addressed. Publication bias is a serious threat to the validity of meta-analytic results. Almost half (k=66) of the 133 reports analyzed in this study were unpublished reports, mostly doctoral dissertations or master's theses.
- *Foreign language studies*. While we did not conduct an exhaustive search for studies published in languages other than English, our search surfaced a handful of reports published in other languages (i.e., German, Dutch, Afrikaans). When this occurred, we employed translators to help us code in order to include these studies in our meta-analysis.

Computation and Reporting of Effect Sizes

- *Effect size.* We employed the s*tandardized mean gain score difference* (for experimental and quasi-experimental designs) or s*tandardized mean gain* (for one-group/pre-post designs).
- *Weighting.* Effect sizes of each study were weighted by the inverse variance (squared standard error).
- *Study level.* We calculated effect sizes aggregated to the study level because many studies included multiple outcomes related to communication. Similarly, studies reported data in different units of analysis, including separately for spouses/partners, couple scores, or individual scores (undifferentiated by gender). As appropriate, we aggregated these reports to the study level (except when we analyzed reports by gender).
- *Calculations.* We used Biostat's Comprehensive Meta Analysis II for calculations.
- *Success rate.* We also report the Binomial Effect Size Display (BESD) success rate difference (Rosenthal & Rubin, 1982).
- *Random effects estimates.* We employed random effects estimates. The fixed effects model assumes that random error results only from subject-level sampling error in the individual studies, whereas the random effects model allows for the possibility that differences in effect sizes from study to study are associated not only with subject-level sampling error but also with variations such as study and intervention methods (Lipsey & Wilson, 2001).
- *Effect size pattern.* We faced the challenge of interpreting a set of effect size estimates rather than a single estimate. That is, we generated a set of 6 effect sizes for each outcome (3 [design group] x 2 [time points]) rather than a single effect size. Thus, we attempt to interpret the *pattern* of effect sizes. Fortunately, this potential complexity was reduced by the fact that the three design groups and the two time points usually pointed in similar directions and seldom produced significant differences.

SUMMARY OF FINDINGS

General Outcomes (see Table 1)

- *Q1: Relationship Satisfaction/Quality.* MRE program effects on marital satisfaction/quality were modest but significant—generally ranging from .30-.40. Moreover, effects were maintained over time.
- *Q2: Undifferentiated Relationship Communication.* MRE program effects on undifferentiated relationship communication were modest but significant—generally in the .40-70 range. Moreover, communication effects were maintained over time.
- *Study design differences.* Differences in effect sizes due to study designs were non-significant for follow-up assessments for both satisfaction/quality and communication; but at immediate post-assessment, effect sizes for one-group/pre-post studies appeared inflated, perhaps due to a concentration of high-intensity programs in this group.

Moderators (see Table 2)

- *Publication bias.* We found no evidence of publication bias.
- *Racial/ethnic, economic diversity.* There was limited ability to test for differences because of small cell sizes, but we found no evidence of racial/ethnic or economic-group differences.
- *Relationship distress.* There was limited ability to test for differences because of small cell sizes, but we found no evidence of differences due to sample couple distress.
- *Gender.* We found no evidence that MRE favors women or men.
- *Relationship length.* We found modest evidence that samples with couples married on average 6-10 years had stronger effects sizes than couples together less than 6 years.
- *Premarital vs. enhancement.* We found evidence that enhancement programs produced stronger effects for relationship satisfaction/quality than premarital programs (a possible ceiling effect). But this same pattern did not hold for communication.
- *Program content.* We found clear evidence that programs that emphasized communication-skills training produced larger effect sizes for both outcomes than programs that emphasized relationship expectations and knowledge.
- *Program intensity.* We found clear evidence that moderate-dosage programs (9-20 hours) produced stronger effects for both outcomes compared to low-dosage programs (<9 hours).
- *Program setting.* We found no evidence that programs delivered in university/clinical settings produced stronger effects sizes than programs delivered in religious settings.

CRITIQUES AND IMPLICATIONS

Critiques

MRE researchers' focus has been too narrow. Marriages and intimate relationships are complex systems that demand understanding from multiple perspectives. Accordingly, we recommend that future MRE researchers include measures of relationship stability, violence, and virtues. This implies that programs that do not address these issues should consider doing so.

- *Marital stability.* Only a handful of studies have examined indicators of marital stability or propensity to divorce. For policy purposes, this is a critical outcome because the stability of the relationship has important consequences beyond its quality (Amato et al., 2007).
- *Relationship aggression.* Similarly, even fewer MRE studies include measures of relationship aggression or violence. This is a crucial outcome relevant to the quality of the couple relationship and the well-being of children in that relationship, and it is an important concern for policy makers supporting MRE with public funds (Roberts, 2006).
- *Virtues.* MRE studies of important relationship virtues, such as commitment, sacrifice, and forgiveness are limited. Communication skills are important, but if relationship virtues can be improved with interventions then MRE practitioners should give more attention to relationship virtues.

Implications

Policy makers have been exploring ways to help couples form and sustain healthy marriages, and MRE recently has become the primary tool of social policy to further this goal. With public funding for MRE comes greater public scrutiny. Our comprehensive meta-analytic study suggest several implications for MRE researchers and practitioners:

- *Overall.* MRE produces modest but reliable effects generally in the range of other psychoeducational interventions of interest to policy makers (e.g., pregnancy prevention). These effects are maintained for

short-term follow-up assessments. MRE appears to work for both women and men. While these findings provide policy makers greater confidence, still . . .

- *Limitations.* The current body of MRE evaluation research still falls short of fully answering important questions policy makers have. Perhaps most important is the question of effects for diverse samples; programs for racially and economically diverse groups, as well as programs targeting the unique needs of remarried couples, are limited, and these are important groups that policy makers are targeting. While we found no early evidence to suggest that MRE is ineffective for more diverse samples, further work is needed. And fortunately, much of that work is underway. And we need to know better how MRE impacts relationship stability and aggression.
- *Study design choices.* Not all evaluation studies need to be rigorous, randomized control trials. Non-experimental designs appear to produce similar effect sizes as experimental designs, at least for follow-up assessments.
- *Communication emphasis.* A focus on improving communication skills seems justified. Stronger effects are apparent for programs that emphasize communication skills over relationship expectations/knowledge. Premarital programs should target communication skills. Relationship satisfaction is at its peak and relatively unaffected by premarital programs, while improvement in communication skills is possible.
- *Moderate dosage.* Moderate-intensity programs (9-20 hours) appear to be the right dosage. Lower-dosage programs yield smaller effect sizes; high-intensity programs do not appear to add further benefit.
  Settings. Programs in religious settings are as effective as those in university/clinic settings.

REFERENCES

Amato, P. R., Booth, A., Johnson, D. R., Rogers, S. J. (2007). *Alone together: How marriage in American is changing.* Cambridge, MA: Harvard University.

Butler, M. H., & Wampler, K. S. (1999). A meta analytic update of research on the Couple Communication program. *The American Journal of Family Therapy, 27*, 223-237.

Carroll, J. S., & Doherty, W. H. (2003). Evaluating the effectiveness of premarital prevention programs: A meta-analytic review of outcome research. *Family Relations, 53,* 105-118.

Giblin, P., Sprenkle, D. H., & Sheehan, R. (1985). Enrichment outcome research: A meta-analysis of premarital, marital and family interventions. *Journal of Marital and Family Therapy, 11*(3), 257-271.

Hight, T. L. (2000). Do the rich get richer? A meta-analysis of the methodological and substantive moderators of couple enrichment, Virginia Commonwealth University, 2000). *Dissertation Abstracts International.*

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis.* Thousand Oaks, CA: Sage.

Reardon-Anderson, J., Stagner, M., Macomber, J. E., Murray, J. (2005). *Systematic review of the impact of marriage and relationship programs.* Washington D.C.: Administration for Children and Families, U.S. Department of Health & Human Services.

Roberts, P. (2006, September). Building bridges between the healthy marriage, responsible fatherhood, and domestic violence movements: Issues, concerns, and recommendations. *CLASP Policy Brief,* Brief No. 7, Washington D.C.: Center for Law and Social Policy.

Rosenthal, R., & Rubin, D. B. (1982). A simple, general purpose display of magnitude of experimental effect. *Journal of Educational Psychology, 74,* 166-169.

Table 1. Effect Size Estimates of Marriage and Relationship Education on Marital Satisfaction/Quality and Undifferentiated Communication, by Study Design.

Outcome/Time	Total *k*	A. Experimental				B. Quasi-Experimental				C. 1-Group Pre-Post				Design Difference	
		k	*d*	*p*	BESD diff.	*k*	*d*	*p*	BESD diff.	*k*	*d*	*p*	BESD diff.	*Q* (df=2)	*p*
Satisfaction/ Quality															
1. Pre-Post	144	52	.370	***	19%	51	.355	***	18%	41	.655	**	31%	6.48	*
2. Pre-Follow-up	85	23	.326	***	16%	35	.364	***	18%	27	.317	***	16%	0.21	ns
Pre-Post-Follow-up Studies only:															
3. Pre-Post	68	18	.247	*	12%	29	.437	***	22%	21	.331	***	17%	1.85	ns
4. Pre-Follow-up	69	18	.315	**	16%	30	.355	***	18%	21	.319	***	16%	0.13	ns
5. Post-Follow-up	57	14	-.072	ns	--	26	-.028	ns	--	17	-.041	ns	--	0.25	ns
Undifferentiated Communication															
6. Pre-Post	140	41	.532	***	25%	48	.380	***	19%	51	.731	***	39%	6.90	*
7. Pre-Follow-up	77	18	.451	***	22%	30	.390	***	19%	29	.556	***	26%	1.93	ns
Pre-Post-Follow-up Studies only:															
8. Pre-Post	83	20	.789	***	40%	34	.443	***	22%	29	.572	***	27%	6.17	ns
9. Pre-Follow-up	70	16	.468	***	23%	28	.410	***	20%	26	.593	***	28%	2.24	ns
10. Post-Follow-up	55	13	-.137	ns	--	21	-.131	ns	--	21	-.077	ns	--	0.38	ns

Table 2. Summary of Moderator Analyses

Moderator Variable	Weak (w), Moderate (m), or Strong (s) Test of Differences	Pattern Suggests No Significant Differences (√)	Pattern Suggests Some Significant Differences (√)	Comments
A. Publication source	s	√		No evidence of publication bias.
B. Racial/ethnic diversity	w	√		Racial/ethnic diversity in body of work too limited for fair test of differences.
C. Economic diversity	w	√		Economic diversity in body of work too limited for fair test of differences.
D. Couple distress	w	√		Relationship distress in body of work too limited for fair test of differences.
E. Gender	m	√		No evidence of pattern of gender differences, but non-independent pairs.
F. Relationship length	m		√	Couples married 6-10 years may benefit most compared to couples married shorter or longer periods.
G. Program type	m		√	Premarital programs produce smaller effects for relationship satisfaction but not communication; possible ceiling effect for premarital couples.
H. Program content	m		√	Communication programs produce larger effects.
I. Program intensity	m		√	Programs 9-20 hours produce larger effects than less intense programs, but longer programs do not produce even stronger effects.
J. Program setting	m	√		Little difference between programs in religious and university/clinic settings.
K. Program delivery	m		√	Weak evidence that specialized delivery produces larger than integrated delivery.