
# Impact Evaluation of ELEVATE and Couples Connecting Mindfully in Alabama: Final Impact Evaluation Report for Auburn University

December 17, 2020

Prepared by

Julianne McGill, Auburn University

Francesca Adler-Baeder, Auburn University

Kim D. Gregson, Auburn University

Recommended Citation:

McGill, J., Adler-Baeder, F., and Gregson, K. (2020). Impact Evaluation of ELEVATE and Couples Connecting Mindfully in Alabama.

Acknowledgements:

A special thank you to the staff and students of the Alabama Healthy Marriage and Relationship Education Initiative, as well as our state partners. Your enthusiasm, work-ethic, and follow-through have made this impact evaluation of the ELEVATE and Couples Connecting Mindfully programs a success. Most importantly, our heartfelt appreciation goes out to the couples who participated in the programs and the study. Your commitment to informing the research on relationship education programs in the community is commendable and the time you invested in this process is valued. We are grateful to Derekh Cornwell with Mathematica for your wonderful technical assistance support throughout the project. And we are enormously grateful to our Federal Program Specialists, Rosalind Ginyard (retired) and Tanya Howell, along with Robin McDonald, OFA Division Director, for your individual and collective leadership, commitment, and support of our project and research every step of the way.

Disclosure:

The evaluation team and portions of the program team are both affiliated with Auburn University; however, the teams have separate leadership and oversight.

This publication was prepared under Grant Number 90FM0082 from the Office of Family Assistance (OFA) within the Administration for Children and Families (ACF), U. S. Department of Health and Human Services (HHS). The views expressed in this report are those of the authors and do not necessarily represent the views or policies of HHS, ACF, or OFA.

**Structured Abstract:** Impact Evaluation of ELEVATE and Couples Connecting Mindfully in Alabama

**Objective.** The Alabama Healthy Marriage and Relationship Education Initiative (AHMREI), a partnership between Auburn University and 9 family resource centers, used a longitudinal randomized control trial (RCT) to implement and evaluate two relationship education curricula, ELEVATE and Couples Connecting Mindfully (CCM), among a diverse population of adult couples.

**Study design.** Couples (adults age 19 and older; $N = 930$ couples) were enrolled into the RCT in five separate cohorts over two and a half years. Couples completed baseline surveys and were randomly assigned to either one of the two program groups or to the control group by implementation site. Program participants were offered 8-9 hours of the ELEVATE or CCM curricula focused on healthy relationship skills; control respondents were provided only with a list of community resources. In order to address primary research questions, follow-up surveys were collected at two months and six months after the baseline survey; we obtained one-year post-baseline follow-up surveys to address additional research questions.

**Results.** Our primary findings indicated significant immediate (2-month) program impacts of both ELEVATE and CCM on couple relationship skills compared to the control group. Additionally, at six months post-baseline (and up to one year for additional analyses), the ELEVATE group reported significant gains in mental health and couple satisfaction, while the control group experienced no change in mental health and declines in couple satisfaction. There were no significant differences in growth over six months for CCM participants versus control respondents in mental health or couple satisfaction; however, in additional analyses, program impacts of CCM on couple satisfaction emerged at one-year follow-up. Further, for both ELEVATE and CCM participants, the immediate improvements in couple relationship skills predicted later enhancements in couple satisfaction.

**Conclusion.** The AHMREI utilized effective recruitment, engagement, and retention strategies, and the AHMREI facilitators implemented both ELEVATE and CCM with high fidelity and were viewed by participants as knowledgeable, effective, and engaged. These favorable implementation factors are likely associated with the robustness and longevity of program impacts on couple functioning (for both ELEVATE and CCM) and individual mental health (for ELEVATE) among an economically and educationally diverse population in Alabama.

# Contents

# Tables

# Figures

# Impact Evaluation of ELEVATE and Couples Connecting Mindfully in Alabama

## I. INTRODUCTION

### A. Introduction and study overview

#### 1. Motivation for Healthy Marriage and Relationship Education (HMRE) in Alabama

Alabama has a persistent history of high levels of marital and family instability and individuals and families encounter barriers to economic self-sufficiency. Accordingly, Alabama's children face tremendous risks to their healthy development and well-being. In the 2014 state rankings on measures of child well-being, Alabama was rated 44 out of the 50 states, indicating some of the poorest conditions and prospects for children of any state (Kids Count, 2014). Further, 41% of children living in Alabama resided in single-parent families, compared to 35% nationally, and 1 in 3 children lived with caregivers who were not regularly employed (Kids Count, 2014); unemployment rates in Alabama were in the top 10 nationally (Bureau of Labor Statistics, 2014). Just under a quarter (24%) of all families with children in Alabama were living below the poverty line, including 10% of married couples with children and 46% of single-parent families (Bureau of Labor Statistics, 2014).

Responding to these specific needs in our state, the Administration for Children and Families (ACF) provided funding from 2015-2020 for the Alabama Healthy Marriage and Relationship Education Initiative (AHMREI), a large-scale partnership among Auburn University and 9 additional implementation partners at family resource centers and community agencies across the state of Alabama. Comprehensive marriage/relationship education programs for economically and educationally disadvantaged couples have been consistently linked with couple- and family-level benefits (Arnold and Beelmann, 2019; Hawkins and Erickson, 2015). Specifically, evidence from our past ACF-funded projects offering HMRE programming to diverse populations of couples across the state of Alabama demonstrates HMRE-related improvements in individual functioning (i.e., depression, anxiety, and individual empowerment; Adler-Baeder et al., 2010; Bradford et al., 2014; McGill et al., 2016) and couple/marriage functioning (i.e., relationship quality, stability, and commitment; Adler-Baeder et al., 2010; 2014; Bradford et al., 2017; Rauer et al., 2014). In addition, the AHMREI found some evidence in previous grant cycles for family-level benefits and ripple effects of community-based HMRE. Among program participants, it was found that positive changes in couple functioning were related to positive changes in coparenting agreement (Adler-Baeder et al., 2013), and that improvements in coparenting agreement were associated with more positive parenting and predicted improvements in young children's social competence over a one-year period (Adler-Baeder et al., 2016). More vulnerable subgroups, including relationally unstable or distressed participants

(Bradford et al., 2017; McGill et al., 2016), incarcerated adults (Harcourt et al., 2017), and stepparents (Garneau et al., 2015) also benefit from HMRE, as our ACF-funded projects using one-sample or quasi-experimental designs have shown.

With this wide base of suggestive evidence regarding program benefits at multiple levels of functioning (i.e., individual, couple, family) among more vulnerable and racially/economically/educationally diverse couples, the AHMREI refined the development of two curricula and prepared to extend the outreach across sites in Alabama in an effort to implement a large-scale impact evaluation. The two curricula, ELEVATE and Couples Connecting Mindfully (CCM), were specifically designed to address the diverse needs of a broad range of couples in an effort to promote healthy marriages and relationships in Alabama.

## 2. Programs studied

ELEVATE: Taking Your Relationship to the Next Level (Futris et al., 2015) was a featured curriculum of the 2015-2020 AHMREI project and was developed explicitly based on the evidence-informed National Extension Relationship and Marriage Education Model (NERMEM; http://www.fcs.uga.edu/nermen/nermem). ELEVATE is an 8-module, scripted HMRE curriculum that can be delivered in 8-12 hours. It utilizes brief informational sessions, activities, skills practices, and videos to facilitate knowledge and skills for developing healthy relationships and marriages. Along with an introductory module, there is one module focused on each of the seven core skills associated with healthy relationships, including the following: caring for self and managing stress, choosing committed behaviors, developing and maintaining intimate knowledge of your partner, sharing interests and activities, demonstrating love/affection, communicating effectively and managing conflict/finances, and connecting to community and other supports (Futris and Adler-Baeder, 2013). A quasi-experimental design study demonstrated the first validation of ELEVATE programming among a diverse, community-based sample (McGill, Adler-Baeder, and Garneau, in press). Specifically, we found sustained 6-month program impacts of ELEVATE on several relationship skills (i.e., intimate knowledge of partner, social connections, and conflict management subscales of the Couple Relationship Skill Inventory [Adler-Baeder et al., 2019]), as well as on couple relationship quality (measured with the Quality of Marriage Index, Norton, 1983).

The second HMRE curriculum used in the 2015–2020 AHMREI project, Couples Connecting Mindfully (CCM; McGill, Ketring and Adler-Baeder, 2015), was developed and refined immediately prior to the current project. CCM is an evidence-informed, scripted, 6-module HMRE curriculum that can be delivered in 9-12 hours. It emphasizes physiology, emotion, and the use of mindfulness-based stress reduction skills to address personal stress and to facilitate emotion regulation and healthy interactional patterns in couple relationships. Topics include: effects of stress and past trauma, usefulness of mindful practices, stress physiology, mindful stress/conflict management, compassion development, loving kindness skills, and mindful strategies for parenting/financial management. Similar to ELEVATE, CCM also emphasizes the NERMEM core relationship skills combined with a particular emphasis on stress management and emotion regulation through the use of mindfulness-based stress reduction strategies. The last

few decades of research document the prevalence of early trauma and toxic stress among more vulnerable populations (Fathers and Families Coalition of America Policy Brief, 2015); accordingly, ACF has emphasized considering early trauma exposure, toxic stress, and emotion regulation issues in HMRE work. Importantly, mindfulness practices have been shown to be useful for stress reduction among individuals (e.g., Grossman, Niemann and Schmidt, 2004; Kabat-Zinn, 1990), and moreover, recent research suggests potential relationship benefits among couples, with higher levels of mindfulness linked with higher couple relationship quality (McGill and Adler-Baeder, 2019) and more positive parenting (Burke et al., 2019). Based on previous research (DeMaria, 2005) and our participant records in previous AHMREI demonstration projects indicating that couples often enter the program with notable levels of distress and relational instability (DeMaria, 2005), this added emphasis on stress management was warranted.

## 3. Motivation for AHMREI impact evaluation

We chose to implement and test the effectiveness of the ELEVATE and CCM curricula because of their unique suitability to the needs of the economically and educationally diverse population of couples we expected to serve. Both curricula paid similar attention to economic/educational/racial diversity (in examples, pictures, videos, etc.), and both were developed through a theory-driven, evidence-informed process, distilling research knowledge into programming content (using the NERMEM framework, http://www.fcs.uga.edu/nermen/nermem). The two curricula differ, however, in their emphasis on stress management, with both including stress management skills, but CCM particularly focused on mindfulness-based stress management practices. Thus, the intent of the impact evaluation was to understand the efficaciousness of different types of programming among diverse couples. Additionally, both curricula are fairly new to the HMRE research field, and our aim was to further validate the evidence base for use of the ELEVATE curriculum and, in parallel, to conduct the first efficacy test of the CCM curriculum. To date, only a handful of studies have evaluated mindfulness-based HMRE interventions and have primarily used small samples of homogenous, non-distressed couples, without a control group. These studies suggest that mindfulness practices lead to lower stress, better emotion regulation, more empathy, and enhanced couple functioning and quality (Carson, et al., 2004; 2007; Gambrel and Piercy, 2014a; 2014b). To our knowledge, the current project is the first randomized control trial (RCT) impact evaluation of a mindfulness-based HMRE curriculum, CCM, with a large, economically and educationally diverse sample of couples and could have important implications for HMRE implementation among more vulnerable populations.

While the last decade has seen an explosion of research on HMRE programs, there is still much to learn (Wadsworth and Markman, 2012). Only a small portion of the HMRE projects that serve economically diverse populations of couples have participated in rigorous evaluations of program outcomes using RCT designs (e.g., a total of 16 research reports in the recent meta-analysis by Arnold and Beelmann, 2019). In addition, follow-up with program participants has been limited as there has been little exploration of relationships among outcomes over time after HMRE programming has ended (Wadsworth and Markman, 2012; Sher, 2012). Furthermore,

focused examinations of specific curricula with diverse populations are needed in the field of HMRE evaluation (Hawkins and Fackrell, 2010; Wadsworth and Markman, 2012). Prior evaluation studies tended to consider HMRE curricula as a group – as in meta-analyses—rather than emphasize distinctions or test curricula simultaneously. Thus, our impact evaluation represents a step forward in the development of the research base on HMRE.

The contributions of our impact evaluation include: 1) The conduct of a relatively rare RCT of HMRE program impacts, along with an assessment of implementation (which informs impact findings), delivered through a multi-site collaborative network of state and community-based organizations to a large sample of couples; 2) the simultaneous test of two similar, yet distinct, HMRE curricula (ELEVATE and CCM) and the evaluation of treatment effects across relational and individual outcomes for educationally and economically diverse Black and White couples, thus illuminating the utility of these new, research-informed curricula for use with these audiences; 3) the use of long-term follow-up assessments, which serves to inform policymakers, practitioners, and other researchers about sustained or delayed treatment effects of both curricula; and 4) a prospective test of the influence expected between gains in couple relationship skills targeted in the curricula on later improvements in couple satisfaction. Results from this project may contribute to the advancement of empirically-grounded models of best practices for working with diverse populations. Overall, the results of our impact evaluation will enhance understanding of the value of ELEVATE and CCM attendance and provide new information for the HMRE field.

## B. Primary impact research questions

Six primary research questions (RQs) were generated from our logic model (Appendix A) and were assessed in the impact evaluation of ELEVATE and CCM. Our logic model proposed that HMRE for adult couples will lead to immediate, short-term improvements in their healthy relationship skills. Primary Impact RQs #1 and #2 examined these short-term (2 months post-baseline, at program completion) impacts on couple relationship skills emphasized in the two curricula tested. Our logic model also hypothesized that short-term improvements in couple functioning would lay the foundation for long-term, sustained gains in individual and relationship functioning. Primary Impact RQs #3-6 assessed intermediate (six-month post-baseline) program impacts on individual and couple functioning. Additional secondary RQs (see Section C) assessed long-term (one-year post-baseline) program impacts. Specifically, we analyzed the following primary research questions:

1. What is the impact of ELEVATE when compared to the no-program control group on change in *couple relationship skills* from baseline to two-month post-baseline follow-up?
2. What is the impact of CCM when compared to the no-program control group on change in *couple relationship skills* from baseline to two-month post-baseline follow-up?
3. What is the impact of ELEVATE when compared to the no-program control group on change in *individual mental health* from baseline to six-month post-baseline follow-up?

4. What is the impact of CCM when compared to the no-program control group on change in *individual mental health* from baseline to six-month post-baseline follow-up?
5. What is the impact of ELEVATE when compared to the no-program control group on change in *couple satisfaction* from baseline to six-month post-baseline follow-up?
6. What is the impact of CCM when compared to the no-program control group on change in *couple satisfaction* from baseline to six-month post-baseline follow-up?

This study is registered on clinicaltrials.gov: https://clinicaltrials.gov/ct2/show/NCT03158714

## C. Additional research questions

In order to address other specific components of our logic model (Appendix A), we examined six additional research questions. Our logic model posits that long-term, sustained gains in individual and relationship functioning will stem from improved short-term couple functioning. Indeed, many HMRE evaluation studies have assessed program impacts only at short-term follow-up (i.e., immediate post-program) or have not found treatment effects that persist beyond six months (see meta-analysis by Hawkins et al., 2008), whereas other studies demonstrated gradual reductions in treatment gains at long-term follow-up (Gubits et al., 2014). Primary Impact RQs #3-6 (see Section I.B.) assessed treatment effects in individual and relationship functioning at six-month follow-up after baseline; thus, to better assess and understand possible long-term, sustained impacts, we extended the follow-up survey to one-year post-baseline to address our Additional RQs #1-4:

1. What is the impact of ELEVATE when compared to the no-program control group on change in *individual mental health* from baseline to one-year post-baseline follow-up?
2. What is the impact of CCM when compared to the no-program control group on change in *individual mental health* from baseline to one-year post-baseline follow-up?
3. What is the impact of ELEVATE when compared to the no-program control group on change in *couple satisfaction* from baseline to one-year post-baseline follow-up?
4. What is the impact of CCM when compared to the no-program control group on change in *couple satisfaction* from baseline to one-year post-baseline follow-up?

Further, in order to better illuminate the pathway proposed in our logic model (Appendix A) linking immediate short-term growth in couple relationship skills with longer-term gains in couple satisfaction, we examined a path model for each curriculum (see Appendix G, Figure 1 for illustration), associated with the following two Additional RQs:

5. Do changes immediately following ELEVATE participation in *couple relationship skills* (from baseline to two-month post-baseline follow-up) predict *couple satisfaction at* six-month post-baseline follow-up, accounting for baseline?
6. Do changes immediately following CCM participation in *couple relationship skills* (from baseline to two-month post-baseline follow-up) predict *couple satisfaction* at six-month post-baseline follow-up, accounting for baseline?

The remainder of the report is structured as follows: Section II discusses the intervention and counterfactual conditions and elaborates on the key research questions; Section III describes the study design, sample formation, and data collection methods; Section IV provides an overview of our analysis methods and discusses key characteristics of the analytic sample; Section V details the implementation and impact findings for our primary research questions, as well as sensitivity analyses and other analyses addressing additional research questions; and Section VI concludes with a discussion of key implementation and impact findings.

# II. INTERVENTION AND COUNTERFACTUAL CONDITIONS

This section describes the intended programs (i.e., components, content, implementation, and target population) and counterfactual conditions, as well as our research questions about the implementation of intervention and counterfactual conditions.

## A. Description of program as intended

For the program condition, this project used two similar, yet distinct, HMRE programs—ELEVATE and CCM—and delivered these multi-session programs to couples in community-based educational settings. Each curriculum's components included group-based classes in which couples received didactic content and interactive practice on relationship skills, workbooks/worksheets for couples to take home, and homework assignments for them to practice between sessions. The content of each of the two curricula is focused on key relationship skills and information that promote couple quality and stability (see NERMEM; www.nermen.org). These include: *self-care* (i.e., augmenting individual strengths and maintaining one's physical, sexual, emotional, and spiritual wellness); *intentionality* (i.e., demonstrating commitment and effort in the relationship); *conflict and stress management* (i.e., using strategies to maintain calm, have proactive conversations, mitigate unhealthy stress responses, and attentively listen); *intimate partner knowledge* (i.e., developing awareness and understanding of each other's day-to-day and long-term experiences, worries, needs, expectations, etc.); *couple identity* (i.e., sharing in each other's lives, finding common interests/goals, incorporating meaningful joint experiences); *caring couple behaviors* (i.e., expressing kindness through nurturing, affectionate behaviors without regard for reciprocity); and *supportive social connections* (i.e., developing and engaging in extended support networks with family, friends, or community organizations). As noted previously, CCM had an added emphasis on the use of mindfulness-based stress reduction strategies.

Program implementation was identical for both curricula, consisting of six sessions delivered in separate consecutive weekly sessions of approximately 2 hours each. Programs were delivered by 52 trained facilitators (male/female teams) in accessible community-based facilities that support a productive learning environment. Most facilitators had a background in family services and/or education and all had a minimum of a Bachelor's degree. Specifically, 73% of facilitators had a bachelor's degree; 24% had a Master's degree; and 3% had a doctorate. Each facilitator was provided a comprehensive 2-day training on each program curriculum (i.e., total of 4 days of training for the two curricula) by program developers prior to implementation. Facilitators then completed weekly fidelity checklists following each program session to track program components that were implemented in that session. Project staff also monitored classes for quality on a monthly basis (via videos, in-person observations, phone debriefings, etc.), provided ongoing technical/program assistance via phone and email on an as-needed basis, typically weekly, and conducted informative webinars to refresh or update strategies/protocols semiannually.

The target population included the following criteria: participants were legally adults in Alabama (age 19 or older) and self-defined as in a committed couple relationship (married or nonmarried). We did not explicitly target or screen for low-resource, disadvantaged couples (e.g., lower income or education). However, informed by our previous years of experience implementing HMRE, we expected that a sizable portion of our sample would be economically distressed and have lower educational attainment, due to the larger proportions in the general population of Alabama and to the characteristics of couples typically served by our family resource centers located in high-need counties. In order to recruit our target population, a number of public awareness methods were utilized (see Logic Model, Appendix A for more details), including media campaigns (i.e., digital, website, social media, billboard), community events/workshops in targeted communities served by the 10 implementation sites, and hard copies of past project reports and informative brochures about upcoming classes distributed to past participants, stakeholders, and community members at partner sites and other public places.

**Table II.1. Description of intervention and counterfactual program components, content, implementation, and target populations**

| Component | Program Content | Implementation | Target Population |
|---|---|---|---|
| **Intervention** | | | |
| Relationship skills sessions | CCM: Healthy relationships curriculum: communication skills; relationship skills; mindfulness-based stress relief skills<br>ELEVATE: Healthy relationships curriculum; communication skills; relationship skills, self-care skills | 6 weekly sessions (2 hours each)<br>Group lessons provided at the intervention's facilities by 2 trained facilitators in every session | Adult couples in a (self-defined) committed couple relationship |
| **Counterfactual** | | | |
| Location-specific resource list | Control respondents were provided with location specific resource lists. Resources covered: employment/work, housing, child services, mental and physical health services, education, addiction treatment and support, legal services, and multi-need services. | Resource list and questionnaire sent via email at baseline | Adult couples in a (self-defined) committed couple relationship |

**Table II.2. Staff training and development to support intervention and counterfactual components**

| Component | Education and initial training of staff | Ongoing training of staff |
|---|---|---|
| **Intervention** | | |
| Relationship skills sessions | Facilitators were male and female, held at least a bachelor's degree, and received four days of initial training (2 days per curriculum). | Project and evaluation staff observed facilitators and provided feedback sessions on a monthly basis, as well as offered ongoing technical/program assistance on an as-needed basis. Facilitators also participated in semiannual webinars for skill refreshers and updated methods/protocols. |
| **Counterfactual** | | |
| Location-specific resource list | Staff at partner agencies developed the resource list for their specific community. | Staff at partner agencies were able to update the resource list before each cohort recruitment period. |

## B. Description of counterfactual condition as intended

As soon as participants completed the baseline survey and were randomly assigned to the counterfactual condition (control group), they were emailed a list of location-specific family resources (e.g., employment, housing, child services, health, education, addiction treatment, and legal services) by the staff (i.e., program facilitator or case manager) at each site. No other programs and services were provided to the control group as part of the study.

## C. Research questions about the intervention and counterfactual conditions as implemented

Six Implementation research questions (RQs) focused only on the intervention groups (both ELEVATE and CCM), and two additional questions focused on the intervention and counterfactual conditions.

Two Implementation RQs addressed the implementation element of fidelity*:*

**1a. What is the average rate of adherence to curriculum content, as reported by ELEVATE facilitators?**

**1b. What is the average rate of adherence to curriculum content, as reported by CCM facilitators?**

Two Implementation RQs addressed dosage*:*

**2a. Of those assigned to the ELEVATE program group, what percentage of individuals attended none, 1-2 sessions, 3-5 session, and all 6 of the sessions offered?**

**2b. Of those assigned to the CCM program group, what percentage of individuals attended none, 1-2 sessions, 3-5 sessions, and all 6 of the sessions offered?**

Two Implementation RQs addressed the quality of the implementation:

**3. What is the average rating of facilitator quality (i.e., knowledge and ability to communicate effectively)?**

**4. What is the average rating of the facilitator-participant alliance?**

In order to assess engagement in intervention services we addressed two Implementation RQs:

**5. What is the average rating of self-engagement during the class?**

**6. What is the average rating of perceived engagement of other class members as a group?**

Finally, to address context we assessed two RQs, one for the two intervention groups and one for the control group:

7. **What proportion of the ELEVATE and CCM participants reported participating in other Healthy Marriage and Responsible Fatherhood (HMRF) programming or outside services from baseline to 1-year post-baseline follow-up?**

8. **What proportion of the control group reported participating in other HMRF programming or outside services from baseline to 1-year post-baseline follow-up?**

# III. STUDY DESIGN

This section describes the sample formation, research design (i.e., eligibility criteria, purposeful sampling, consent process, RCT research design), and data collection for both implementation and impact analyses, including engagement strategies used to reduce attrition and retain participants for the duration of the 2-year study.

## A. Recruitment and eligibility

The recruitment of participants for the efficacy study was conducted across the state and used broad eligibility requirements. Specifically, couples were recruited into our study in five separate cohorts over a two-and-a-half-year period, beginning in August 2016 and ending with one-year post-baseline follow-up surveys completed for the last cohort in February 2020. Recruitment across the 10 implementation sites began approximately a month before the start of programming for each cohort. The 10 implementation sites that covered a broad segment of the state geographically were: 1) Alabama Cooperative Extension System in Elmore County; 2) Circle of Care; 3) Family Guidance Center; 4) Family Success Center; 5) Hope Place; 6) IMPACT Family Counseling; 7) Auburn University; 8) Parents and Children Together (PACT); 9) Sylacauga Alliance for Family Enhancement (SAFE); and 10) Tuscaloosa's One Place (TOP). Recruitment of couples occurred through distribution of flyers (approved by the Auburn University Institutional Review Board [IRB]) at family resource centers, local churches, libraries, and other settings where community members were likely to frequent. Additionally, flyer information was posted online via the AHMREI Facebook page and website. All adults who indicated they were in a committed couple relationship (married or nonmarried) and who expected to attend as a couple if selected for a program group were eligible to participate in the RCT.

The following data collection procedures and consent process were approved by the Auburn University IRB that oversees Human Subjects protection in research. The original approval was provided on August 15, 2016 and is currently up to date after yearly renewals. The evaluation team conducted all steps of the consent and sample formation process. When couples or a member of a couple called the phone numbers (specific to each partner site) listed on AHMREI fliers and expressed interest in participating, we collected basic enrollment information and explained the nature of the evaluation project using a recruitment script. We explained random assignment procedures, reminding couples they had a 2 out of 3 chance to be assigned to one of the program groups and a 1 out of 3 chance they would be assigned to the control (no program) condition. We explained they would be asked to complete a series of surveys on the web-based system, Qualtrics, at the following times: before random assignment (baseline), 2-month follow-up after baseline (i.e., immediate post-program), 6-month follow-up after baseline, 1-year follow-up after baseline, and 2-year follow-up after baseline (although 2-year follow-up data is not included in this report). Participants were told to complete the surveys individually rather than as a couple. We offered the option of filling out these surveys through an emailed link or by coming to the nearest agency/center to complete the survey on a tablet. We also explained the

participant compensation plan: participants who completed a survey received $50 each at baseline, 2-month, 6-month, and 1-year follow-up, and $100 at 2-year follow-up.

Once a couple agreed to participate, they provided contact information (name, birthdate, address, phone, email) to implementation staff that was used to generate a participant profile in the sample management system (known as "nFORM"). The nFORM system is an online management information system developed by the Office of Planning, Research, and Evaluation and Mathematica that collects and stores performance measurement data on workshop attendance, case management activities, and participant outcome data on all HMRF participants in real time.

Enrollment took place both on the phone and in person. Information from each member of the couple was used to create an individual profile, and to generate both a participant and couple ID in the nFORM system. Immediately following enrollment in the nFORM system, each participant indicated their willingness to participate in the study by reading and signing the informed consent letter approved by the Auburn University IRB. Two weeks prior to program start, all study participants who signed and returned an informed consent letter were emailed a link to a baseline Qualtrics survey to be completed within a week. Following one or both partners' completion of informed consent and baseline survey, the couple was randomly assigned to one of the three groups (CCM, ELEVATE, or control).

A key feature of our evaluation design was the use of block random assignment procedures, the preferred method for random assignment when multiple sites are used in an evaluation study. Sites recruited couples for the study and participants were coded according to the implementation site block. For all consenting couples during each cohort period, if at least one of the individuals in a given couple completed the baseline survey, that couple was grouped with the other couples recruited by that site and randomly assigned to either one of the two program groups or to the no-program control group. This block random assignment approach was used to ensure a balance of randomly assigned groups at each implementation site and to best ensure an equal number of couples within each equal group in the final randomly assigned sample.

Each step of the random assignment process was implemented by evaluation staff. Couple IDs developed by the nFORM system were entered into separate site-specific SPSS files by the evaluation staff, and then the SPSS random number generator was used to code the couples as a 1 (ELEVATE), 2 (CCM), or 3 (control) in order to randomly assign the study participants. Overall, couples had a 2 out of 3 chance of being assigned to one of the program groups and a 1 out of 3 chance of being assigned to the control (no program) condition. This assignment probability remained constant across the 10 implementation sites and across all five cohorts.

The timeline progressed as follows. Recruitment efforts for each cohort were concentrated in one month. Following the recruitment period, the baseline survey was disseminated to all enrollees, who were then given two weeks to complete the survey. Following the deadline for completing the baseline survey, random assignment was conducted, and participants were notified by email of their assignment a week prior to program start. Thus, recruitment to random assignment spanned approximately one month. For those in the program groups, the notification email

included times and location of the assigned class. For those in the control group, the notification email included a list of community resources and the reminder that they would not be attending ELEVATE or CCM.

## B. Data collection

In this section, we first describe data collection procedures for the implementation analysis assessing fidelity, dosage, quality, engagement, and context for both intervention and counterfactual conditions. Next, we detail the data collection procedures for the impact analysis, including engagement strategies used to retain participants for the remainder of the 2-year study.

### 1. Implementation analysis

The following data sources were used to answer the Implementation RQs: nFORM workshop sessions, Qualtrics surveys from facilitators after workshop completion, and Qualtrics follow-up surveys from participants (see Table B.1. in Appendix B). To assess fidelity (Implementation RQs 1a, 1b), within days of completing each class session in a class series, facilitators completed survey items that pertained to their adherence to the curriculum content of ELEVATE and CCM. To assess dosage (Implementation RQs 2a, 2b), intervention staff tracked attendance in nFORM for individuals at all sessions during each of the 6-week program delivery periods. To assess quality (Implementation RQs 3, 4) and engagement (Implementation RQs 5, 6), study respondents provided information to evaluation staff on a Qualtrics survey that was sent via email at 2-month follow-up (immediately following the class series) related to facilitator quality and knowledge, the alliance between participant and facilitator, their own engagement in the classes, as well as the perceived engagement of other class members as a group. Intervention group participants had two weeks to complete the survey. Finally, to assess context (Implementation RQs 7, 8), both intervention and control study individuals responded to Qualtrics surveys sent via evaluation staff at 2 months, 6 months, 1 year, and 2 years after baseline data collection, and responded to questions related to additional services they may have accessed from baseline to 1-year follow-up other than the HMRE programs offered to the intervention group.

The data collection procedures related to the Implementation RQs are detailed in this report in Section II.C and in Appendix B, Table B.1. The Implementation RQs were addressed using either composite measures, average scores, or percentages, depending on the measure and the question. Composite scores were created for multi-item measures by averaging the values associated with each response. Individuals who responded to at least 80% of items on the measure(s) of focus in each Implementation RQ were retained as part of the analytic sample for that RQ, and mean imputation was used for missing items; however, two measures (Participant Involvement and Group Involvement) have only three items, so any participants who did not answer all three items were deleted from analyses.

## 2. Impact analysis

Key details about the survey collection timing and sources are in Table B.2. in Appendix B. Baseline (2 weeks prior to program start) and follow-up surveys (2-month, 6-month, 1-year, and 2-years post-baseline) were distributed via the online Qualtrics system and tracked in a master Excel list by evaluation staff. All impact study participants were emailed a link to the 2-month follow-up Qualtrics survey at the same time (within one week after the class series ended for program participants and approximately two months after the baseline survey was completed for control participants). This survey process was repeated at 6-month, 1-year, and 2-year follow-ups after baseline survey completion for all five cohorts. For follow-up surveys, participants were initially contacted one month prior to the data collection time point and given until one month after the data collection time point to complete the survey. Several prompts and reminders to complete the survey were sent during the "open" period and are further discussed below. On a daily basis the evaluation staff tracked data in a master Excel list as it was collected. Once the survey was submitted in the online Qualtrics system, the evaluation team added the date of completion for each survey in the row corresponding to the appropriate participant ID. The final master Excel list included everyone in the study and indicated which surveys were completed, when they were compensated for the survey completion, and for the program group, which classes were attended. Compensation was initiated via mailed checks within the week of receiving a completed survey from a participant.

Evaluation staff used the following engagement strategies to reduce attrition after enrollment: weekly emails and/or texts to remind participants of study activities and upcoming deadlines; emails and/or texts throughout the week of survey completion until the survey was submitted; and an email and/or text one day prior to class start for the program groups. Furthermore, evaluation staff also used the following strategies to retain participants up to the 2-year follow-up survey: birthday cards emailed to participants in the month of their birthday; anniversary cards (based on the date collected on baseline survey) emailed to participants in the month of their anniversary; "save the date" notices (via email, mail, and/or text) sent one week prior to emailing follow-up survey links; and reminder texts/emails sent during the "open" period for survey completion, which also included an offer to schedule appointments to come to the implementation site that recruited them and complete the survey on an iPad at that location.

# IV. ANALYSIS METHODS

This section describes the construction of the sample used to estimate the impact of the intervention on each targeted outcome. Information on the outcome measures used to address the Impact RQs and the baseline equivalence of the final treatment and control groups used in analyses are also presented.

## A. Analytic sample for impact evaluation

The analytic samples differed based on the specific RQ for two reasons: (1) response rates differed at 2-month and 6-month follow-up periods and (2) item nonresponse differed based on the outcome of interest. Attendance at the program (for program participants) was not a condition for inclusion in the analytic sample because we used an "intent to treat" (ITT) approach. The ITT approach is the most rigorous approach to evaluation and yields the most reliable efficacy results for testing a program in a "real world" setting (Weiss and Jacobs, 2008; Wood et al., 2014). The assessment of program impacts was based on the conditions to which participants were originally randomly assigned regardless of whether participants actually completed programming or whether control respondents crossed over to a program group. There were, however, no crossovers from the control group to either program group.

We originally planned to randomly assign 750 couples (250 couples assigned to each group); however, we successfully over-recruited participants for the study. Of the 2,308 individuals who contacted implementation sites with interest and were sent the informed consent letter and baseline survey by the evaluation team, 1,796 completed the baseline survey. If at least one individual in the couple completed the baseline survey, the couple was included in the study and was randomly assigned to one of three study conditions ($N$ = 930 couples or 1,860 individuals in total). Couples were randomly assigned to one of the program groups – either ELEVATE ($N$ = 313 couples/626 individuals) or CCM ($N$ = 313 couples/626 individuals) – or the control group ($N$ = 304 couples/608 individuals). Our final sample of 930 couples is 124% of our recruitment goal for the study.

For analyses of each Impact RQ, we included data from respondents who completed surveys at all time points relevant to the RQ and who had very little missing data on the relevant measures. Specifically, in order to minimize bias in the results, we included only respondents who completed 80% or more of the measure items. For each primary Impact RQ we provide CONSORT Images A, B, and C in Appendix B for details about how the final samples for these analyses were selected. *Overall attrition* of program and control participants from random assignment to inclusion in the final analytic sample is indicated by the percentage of randomly assigned participants who are not included in the analytic sample, and the *differential attrition* rate between program and control groups is the difference in attrition rates between groups. The What Works Clearinghouse (WWC) guidelines provide demarcations to inform acceptable to high levels of attrition. Because attrition can introduce bias into the analysis, the WWC offers evidence-based guidelines to demarcate acceptable combinations of overall and differential attrition. For a given overall attrition rate, there is a range of what is considered acceptable

differential attrition, with the cautious boundary representing the most stringent level of allowable attrition. In general, the higher the overall attrition rate, the narrower the differential attrition rate needs to be to avoid potential bias. Because we had very high retention standards for the analytic sample (i.e., 80+% of outcome measure completed at all time points relevant to the RQ) there were 3 out of 6 instances when allowable differential attrition rates did not meet the cautious boundary. This is attributable to the conditions of the completion of all survey timepoints and completing either 80+% or all the measure items.

In sum, attrition analyses related to ELEVATE research questions (#1, 3, and 5) indicated the combination of overall and differential attrition was high (in relation to the WWC's cautious boundary), whereas attrition analyses related to CCM research questions (#2, 4, and 6) indicated the combination of overall and differential attrition was acceptable (in relation to the WWC's cautious boundary). Table IV.1a. provides details about overall, differential, and WWC guidelines regarding attrition rates, as well as whether or not the results are acceptable or considered high. Differential attrition rates below the WWC cautious boundary indicate acceptable levels of attrition while rates above the boundary indicate high attrition. Specific details related to each research question can be found in Section VIII.C.

**Table IV.1a. Attrition rates (overall, differential, and WWC accepted rates) and conclusion on attrition rates for RQs #1-6**

| Attrition rates and WWC conclusion on attrition | RQ #1 | RQ #2 | RQ #3 | RQ #4 | RQ #5 | RQ #6 |
|---|---|---|---|---|---|---|
| Overall attrition rate | 18% | 16% | 35% | 32% | 33% | 30% |
| Differential attrition rate (between program and control groups) | 7% | 3% | 6% | 0% | 5% | 1% |
| WWC accepted attrition rate (i.e., cautious boundary) | 5.7% | 5.9% | 3.3% | 3.8% | 3.6% | 4.1% |
| Conclusion on study attrition (based on WWC accepted rate) | High | Acceptable | High | Acceptable | High | Acceptable |

RQ = Research Question; WWC = What Works Clearinghouse.

## B. Outcome measures

The following outcome measures were used for both primary and additional Impact RQs:

### 1. Primary RQs #1 and #2 and additional RQs #5 and #6

*Couple relationship skills* were measured using 32 items from the Couple Relationship Skills Inventory (CRSI). This measure was constructed to match the core relationship skills and predictors of couple quality emphasized in the HMRE programs provided. Items were taken from established and validated social science measures assessing commitment, intimate partner knowledge, friendship, caring behaviors, conflict management, and external support. Example items include, "I commit effort every day to making my relationship work," "I know my partner's current life stresses," "When things 'get heated' I suggest we take a break to calm down," and "I initiate physical affection with my partner." Items ranged from 1 (Very strongly

disagree) to 7 (Very strongly agree). Composite sum scores were created at the individual level. Higher scores indicate *greater couple relationship skills*. As noted, program and control participants were only included in the analytic sample if they had responded to at least 80% of items (at least 25 of 32 total). Item nonresponse was addressed by mean imputation. Psychometric analyses confirmed the validity and reliability of the full measure for our sample (Adler-Baeder et al., 2019). To assess model fit, we used a Bayesian variant of the Root Mean Squared Error (BRMSEA; Hoofs et al., 2018), and incremental fit indices including the Bayesian Confirmatory Factor Index (BCFI), Bayesian Tucker Lewis Index (BTLI), and Bayesian Normed Fit Index (BNFI). A ppp value of around .10 (Cain and Zhang, 2018), a BRMSEA value smaller than .08 (Hoofs et al., 2018), and BCFI, BTLI, and BNFI values above .95 indicate good model fit and values above .90 indicating acceptable model fit (Asparovhov and Muthén, 2019). The Bayesian confirmatory factor analyses indicated a high degree of fit for the measurement model (ppp < .001; BRMSEA = .043; BCFI =.949; BTLI = .932; BNFI = .920; whereby non-informative priors were used and model fit was assessed by using posterior predictive p-value (ppp), The assessment of Cronbach's alpha indicated high reliability at baseline and 2-month follow-up for both men (α = .91; .93) and women (α = .92; .93).

### 2. Primary RQs #3 and #4 and additional RQs #1 and #2

*Individual mental health* was measured using the SF-12 Mental Component Summary score (Ware, Kosinski, and Keller, 1996). The measure assessed mental health through items pertaining to anxiety, depression, sense of wellbeing, and social/emotional functioning. Per the measure's design, items were asked on differing scales (e.g., 1 to 3 or 1 to 5) with different anchor responses (e.g., all of the time to none of the time, or not at all to extremely). Example items include, "During the past month how much of the time have you accomplished less than you would like?" or "During the past month how often have you felt downhearted and depressed?" The SF-12 Mental Component Summary score has shown scale reliability and validity in empirical studies (Gandek et al., 1998; Ware, Kosinski, and Keller, 1996). Following the SF-12 scoring instructions, item responses were first standardized, then summed, and standardized further by adding 60.75781 to create a population-normed score (Maruish, 2012). Scores range from 0 to 100 with a mean of 50 and standard deviation (SD) of 10 in the general U.S. population (Gandek et al., 1998; Ware et al., 1996). Higher scores indicate *better individual mental health*. Due to the complex scoring of the SF-12 measure, mean imputation methods cannot be used for missing items; thus, only participants who responded to 100% of the 12 items were retained.

### 3. Primary RQs #5 and #6 and additional RQs #3, #4, #5, and #6

*Couple satisfaction* was measured using an abbreviated version of the Couple Satisfaction Index (CSI; Funk and Rogge, 2007), which was utilized in previously published studies (Galovan, Drouin, and McDaniel, 2018; Kevin and Risla, 2020). The 3 items (rated from 1 [Not at all] to 6 [Completely]) are: "I have a warm and comfortable relationship with my partner," "How rewarding is your relationship with your partner?" and "In general, how satisfied are you with your relationship?" Thus, higher scores indicate *higher couple satisfaction.* Program and control

participants were only retained if they responded to all 3 items since completing 2 of the 3 items would not meet the 80+% criterion for mean imputation used in this study. As such, no scores were imputed. Composite sum scores were created at the individual level since they reflect the individual's satisfaction with their couple relationship, and assessment of Cronbach's alpha indicated high reliability at baseline, 2-month, and 6-month follow-up for both men (α = .92; .95; .95) and women (α = .94; .95; .96).

**Table IV.2. Outcome measures used for primary and additional impact analyses research questions**

| Outcome measure | Description of the outcome measure | Example Item | Source | Timing for primary questions | Timing for additional questions | Citation |
|---|---|---|---|---|---|---|
| Couple relationship skills | The outcome measure is an individual composite score of 32 questions asked on a 7-point Likert scale ($\alpha_{men}$ = .91; $\alpha_{women}$ = .92) | I commit effort every day to making my relationship work | Qualtrics survey | Baseline and 2-month follow-up | Baseline and 2-month follow-up | Adler-Baeder et al., 2019 |
| Mental health | The outcome measure is an individual score calculated from 12 questions per the SF-12 instructions. | During the past month how much of the time have you accomplished less than you would like | Qualtrics survey | Baseline, 2-month, and 6-month follow-up | Baseline, 2-month, 6-month, and 1-year follow-up | Ware, Kosinski, and Keller, 1996 |
| Couple satisfaction | The outcome measure is the couples' composite score from 3 questions on a 5-point Likert scale. ($\alpha_{men}$ = .92; $\alpha_{women}$ = .94) | I have a warm and comfortable relationship with my partner. | Qualtrics survey | Baseline, 2-month, and 6-month follow-up | Baseline, 2-month, 6-month, and 1-year follow-up | Funk and Rogge, 2007 |

## C. Sample characteristics and baseline equivalence

### 1. Sample characteristics

The sample characteristics did not statistically differ across analytic samples for Impact RQs #1-6, thus we present the sample description of the largest groups here (i.e., the samples for Impact RQs #1 and 2). The sample was composed of men (47%) and women (53%), of whom the majority (71%) were married, whereas 29% reported being in a committed non-marital relationship. Almost three-quarters (73%) of the sample reported being a parent (e.g., biological, step, adoptive, etc.). The majority of the sample reported their race as White (63%), while a little over a third were either Black (32%) or other (5%; Native-American, Asian-American, mixed, etc.). Almost a quarter (24%) of respondents reported having a high school diploma, GED, or less; 12% had a 2-year degree (i.e., vocational, technical, or associate's degree); 20% had some college experience, but had not completed college; 24% possessed a bachelor's degree; and 20% had an advanced degree (i.e., Master's or doctorate). With regard to socioeconomic status, the following proportions of household income were reported: 26% reported earning less than $24,999; 15% reported earning $25,000-39,999; 30% reported earning $40,000-74,999; 15%

reported $75,000-99,999; and 14% reported greater than $100,000. Thus, the educational background and household income of study respondents were diverse, but, in general, the sample was majority lower-resource individuals, representative of Alabama's population.

## 2. Baseline equivalence

We tested baseline equivalence between groups in the analytic samples for two primary reasons. First, for those outcomes that exhibited high attrition, establishing baseline equivalence helps reduce the threat of bias, which is increased when attrition is high. Second, for other outcomes with low attrition, baseline equivalence between groups provides an assessment of how successful the randomization process worked to achieve balance between the groups. Testing for whether there were any observed substantial differences between the treatment groups and the control group enhances the validity of the assumption that tests of differences in outcomes were due to program effects and not to systematic differences between groups or because of high attrition rates.

Differences between each program group and the control group on baseline demographics (i.e., race, age, income, marital status and parent status) and the baseline levels of the outcome variables targeted for the impact evaluation (i.e., couple relationship skills, mental health, and couple satisfaction) were assessed using independent sample *t*-tests for continuous variables and crosstabs with chi-square tests for categorical variables wherein statistical significance in differences is determined at $p < .004$. We utilized a critical *p* value of .004 or less instead of the traditional .05 to adjust for multiple comparisons (McDonald, 2014) since 12 tests (6 for men and 6 for women) were conducted to test differences for each research question. If the test indicated statistically significant results, the groups were considered different from each other on that demographic variable or in reports at baseline. To test whether statistically significant differences were substantively meaningful, either the Fisher's *z* (for categorical variables) or the Cohen's *d* (for continuous variables) effect size for group differences was calculated based on the analytic sample. These tests were conducted for the analytic samples for each research question with non-imputed data, and for men and women separately to address the clustering of data.

Tables IV.3, IV.4, and IV.5 summarize the baseline equivalence results for all impact research questions. Overall, the random assignment process was successful in creating nearly equivalent groups at baseline, and this is also true even for the outcomes with high attrition. Specifically, there were differences related to relationship status; however, these differences were only relevant for two of the six impact RQs (more details are provided below). Moreover, for all the impact RQs, no significant differences on baseline levels of outcomes (i.e., CRSI, mental health, couple satisfaction) were found. We utilized the results of these hypothesis tests to make decisions about inclusion of covariates in subsequent models. Per the results of the tests, marital status was included as a covariate in subsequent model testing because it was the only statistically significant indicator of group differences.

- **Impact RQ #1:** This research question was impacted by high attrition rates, and results indicate female ELEVATE participants were more likely to be married compared to female control respondents (76% vs. 65%; *Chi-Square* = 8.17, *p* = .004; see Table IV.3).
- **Impact RQ #2:** No statistically significant baseline differences were apparent.
- **Impact RQ #3:** This research question was impacted by high attrition rates, and results indicate female ELEVATE participants were more likely to be married compared to female control respondents (79% vs. 64%; *Chi-Square* = 11.64, *p* = .001; see Table IV.4).
- **Impact RQ #4:** No statistically significant baseline differences were apparent.
- **Impact RQ #5:** No statistically significant baseline differences were apparent.
- **Impact RQ #6:** No statistically significant baseline differences were apparent.

**Table IV.3. Summary statistics of key baseline measures and baseline equivalence across study groups, for individuals completing Couple Relationship Skills Inventory (CRSI)**

| CRSI baseline measure | ELEVATE | CCM | Control | ELEVATE versus control difference (*p*-value) | *Effect size* | CCM versus control difference (*p*-value) | *Effect size* |
|---|---|---|---|---|---|---|---|
| **Race (%)** | | | | | | | |
| African American/Black | 71 (30%)<br>**73 (28%)** | 74 (31%)<br>**89 (32%)** | 86 (35%)<br>**93 (34%)** | | | | |
| Caucasian/White | 154 (66%)<br>**168 (64%)** | 154 (63%)<br>**173 (63%)** | 148 (60%)<br>**166 (60%)** | $X^2$ = 2.08 (.72)<br>**$X^2$ = 3.76 (.44)** | *z* = .07<br>***z* = .08** | $X^2$ = 2.36 (.67)<br>**$X^2$ = 4.23 (.52)** | *z* = .07<br>***z* = .09** |
| Other | 10 (4%)<br>**21 (8%)** | 15 (6%)<br>**13 (5%)** | 11 (5%)<br>**17 (6%)** | | | | |
| Age mean (SD) | 39.26 (12.53)<br>**35.98 (11.05)** | 40.45 (13.58)<br>**37.17 (12.04)** | 37.46 (11.48)<br>**35.45 (11.00)** | *t* = 1.64 (.10)<br>***t* = .56 (.58)** | *d* = .15<br>***d* = .05** | *t* = 2.61 (.01)<br>***t* = 1.75 (.08)** | *d* = .24<br>***d* = .15** |
| Total household income before taxes in current year mean (SD) | 4.62 (1.65)<br>**4.54 (1.67)** | 4.72 (1.63)<br>**4.60 (1.61)** | 4.57 (1.68)<br>**4.42 (1.73)** | *t* = .32 (.75)<br>***t* = .80 (.42)** | *d* = .03<br>***d* = .07** | *t* = .98 (.33)<br>***t* = 1.24 (.22)** | *d* = .09<br>***d* = .12** |
| **Marital status number (%)** | | | | | | | |
| Married | 179 (76%)<br>**200 (76%)** | 179 (73%)<br>**194 (70%)** | 162 (66%)<br>**179 (65%)** | $X^2$ = 5.89 (.02)<br>**$X^2$ = 8.17* (.004)** | *z* = .11<br>***z* = .12** | $X^2$ = 3.04 (.08)<br>**$X^2$ = 1.70 (.19)** | *z* = .08<br>***z* = .06** |
| Non-married | 56 (24%)<br>**62 (24%)** | 65 (27%)<br>**82 (30%)** | 83 (34%)<br>**96 (35%)** | | | | |
| **Parent status number (%)** | | | | | | | |
| Parent | 161 (69%)<br>**194 (74%)** | 167 (68%)<br>**203 (74%)** | 176 (72%)<br>**206 (75%)** | | | | |
| Non-parent | 72 (31%)<br>**65 (25%)** | 72 (30%)<br>**71 (26%)** | 65 (26%)<br>**64 (23%)** | $X^2$ = 2.57 (.28)<br>**$X^2$ = 1.29 (.53)** | *z* = .07<br>***z* = .05** | $X^2$ = .70 (.70)<br>**$X^2$ = 3.05 (.22)** | *z* = .04<br>***z* = .07** |
| Prefer not to answer | 1 (< 1%)<br>**2 (1%)** | 5 (2%)<br>**1 (< 1%)** | 4 (2%)<br>**5 (2%)** | | | | |
| CRSI mean (SD) | 5.13 (.75)<br>**5.11 (.75)** | 5.18 (.75)<br>**5.10 (.75)** | 5.25 (.61)<br>**5.09 (.74)** | *t* = -1.88 (.06)<br>***t* = .30 (.76)** | *d* = .17<br>***d* = .03** | *t* = -1.00 (.32)<br>***t* = .11 (.91)** | *d* = .09<br>***d* = .01** |
| Sample size | 235<br>**262** | 244<br>**276** | 245<br>**276** | n.a. | n.a. | n.a. | n.a. |

Notes: Women's results are in bold. Income is reported on a scale from 1 to 7. 1 represents "< $7,000," 4 represents "$25,000–$39,999," and 7 represents "≥ $100,000." CRSI = Couple Relationship Skills Inventory; n.a. = not applicable; SD = standard deviation; $X^2$ = chi-square statistic; t = t statistic; z = Fisher's z; d = Cohen's d. * p < .004.

**Table IV.4. Summary statistics of key baseline measures and baseline equivalence across study groups, for individuals completing SF-Mental Health**

| Mental Health baseline measure | ELEVATE | CCM | Control | ELEVATE versus control difference (*p*-value) | *Effect size* | CCM versus control difference (*p*-value) | *Effect size* |
|---|---|---|---|---|---|---|---|
| **Race (%)** | | | | | | | |
| African American/Black | 52 (30%)<br>**62 (29%)** | 64 (32%)<br>**73 (32%)** | 65 (34%)<br>**74 (33%)** | | | | |
| Caucasian/White | 114 (65%)<br>**134 (63%)** | 122 (62%)<br>**141 (63%)** | 118 (63%)<br>**137 (60%)** | $X^2$ = 2.40 (.66)<br>**$X^2$ = 2.05 (.73)** | $z$ = .11<br>**$z$ = .07** | $X^2$ = 3.56 (.47)<br>**$X^2$ = 5.18 (.40)** | $z$ = .14<br>**$z$ = .11** |
| Other | 9 (5%)<br>**18 (8%)** | 12 (6%)<br>**10 (4%)** | 6 (3%)<br>**16 (7%)** | | | | |
| Age mean (SD) | 38.56 (11.28)<br>**36.53 (11.15)** | 40.03 (13.68)<br>**36.94 (11.52)** | 38.11 (11.91)<br>**35.94 (11.22)** | $t$ = .37 (.71)<br>**$t$ = .55 (.58)** | $d$ = .04<br>**$d$ = .05** | $t$ = 1.46 (.14)<br>**$t$ = .93 (.35)** | $d$ = .04<br>**$d$ = .09** |
| Total household income before taxes in current year mean (SD) | 4.69 (1.61)<br>**4.58 (1.66)** | 4.69 (1.65)<br>**4.69 (1.62)** | 4.63 (1.65)<br>**4.50 (1.71)** | $t$ = .33 (.74)<br>**$t$ = .49 (.62)** | $d$ = .03<br>**$d$ = .05** | $t$ = .30 (.76)<br>**$t$ = 1.17 (.24)** | $d$ = .03<br>**$d$ = .11** |
| **Marital status number (%)** | | | | | | | |
| Married | 131 (76%)<br>**165 (79%)** | 141 (71%)<br>**154 (69%)** | 127 (67%)<br>**144 (64%)** | $X^2$ = 3.55 (.06)<br>**$X^2$ = 11.64** (.001)** | $z$ = .14<br>**$z$ = .16** | $X^2$ = .87 (.35)<br>**$X^2$ = 1.61 (.21)** | $z$ = .08<br>**$z$ = .06** |
| Non-married | 41 (24%)<br>**45 (21%)** | 56 (28%)<br>**68 (31%)** | 62 (33%)<br>**82 (36%)** | | | | |
| **Parent status number (%)** | | | | | | | |
| Parent | 119 (68%)<br>**155 (72%)** | 130 (66%)<br>**160 (71%)** | 133 (70%)<br>**169 (74%)** | | | | |
| Non-parent | 54 (31%)<br>**57 (27%)** | 63 (32%)<br>**64 (28%)** | 54 (29%)<br>**54 (24%)** | $X^2$ = 4.92 (.78)<br>**$X^2$ = .97 (.62)** | $z$ = .16<br>**$z$ = .05** | $X^2$ = 1.80 (.41)<br>**$X^2$ = 2.89 (.24)** | $z$ = .10<br>**$z$ = .08** |
| Prefer not to answer | 1 (1%)<br>**2 (< 1%)** | 5 (2%)<br>**1 (< 1%)** | 2 (1%)<br>**4 (2%)** | | | | |
| Mental Health mean (SD) | 44.58 (9.87)<br>**42.00 (10.49)** | 45.06 (9.50)<br>**42.37 (10.58)** | 45.59 (10.25)<br>**42.59 (10.20)** | $t$ = -.95 (.34)<br>**$t$ = -.60 (.55)** | $d$ = .10<br>**$d$ = .06** | $t$ = -.53 (.60)<br>**$t$ = -.22 (.82)** | $d$ = .05<br>**$d$ = .02** |
| Sample size | 175<br>**214** | 198<br>**225** | 189<br>**227** | n.a. | n.a. | n.a. | n.a. |

Notes: Women's results are in bold. Income is reported on a scale from 1 to 7. 1 represents "< $7,000," 4 represents "$25,000–$39,999," and 7 represents "≥ $100,000." n.a. = not applicable; SD = standard deviation; $X^2$ = chi-square statistic; t = t statistic; z = Fisher's z; d = Cohen's d. ** p < .004.

**Table IV.5. Summary statistics of key baseline measures and baseline equivalence across study groups, for individuals completing Couple Satisfaction Index (CSI)**

| CSI baseline measure | ELEVATE | CCM | Control | ELEVATE versus control difference (p-value) | *Effect size* | CCM versus control difference (p-value) | *Effect size* |
|---|---|---|---|---|---|---|---|
| **Race (%)** | | | | | | | |
| African American/Black | 58 (31%)<br>**58 (27%)** | 61 (30%)<br>**74 (33%)** | 65 (34%)<br>**76 (33%)** | | | | |
| Caucasian/White | 121 (64%)<br>**143 (65%)** | 131 (63%)<br>**142 (63%)** | 121 (62%)<br>**140 (60%)** | $X^2$ = .64 (.96)<br>**$X^2$ = 3.16 (.53)** | $z$ = .04<br>**$z$ = .08** | $X^2$ = 2.57 (.63)<br>**$X^2$ = 5.85 (.32)** | $z$ = .08<br>**$z$ = .11** |
| Other | 9 (5%)<br>**17 (8%)** | 14 (7%)<br>**9 (4%)** | 9 (4%)<br>**16 (7%)** | | | | |
| Age mean (SD) | 39.37 (11.81)<br>**35.81 (10.43)** | 39.98 (13.70)<br>**37.48 (12.38)** | 37.65 (11.54)<br>**35.64 (10.82)** | $t$ = 1.44 (.15)<br>**$t$ = .17 (.87)** | $d$ = .15<br>**$d$ = .02** | $t$ = 1.83 (.07)<br>**$t$ = 1.70 (.09)** | $d$ = .18<br>**$d$ = .16** |
| Total household income before taxes in current year mean (SD) | 4.68 (1.62)<br>**4.55 (1.66)** | 4.69 (1.64)<br>**4.67 (1.64)** | 4.63 (1.66)<br>**4.50 (1.70)** | $t$ = .31 (.76)<br>**$t$ = .32 (.75)** | $d$ = .03<br>**$d$ = .03** | $t$ = .39 (.70)<br>**$t$ = 1.10 (.27)** | $d$ = .04<br>**$d$ = .10** |
| **Marital status number (%)** | | | | | | | |
| Married | 146 (78%)<br>**170 (78%)** | 149 (72%)<br>**163 (72%)** | 131 (68%)<br>**154 (66%)** | $X^2$ = 4.92 (.03)<br>**$X^2$ = 7.51 (.01)** | $z$ = .11<br>**$z$ = .13** | $X^2$ = 1.10 (.30)<br>**$X^2$ = 1.77 (.18)** | $z$ = .05<br>**$z$ = .06** |
| Non-married | 42 (22%)<br>**48 (22%)** | 57 (28%)<br>**63 (28%)** | 63 (32%)<br>**78 (34%)** | | | | |
| **Parent status number (%)** | | | | | | | |
| Parent | 132 (71%)<br>**157 (72%)** | 136 (66%)<br>**163 (73%)** | 137 (71%)<br>**171 (74%)** | | | | |
| Non-parent | 54 (29%)<br>**58 (27%)** | 65 (32%)<br>**61 (27%)** | 52 (27%)<br>**57 (25%)** | $X^2$ = 1.07 (.59)<br>**$X^2$ = .37 (.83)** | $z$ = .05<br>**$z$ = .03** | $X^2$ = 1.46 (.48)<br>**$X^2$ = 1.25 (.54)** | $z$ = .06<br>**$z$ = .05** |
| Prefer not to answer | 1 (< 1%)<br>**2 (1%)** | 5 (2%)<br>**1 (< 1%)** | 3 (2%)<br>**3 (1%)** | | | | |
| CSI mean (SD) | 4.84 (1.16)<br>**4.75 (1.28)** | 4.81 (1.03)<br>**4.75 (1.17)** | 5.01 (1.05)<br>**4.82 (1.17)** | $t$ = -1.40 (.12)<br>**$t$ = -.62 (.54)** | $d$ = .15<br>**$d$ = .06** | $t$ = -1.85 (.07)<br>**$t$ = -.58 (.56)** | $d$ = .19<br>**$d$ = .05** |
| Sample size | 188<br>**218** | 206<br>**226** | 194<br>**232** | n.a. | n.a. | n.a. | n.a. |

Notes: Women's results are in bold. Income is reported on a scale from 1 to 7. 1 represents "< $7,000," 4 represents "$25,000–$39,999," and 7 represents "≥ $100,000." CSI = Couple Satisfaction Index; n.a. = not applicable; SD = standard deviation; $X^2$ = chi-square statistic; t = t statistic; z = Fisher's z; d = Cohen's d. * p < .004.

# V. FINDINGS AND ESTIMATION APPROACH

## A. Implementation evaluation

### Key Findings:

- ELEVATE and CCM facilitators adhered to the curriculum content, reporting on average that they "somewhat agree" that they followed the program as designed.
- ELEVATE and CCM participants had high attendance rates as well over half of participants received 50 percent or more of the content.
- Program participants highly rated the facilitation quality (on average, 92 out of 100) and their alliance with their facilitators (on average, 93 out of 100).
- Program participants were actively engaged in the class and perceived the others in the class to be actively engaged in the material.
- Approximately a quarter of both program (ELEVATE and CCM) and control respondents participated in outside therapeutic/counseling services from baseline to one-year follow-up).

**FIDELITY (Implementation RQs #1a, #1b).** Within two weeks of completing each class series, facilitators completed a survey assessing their adherence to the curriculum content of ELEVATE or CCM. During the period of time classes were being implemented for the study, we completed 92 class series (either ELEVATE or CCM) and obtained 158 completed facilitator surveys for a response rate of 86%. To measure fidelity, facilitators responded to 10 items assessing their adherence to curriculum design and content (e.g., "I followed the curriculum guide in the process of teaching the program;" rated 1 = Very Strongly Disagree to 7 = Very Strongly Agree), with higher scores indicating a greater level of adherence to the curriculum. Average scores were computed, and the fidelity scale was internally consistent ($\alpha = .80$). ELEVATE facilitators reported a mean fidelity score of 4.72 and CCM facilitators reported a mean fidelity score of 4.59, suggesting that, on average, they "somewhat agree" that they followed the program as designed.

**DOSAGE (Implementation RQs #2a, #2b).** Intervention staff tracked attendance for individuals at all sessions during each of the 6-week program delivery periods. To evaluate dosage, we created groups of participants based on number of sessions attended. Specifically, we recoded into a new categorical variable: 1) those who were randomly assigned to a program but never attended a session; 2) those who were randomly assigned to a program and attended 1 or 2 sessions; 3) those who were randomly assigned to a program and attended 3, 4, or 5, sessions; and 4) those who were randomly assigned to a program and attended all (6) of the sessions. We calculated frequencies of these groupings and found that of those randomly assigned to ELEVATE, 25% attended no classes, 9% attended 1 or 2 classes, 39% attended 3 to 5 sessions, and 27% attended all classes offered. The average dose for ELEVATE participants was 3.5

classes. Of those randomly assigned to CCM, 15% attended no classes, 12% attended 1 or 2 classes, 39% attended 3 to 5 sessions, and 34% attended all classes offered. The average dose for CCM participants was 4 classes. For both curricula, the average participant attended more than half the classes, which is considered an "appropriate dosage" for HMRE in a recent meta-analysis (Arnold and Beelman, 2019).

**QUALITY (Implementation RQs #3, #4).** Study participants completed a Qualtrics survey at 2-month follow-up (i.e., immediately following the class series) related to facilitator quality and the alliance between the participant and facilitator. Intervention group respondents had two weeks to complete the survey. Participants' assessment of facilitator quality was assessed with four items (response ranges from 0 to 100 with anchors 0 = "The facilitators did not allow quality class interactions and did not manage time constraints" and 100 = "allowed quality class interactions and managed time constraints well"), and mean composite scores were utilized. Cronbach's alpha for facilitator quality ($\alpha = .92$) indicates high reliability. Additionally, participants reported on 6 items assessing the alliance formed between the facilitator and participant (e.g., 0 = "The facilitators were not approachable nor were they personable in helping us" to 100 = "The facilitators were approachable and personable in helping us"), and mean composite scores were utilized. Cronbach's alpha for facilitator alliance ($\alpha = .92$) indicates high reliability. Higher scores indicate better facilitation quality or alliance. Both ELEVATE and CCM program participants reported a facilitation quality mean score of 92.43, on a scale of 0 to 100, suggesting that program participants perceived they engaged with high-quality facilitators. Similarly, both ELEVATE and CCM program participants reported a facilitator alliance mean score of 92.38, on a scale of 0 to 100, suggesting the program participants perceived a close alliance with their facilitators.

**ENGAGEMENT (Implementation RQs #5, #6).** Study respondents completed a Qualtrics survey at 2-month follow-up (i.e., immediately following the class series) related to their own engagement level in the classes, as well as the perceived engagement of the other class members as a group. Intervention group respondents had two weeks to complete the survey. Separate scores were calculated for engagement level of participants (3 items averaged; e.g., "I actively participated in the class activities" rated 1 = Very Strongly Disagree to 7 = Very Strongly Agree) and engagement of the class participants as a group (3 items averaged; e.g., "Participants in this class took an active part in class discussions" rated 1 = Very Strongly Disagree to 7 = Very Strongly Agree). Higher scores indicate higher participant or group engagement. Cronbach's alpha for participant engagement ($\alpha = .91$) and group engagement ($\alpha = .94$) indicate high reliability. Both ELEVATE and CCM participants reported a self-engagement mean score of 5.74, on a scale from 1 to 7, suggesting that, on average, they "agree" that they were actively engaged in the class content, activities, and discussions. Similarly, both ELEVATE and CCM program participants reported group engagement mean score of 5.88, on a scale from 1 to 7, suggesting that, on average, they "agree" that others in the class were actively engaged in the class content, activities, and discussions.

**CONTEXT (Implementation RQs #7, #8).** Both intervention and control study respondents completed Qualtrics surveys sent via evaluation staff at 2-month, 6-month, and 1-year follow-up

after baseline data collection and responded to questions related to additional services they may have accessed during the period from baseline to 1-year follow-up other than the HMRE programs offered to the intervention group. For both the intervention group and the control group, context was assessed by examining the percentage of respondents in each group who reported participating in other HMRF programming or other outside services from baseline to 1-year follow-up. This dichotomous measure (1 = yes; 0 = no) allowed respondents to identify if they had received any of the following services: individual therapy, couple therapy, family therapy, marriage and family therapy, cognitive behavioral therapy, counselor, life coach, psychologist, religious leader, and other. Note that a participant could respond "yes" to more than one of these. Although there was no cross-over between the control group participants and the program group participants for HMRE attendance, some program participants and some control group participants did report receiving other services offered in the community.

Since case management is offered to all HMRE program participants, some involvement in community services occurred for those randomly assigned to a program group. Specifically, 26% of program participants (24% of ELEVATE participants and 29% of CCM participants) reported they had received therapeutic or counseling services available in the community at some point from baseline to 1-year follow-up. Some participated in one type of outside service, while others participated in and reported multiple types of services. Specifically, of those in the program participant groups who reported receiving outside services (n = 328), 177 participated in individual therapy, 231 participated in couples therapy, and 38 participated in family therapy.

We also tracked whether control participants made use of mental health or couple/family therapy available in their communities during the period from baseline to 1-year follow-up. Overall, 22% of control respondents reported that they had received therapeutic or counseling services that were available in the community at some point in the year following baseline. Some participated in one type of outside service, while others participated in and reported multiple types of services. Specifically, of those in the control group who reported receiving outside services (n = 133), 79 participated in individual therapy, 87 participated in couples therapy, and 24 participated in family therapy.

**Limitations.** Our implementation findings indicate facilitators adhered to program content, participants received a majority of program sessions and were engaged with the material, and only a minority of program or control respondents received outside services from baseline to 1-year follow-up. A few limitations, however, should be noted. First, although facilitators reported general adherence to the Elevate and CCM curricula, we recognize these are subjective self-reports to a general question of fidelity and are not objective, detailed assessments of what exactly was and was not taught. Objective observations of fidelity are not included in this report because this implementation question was not part of the original evaluation design. Additionally, approximately one quarter of study participants were engaged with therapeutic services at some point from baseline to 1-year follow-up, which could have contributed to noted improvements at later timepoints. It is also noteworthy, however, that the proportion of individuals engaging in these additional community services were fairly balanced between control and program groups, lessening the likelihood that differences documented between

groups is attributed to program participants' use of therapeutic services and not HMRE. Another limitation related to the questions of use of community resources is that survey question options focused solely on individual mental health and couple and family therapy and not other family support services (e.g., job skills training, parent education, financial management classes, etc.). It is not known to what extent study participants were involved with these supports during the period from baseline to 1-year follow-up and their role in predicting the outcomes observed.

## B. Primary impact evaluation

### Key Findings:

**The results of the primary impact analyses indicate significant differences between program participants and the no-program control group:**

- ELEVATE participants experienced greater immediate improvements (2-month post-baseline) in couple relationship skills compared to the no-program control group.
- CCM participants experienced greater immediate improvements (2-month post-baseline) in couple relationship skills compared to the no-program control group.
- ELEVATE participants experienced greater improvements in mental health over 6 months compared to the no-program control group, who did not experience improvements (i.e., changes were not statistically significantly distinguishable from zero).
- There were no significant differences between CCM participants and the control group over 6 months for mental health.
- ELEVATE participants experienced greater improvements in couple satisfaction over 6 months compared to the no-program control group, who experienced statistically significant declines in couple satisfaction over 6 months.
- There were no significant differences between CCM participants and the control group over 6 months in couple satisfaction.

### 1. Methods used to estimate the program impact

As noted above, we used an ITT approach to assess the effectiveness of the ELEVATE and CCM programs. The assessment of program impacts was based on the conditions to which participants were originally randomly assigned regardless of whether participants actually completed programming.

Primary Impact RQs #1 and #2 examined the impact of ELEVATE and CCM, respectively, on immediate post-program change (at 2-month follow-up after baseline) in couple relationship skills, the primary area of skills-training for the HMRE intervention. We used SPSS statistical software to fit multi-level regression models to assess differences between groups at the 2-month follow-up time point. While simultaneously accounting for the nesting of individuals within

couples, these models predict the value of the outcome at the 2-month follow-up time point, as predicted by randomly-assigned study condition, while accounting for baseline levels of the outcome. In other words, these models predict future behaviors based on past behaviors and the outcome can be interpreted as residual change (Hyndman and Athanasopoulos, 2018). Findings were considered statistically significant based on $p < .05$, two-tailed test. We used Cohen's *d* effect size to evaluate the magnitude of the effect on the outcomes of interest. Technical details and the equation to calculate the effect size are presented in Appendix E.

Growth curve modeling (GCM) was used to test Primary Impact RQs #3–#6, which focused on the program impact over 6-months. This is a recommended method for randomized trials with multiple follow-up periods because GCM is a variant of mixed-effects models (also commonly referred to as random effects or random-regression) and is a more powerful technique than repeated-measures analysis of covariance (e.g., Schultz, Cowan and Cowan, 2006; Gueorguieva and Krystal, 2004; Amato, 2014). The results of a multilevel GCM provide information on differences at given time points between groups, as well as the program impact on rates of change, while accounting for the nesting of individuals within couples. Findings were considered statistically significant based on $p < .05$, two-tailed test. We note that these tests assume linear growth and do not test for non-linear or quadratic growth that could indicate delayed or diminished growth. Technical details and the equation to calculate the amount of variance explained by program participation can be found in Appendix E.

**PRIMARY Impact RQ #1.** Results of the regression model in which earlier time points for the same measure are accounted for in predicting later timepoints, indicated a significant difference between ELEVATE participants and the no-program control group in the average change in Couple Relationship Skills from baseline to 2-month follow-up ($B = 3.96$, $p < .001$). The Cohen's *d* effect size of ELEVATE on change in Couple Relationship Skills at 2-month follow-up was .19. Our analyses indicated that ELEVATE respondents improved, on average, approximately 7 points on the CRSI ($M = 6.91$; $SD = 17.48$; $t = 8.81$, $p < .001$), whereas control respondents improved, on average, approximately 2 points ($M = 2.39$; $SD = 15.69$; $t = 3.48$, $p = .001$). The range of the measure is based on a seven-point scale of 32 questions, which gives a composite sum score minimum of 7 and a maximum of 224.

**Table V.2a. Estimated effects of ELEVATE on change in couple relationship skills from baseline to 2-month follow-up; Primary Impact RQ #1**

| Outcome measure | ELEVATE mean change (*p*-value) | Control mean change (*p*-value) | ELEVATE compared to control mean change difference (*p*-value) | Cohen's d effect size |
|---|---|---|---|---|
| Couple relationship skills | 6.91 (< .001) | 2.39 (.001) | 3.96 (< .001) | .19 |

Note: This model includes relationship status as a covariate.

**PRIMARY Impact RQ #2.** Results of the regression model in which earlier time points for the same measure are accounted for in predicting later timepoints, indicated a statistically significant difference between CCM participants and the no-program control group in the average change in Couple Relationship Skills from baseline to 2-month follow-up ($B = 4.09$, $p < .001$). The

Cohen's *d* effect size of CCM on change in Couple Relationship Skills immediately following program participation was .18. Our analyses indicated that CCM respondents improved, on average, approximately 7 points on the CRSI ($M = 6.69$; $SD = 17.75$; $t = 8.60$, $p < .001$), whereas control respondents improved, on average, approximately 2 points ($M = 2.39$; $SD = 15.69$; $t = 3.48$, $p = .001$). The range of the measure is based on a seven-point scale of 32 questions, which gives a composite sum score minimum of 7 and a maximum of 224.

**Table V.2b. Estimated effects of CCM on change in couple relationship skills from baseline to 2-month follow-up; Primary Impact RQ #2**

| Outcome measure | CCM mean change (*p*-value) | Control mean change (*p*-value) | CCM compared to control mean change difference (*p*-value) | Cohen's d effect size |
|---|---|---|---|---|
| Couple relationship skills | 6.69 (< .001) | 2.39 (.001) | 4.09 (< .001) | .18 |

Note: This model includes relationship status as a covariate.

**PRIMARY Impact RQ #3.** Results of the GCM including baseline, 2-month follow-up, and 6-month follow-up revealed a significant difference in the positive growth trajectory between the ELEVATE participants and the no-program control group, indicating a positive treatment effect of ELEVATE for change in mental health over 6 months (B = .343, $p = .003$). On average, ELEVATE participants improved .343 points more on the SF-12 mental health index each month compared to the control group. The results of the GCM also indicated the control group did not experience statistically significant change (B = .037, $p = .668$). The adjusted $R^2$ of the effect of ELEVATE on change in mental health is .13; in other words, 13% of the variance associated with change in mental health was due to participation in ELEVATE. The variance associated with change in mental health converts to a Cohen's *d* effect size of .26 (Rosenthal, 1994).

**Table V.2c. Estimated effects of ELEVATE on change in mental health from baseline to 6-month follow-up; Primary Impact RQ #3**

| Outcome measure | ELEVATE mean slope (p-value) | Control mean slope (p-value) | ELEVATE compared to control mean difference (*p*-value) | Change in $R^2$ (Cohen's d) |
|---|---|---|---|---|
| Mental health | .380 (.001) | .037 (.668) | .343 (.003) | .13 (.26) |

**PRIMARY Impact RQ #4.** Results of the GCM including baseline, 2-month follow-up, and 6-month follow-up did not reveal a significant difference in the growth trajectory between the CCM participants and the no-program control group, indicating no treatment effect of CCM for change in mental health over 6 months (B = .172, $p = .152$).

**Table V.2d. Estimated effects of CCM on change in mental health from baseline to 6-month follow-up; Primary Impact RQ #4**

| Outcome measure | CCM mean slope (p-value) | Control mean slope (p-value) | CCM compared to control mean difference (*p*-value) |
|---|---|---|---|
| Mental health | .209 (.014) | .037(.668) | .172 (.152) |

**PRIMARY Impact RQ #5.** Results of the GCM including baseline, 2-month follow-up, and 6-month follow-up revealed a significant difference in the positive growth trajectory between ELEVATE participants and the no-program control group, indicating a positive treatment effect of ELEVATE on change in couple satisfaction over 6 months (B = .033, $p$ = .005). On average, ELEVATE participants improved .033 points more in their couple satisfaction score each month compared to the control group. The results of the GCM also indicated that the control group declined in couple satisfaction (B = -.017, $p$ = .032). The range of the measure is based on a six-point scale. The adjusted $R^2$ of the effect of ELEVATE on change in couple satisfaction is .10; in other words, 10% of the variance associated with change in couple satisfaction was due to participation in ELEVATE. The variance associated with change in couple satisfaction converts to a Cohen's $d$ effect size of .20 (Rosenthal, 1994).

**Table V.2e. Estimated effects of ELEVATE on change in couple satisfaction from baseline to 6-month follow-up; Primary Impact RQ #5**

| Outcome measure | ELEVATE mean slope (p-value) | Control mean slope (p-value) | ELEVATE compared to control mean difference (*p*-value) | Change in $R^2$ (Cohen's d) |
|---|---|---|---|---|
| Couple satisfaction | .016 (.072) | -.017 (.032) | .033 (.005) | .10 (.20) |

**PRIMARY Impact RQ #6.** Results of the GCM including baseline, 2-month follow-up, and 6-month follow-up did not reveal a significant difference in the growth trajectory between the CCM participants and the no-program control group, indicating no treatment effect of CCM for change in couple satisfaction over 6 months (B = .020 p = .101).

**Table V.2f. Estimated effects of CCM on change in couple satisfaction from baseline to 6-month follow-up; Primary Impact RQ #6**

| Outcome measure | CCM mean slope (p-value) | Control mean slope (p-value) | CCM compared to control mean difference (*p*-value) |
|---|---|---|---|
| Couple satisfaction | .002 (.792) | -.017 (.032) | .020 (.101) |

## C. Sensitivity analyses

This section presents findings from two types of sensitivity analyses conducted to check the robustness of the primary impact findings to alternative assumptions. Complete response cases were used to construct the analytic sample for the primary research questions, and this was somewhat relaxed in the sensitivity analyses. First, we tested models that were more *data inclusive* whereby sample members were selected by inclusion in random assignment, completion of at least two of the three surveys relevant to the Impact RQs #3–#6, and completion of at least one of the multi-item measures. Because item missingness/nonresponse was allowed in these models, mean scores were used instead of sum scores (except for the Mental Health measure because all items are required to create a composite) and full information maximum likelihood was used in the growth models. For Impact RQs 1 and 2, an additional 4 ELEVATE participants, 7 CCM participants, and 4 control respondents were retained with this method. For impact RQs 3 and 4, an additional 169 Elevate participants, 125 CCM participants, and 119

control respondents were retained with this method. For impact RQs 5 and 6, an additional 132 ELEVATE participants, 118 CCM participants, and 123 control respondents were retained with this method.

Second, we tested the models using a more advanced method of *multiple imputation* and calculated sum scores. Before imputation was conducted, we assessed item level and scale level missingness (4-7% at baseline; 13-15% at 2-month follow-up; and 21-24% at 6-month follow-up), as well as correlations between missingness and covariates. The levels of missingness were considered low and met the recommended guidelines for imputation (Newman, 2014). Specifically, we utilized multiple imputation by chained equations using classification and regression trees (CART) in the "mice" package of R 3.6.2 (Burgette and Reiter, 2010; van Buuren and Groothuis-Oudshoorn, 2011; R Core Team, 2019). This approach draws information from both categorical and continuous variables which are used as auxiliary variables to inform the algorithm. For impact RQs 1 and 2, an additional 129 ELEVATE participants, 106 CCM participants, and 85 control respondents were retained with this method. For impact RQs 3 and 4, an additional 237 ELEVATE participants, 203 CCM participants, and 190 control respondents were retained with this method. For impact RQs, 5 and 6 an additional 220 ELEVATE participants, 194 CCM participants, and 180 control respondents were retained with this method.

## Key Findings

**Sensitivity analyses, part 1.** Overall, similar results documenting robust program effects were found for the data-inclusive models.

- There were statistically significant differences between ELEVATE participants and control respondents and between CCM participants and control respondents in immediate post-program (2-month follow-up) improvements in relationship skills.
- There were significant differences between ELEVATE participants and control respondents over 6 months on mental health and couple satisfaction.
- There were no significant differences between CCM program participants and control respondents over 6 months on mental health and couple satisfaction.

**Primary Impact RQ #1.** Similar to the results presented in the key findings for the primary impact evaluation section (Section V.B), results of the regression model indicated a significant difference between ELEVATE participants and the no-program control group (B = .131, $p <$ .001), such that ELEVATE participants experienced greater changes than the no-program control group.

**Primary Impact RQ #2.** Similar to the results presented in the primary impact evaluation section (Section V. B), results of the regression model indicated a significant difference between CCM participants and the no-program control group (B = .127, $p < .001$), such that CCM participants experienced greater changes than the no-program control group.

**Primary Impact RQ #3.** Similar to the results presented in the primary impact evaluation section (Section V. B), results of the GCM revealed a significant difference in the growth trajectory between the ELEVATE and no-program control groups, indicating a positive treatment effect of ELEVATE for change in mental health over 6 months (B = .286, $p$ = .009). In other words, on average, ELEVATE participants improved .286 points more on the SF-12 mental health index each month compared to the control group. The GCM also indicated the control group did not experience change (B = .090, p = .258).

**Primary Impact RQ #4.** Similar to the results presented in the primary impact evaluation section (Section V. B), results of the GCM did not reveal a significant difference in the growth trajectory between the CCM and no-program control groups, indicating no treatment effect of CCM on mental health over 6 months (B = .172, $p$ = .131).

**Primary Impact RQ #5.** Similar to the results presented in the primary impact evaluation section (Section V. B), results of the GCM revealed a significant difference in the growth trajectory between the ELEVATE and no-program control group, indicating a positive treatment effect of ELEVATE for change in couple satisfaction over 6 months (B = .024, $p$ = .033). In other words, on average, ELEVATE participants improved .024 points more in their couple satisfaction score each month compared to the control group. The GCM also indicated the control group did not experience change (B = -.01, p = .21).

**Primary Impact RQ #6.** Similar to the results presented in the primary impact evaluation section (Section V. B), results of the GCM did not reveal a significant difference in growth trajectories between CCM participants and the no-program control groups, indicating no treatment effect of CCM for change in couple satisfaction over 6 months (B = .010, $p$ = .403).

## Key Findings

**Sensitivity analyses, part 2.** Overall, similar results documenting program effects were found for the models fit with multiply imputed data.

- There were significant differences between ELEVATE program participants and control respondents and between CCM program participants and control respondents in immediate post-program (2-month follow-up) improvements in relationship skills.
- There were significant differences between ELEVATE program participants and control respondents over 6 months on mental health and couple satisfaction.
- There were no significant differences between CCM program participants and control respondents over 6 months on mental health and couple satisfaction.

**Primary Impact RQ #1.** Similar to the results presented in the primary impact evaluation section (Section V. B), results of the regression model indicated a significant difference between ELEVATE participants and the no-program control group (B = 4.781, $p < .001$), such that ELEVATE participants experienced greater change in couple relationship skills compared to the no-program control group.

**Primary Impact RQ #2.** Similar to the results presented in the primary impact evaluation section (Section V. B), results of the regression model indicated a significant difference between CCM participants and the no-program control group (B = 3.589, $p = .001$), such that CCM participants experienced greater change in couple relationship skills compared to the no-program control group.

**Primary Impact RQ #3.** Similar to the results presented in the primary impact evaluation section (Section V. B), results of the GCM revealed a significant difference in the growth trajectory between the ELEVATE and no-program control groups, indicating a positive treatment effect of ELEVATE for change in mental health over 6 months (B = .419, $p < .001$). In other words, on average, ELEVATE participants improved .419 points more in their SF-12 mental health index each month compared to the control group who did not experience change.

**Primary Impact RQ #4.** Similar to the results presented in the primary impact evaluation section (Section V. B), results of the GCM did not reveal a significant difference in the growth trajectory between the CCM and no-program control groups, indicating no treatment effect of CCM for change in mental health over 6 months (B = .193, $p$ =.104).

**Primary Impact RQ #5.** Similar to the results presented in the primary impact evaluation section (Section V. B), results of the GCM revealed a significant difference in the growth trajectory between the ELEVATE and no-program control groups, indicating a positive treatment effect of ELEVATE on couple satisfaction over 6 months (B = .146, $p = .006$). In other words, on average, ELEVATE participants improved .146 points more in their couple satisfaction score each month compared to the control group who experienced declines.

**Primary Impact RQ #6.** Similar to the results presented in the primary impact evaluation section (Section V. B), results of the GCM did not reveal a significant difference in the growth trajectory between the CCM and no-program control groups, indicating no treatment effect of CCM on couple satisfaction over 6 months (B = .094 $p = .065$).

**Summary.** Overall, the sensitivity analyses support the results of the primary analyses, indicating the robustness of 2-month (for both ELEVATE and CCM) and 6-month (for ELEVATE) effects of program participation on couple relationship skills, mental health, and couple satisfaction to alternative methods of addressing missing values. Specifically, for Primary Impact RQs #1-6, the two sets of sensitivity analyses align with the results of the primary analyses.

**Table V.3. Differences in means between intervention and comparison groups estimated using alternative methods**

| Outcome | Benchmark approach (p value) | Data inclusive model (p value) | Multiple imputation model (p value) |
|---|---|---|---|
| Couple relationship skills (ELEVATE) | 3.96 (< .001) | .131 (< .001) | 4.781 (< .001) |
| Couple relationship skills (CCM) | 4.09 (< .001) | .127 (< .001) | 3.589 (.001) |
| Mental health (ELEVATE) | .343 (.003) | .286 (.009) | .419 (< .001) |
| Mental health (CCM) | .172 (.152) | .172 (.131) | .193 (.104) |
| Couple satisfaction (ELEVATE) | .033 (.005) | .024 (.033) | .146 (.006) |
| Couple satisfaction (CCM) | .020 (.101) | .010 (.403) | .094 (.065) |

Note: Sum scores were used for couple relationship skills for the benchmark approach, the multiple imputation model, and the Bayesian model.

## D. Additional analyses

In addition to the Primary Impact RQs central to the local impact evaluation, there were several other important questions we explored in order to more fully understand the experience of participation in the Elevate and CCM programs at our sites. To address Additional Impact RQs #1-4, assessing more far-reaching (1-year) program effects on mental health and couple satisfaction, we employed the same multi-level modeling methods described above. Again, we note that these growth curve models assume linear growth and do not test for non-linear or quadratic growth that could indicate delayed or diminished growth. These models also were fit with the multiple imputation data (see sensitivity analyses, Section V.C). To address Additional Impact RQs #5 and 6, whether changes immediately following program participation (i.e., 8 weeks after baseline) in couple relationship skills predict couple satisfaction at 6-month follow-up, accounting for baseline level, a path model (See Appendix G, Figure 1) was fit to test the hypothesized change on change pathway derived from our logic model. Specifically, our logic model (See Appendix A) posits that the main target of the program, relationship skills, will improve immediately after program participation (i.e., 2-month follow-up), and that those changes will lead to longer term (6 months to 1 year) improvements in couple relationship satisfaction. Structural equation modeling (SEM) techniques allow for the test of the hypothesis that immediate changes are associated with prospective changes, or specifically for this study changes over 6 months. Goodness of fit indices were assessed to establish model fit.

### Key Findings

**Additional Impact RQ #1.** ELEVATE participants experienced greater improvements in mental health over 1 year compared to the no-program control group, who did not experience improvements.

Results of the GCM revealed a significant difference in the growth trajectory between the ELEVATE participants and the no-program control group, indicating a positive treatment effect of ELEVATE for change in mental health over 1 year (B = .138, $p$ = .008). On average, ELEVATE participants improved .138 points more in their mental health score each month over 1 year compared to the control group. The GCM also indicated the control group did not experience significant growth (B = .004, $p$ = .905).

## Key Findings

**Additional Impact RQ #2.** CCM participants did not experience improvements in mental health over 1 year compared to the no-program control group.

Results of the GCM revealed no significant difference in the growth trajectory between the CCM participants and the no-program control group, indicating no observed treatment effect of CCM for change in mental health over 1 year (B = .083, $p$ = .112). In other words, on average, CCM participants did not improve in mental health over 1 year more or less than the control group, who, based on the GCM results, did not experience change (B = .004, $p$ = .905).

## Key Findings

**Additional Impact RQ #3.** ELEVATE participants experienced greater improvements in couple satisfaction over 1 year compared to the no-program control group, who declined in couple satisfaction.

Results of the GCM revealed a significant difference in growth trajectories between ELEVATE participants and the no-program control group, indicating a positive treatment effect of ELEVATE on couple satisfaction over 1 year (B = .085, $p$ < .001). In other words, on average, ELEVATE participants improved .086 points more in their couple satisfaction score each month over 1 year compared to the control group. The GCM results also indicated the control group statistically significantly declined over time (B = -.054, $p$ = .001).

## Key Findings

**Additional Impact RQ #4.** CCM participants experienced greater improvements in couple satisfaction over 1 year compared to the no-program control group, who declined in couple satisfaction.

Results of the GCM revealed a significant difference in the growth trajectory between CCM participants and the no-program control group, indicating a desired treatment effect of CCM on

couple satisfaction over 1 year (B = .076, $p < .001$). In other words, on average, CCM participants improved .076 points more in their couple satisfaction score each month compared to the control group. The GCM results also indicated the control group statistically significantly declined over time (B = -.054, $p = .001$).

## Key Findings

**Additional Impact RQ #5.** Greater changes in couple relationship skills immediately following ELEVATE participation predicted greater long-term changes in couple satisfaction 6 months after program participation.

The results of the path analysis indicated greater immediate change in couple relationship skills after ELEVATE participation was associated with greater change in couple satisfaction over 6 months ($\beta = .250$, $p < .001$). Appendix G provides information about goodness of fit indices.

## Key Findings

**Additional Impact RQ #6.** Greater changes in couple relationship skills immediately following CCM participation predicted greater long-term changes in couple satisfaction 6 months after program participation.

The results of the path analysis indicated greater immediate change in couple relationship skills after CCM participation was associated with greater change in couple satisfaction over 6 months ($\beta = .205$, $p < .001$). Appendix G provides information about goodness of fit indices.

# VI. DISCUSSION

Overall, the AHMREI implemented the ELEVATE and CCM curricula effectively and demonstrated evidence for immediate and long-term positive program impacts on couple functioning, including, in the case of ELEVATE, sustained impacts on individual mental health. This efficacy study demonstrates that AHMREI programming functions as a community-level support (see transactional ecological family systems model; Bronfenbrenner, 1977) and may promote resilience among economically and educationally diverse couples (Patterson, 2002). Impact findings indicate that both curricula serve as longer-term protective factors, improving couples' relationship functioning and, for ELEVATE participants, the benefits extended to their reported individual well-being.

The AHMREI successfully recruited a large, economically and educationally diverse sample of predominantly White and Black couples and effectively engaged those randomly assigned to relationship education programming. Immediate treatment effects on couple relationship skills were evidenced for both ELEVATE and CCM participants, and ELEVATE couple participants reported significant program impacts on mental health and couple satisfaction over 6 months and up to 1 year, compared with no-program control respondents. Although CCM treatment effects on mental health and couple satisfaction were not evident at 6 months, the longer-term effects of CCM participation emerged with significant program impact on couple satisfaction at the 1-year follow-up period, compared with no-program control respondents. The ELEVATE program demonstrated small-to-moderate effect sizes ($|d|$ = .19 to .26; Cohen, 1988) on relationship skills, mental health, and couple satisfaction up to the 6-month follow-up period, whereas treatment effect sizes for CCM were smaller ($|d|$ = .12 to .18). These effect sizes among a more economically diverse sample of couples are notable, considering that Arnold and Beelmann's (2019) meta-analysis of 48 independent studies of HMRE found effects were generally lower in studies with primarily lower- income, less-educated, younger, and unmarried couples and Hawkins and Erickson's (2015) earlier meta-analysis of HMRE among low-income couples (38 studies with 47 independent samples) found a very small average effect size ($d$ = .06) for more rigorous, long-term RCT studies (similar to the present study).

Assessing implementation elements (Ballard, 2020; Stanley et al., 2019) is key for understanding the benefits of HMRE participation. Although we did not formally link implementation elements to outcomes and test their connection in this report, we can note the parallel implementation analyses findings and hypothesize their connection to the positive program impacts that were detected. We found that AHMREI's community facilitators reported fairly high fidelity to the relationship education curricula, established strong bonds with program participants, promoted engagement with the course content, and were viewed by participants as very knowledgeable and highly effective in their teaching efforts. Recent reviews of best practices in CRE research cite effective implementation strategies as central to program effectiveness (Stanley et al., 2019), even regarding implementation outcomes to be of equal importance to program outcomes (Ballard, 2020). AHMREI staff reported recruitment, engagement, and retention strategies similar to other recent published studies (Liu et al., 2020; Doss et al., 2020; Roddy et al., 2020a;

2020b). The AHMREI staff carefully developed the curricula to be applicable to a broad spectrum of couples, ensuring that examples and illustrations included experiences of racially and economically diverse couples. In addition, facilitators attended rigorous curriculum training sessions with ongoing feedback from supervisors throughout the study duration, in line with best practices (Ballard, 2020; Stanley et al., 2019). Adherence to curriculum content and main teaching points and skills practices were emphasized in trainings. The reports of curriculum fidelity are likely linked with the evidence of immediate program effects on the core couple relationship skills that are the focus of the curricula.

Program attendance also likely explains the positive program effects found. Previous research notes that attendance to half of the HMRE sessions or 6-12 hours is considered an "appropriate" or "adequate" dose (Arnold and Beelmann, 2019). Considerable effort was made by program staff at each site to engage study participants assigned to program groups, and follow-ups through emails and texts as well as the provision of program supports (e.g., childcare, meals) to enhance retention in the program. The AHMREI's overall average attendance rate was 62% among those assigned to the program groups, which meets the "half or more" threshold and is notable since the sample includes couples who never attended one session (i.e., 25% for ELEVATE, 15% for CCM). These no-show rates are much lower than the Building Strong Families (BSF) evaluation of HMRE (Dion et al., 2010; Wood et al., 2014) and comparable to other large-scale RCT impact studies among economically and educationally diverse samples; Lundquist et al., 2014; Moore et al., 2018; Stanley et al., 2014).

The current study offers a unique feature because it tested the parallel efficacy of two newer, evidence-informed curricula using a rigorous RCT design. As such, we offer information to consider both ELEVATE and CCM as evidence-based curricula, given the results for immediate program impacts on couple relationship skills after program participation and enhanced couple satisfaction compared to control group participants at the one-year mark.

## A. Comparing ELEVATE and CCM program effects

We note, however, the unique pattern of treatment effects that emerged for each curriculum at the 6 month and one year marks. Our planned analyses for this report did not include statistical tests of differences between the three groups. As such, we first offer general observations and suggestions regarding the differences noted in the analyses of each curriculum relative to the control condition. We then provide results from some initial explorations of differences between the CCM and the ELEVATE groups. We expect that future analyses using more complex comparative models that include all three groups will further elucidate any actual differences that may exist, as well as variables that may explain them.

Immediate post-program impacts were similar for ELEVATE and CCM, such that couples assigned to either program, compared with the control group, reported gains at 2-month follow-up for a variety of couple relationship skills which were directly related to curriculum content and skills-based exercises (e.g., commitment, intimate partner knowledge, loving-kindness, conflict management). Our logic model posited that enhancements in the common couple skills

that are targeted by both CCM and ELEVATE following participation would be evident, and this expectation was supported. However, when examining longer-term impacts, results differed for each curriculum group in comparison to the control condition. Assignment to the ELEVATE condition predicted intermediate and long-term (6-months and 1-year) improvements in mental health (i.e., anxiety/depression, sense of wellbeing, and social/emotional functioning) as well as reports of couple satisfaction, compared with no change in mental health and declines in couple satisfaction for no-program control respondents. This confirms McGill, Adler-Baeder, and Garneau's (in press) quasi-experimental study validating the ELEVATE curriculum for effecting changes in relationship skills and couple relationship quality, and extends evidence of ELEVATE's impacts to the individual functioning domain. A few prior studies have linked HMRE programs with immediate improvements in depression or anxiety symptoms (Adler-Baeder et al., 2010; Bradford et al., 2014); however, limited evidence exists for RCT studies of HMRE that find sustained benefits in reports of mental health (see Hsueh et al., 2012 for exception). This finding of the mental health benefits of the ELEVATE curriculum is a potentially novel contribution to the field of HMRE research, particularly given the large, diverse sample used in the current study.

CCM treatment effects were evident in this sample only for couple functioning outcomes (i.e., relationship skills at 2-month follow-up and couple satisfaction at 1-year follow-up). This pattern is consistent with a "late-blooming" model of effects with more gradual, long-term impacts and demonstrates the value of continuing to collect follow-up data beyond the immediate post-program time point. Meta-analyses have found that the majority of studies of HMRE collect follow-up data only through the 4-6 month period after program participation (Hawkins et al., 2008). Overall, the longevity of impacts on couple satisfaction for both curricula in the present study is noteworthy. Sensitivity analyses replicated and reinforced the findings. Positive longer-term effects on couple functioning is in line with a handful of studies finding HMRE program participation predicting maintained, stable effects on relationship functioning up to one year later (see Arnold and Beelmann, 2019; Halford and Bodenmann, 2013 for reviews). Other studies, however, have evidenced declining long-term trajectories (Gubits et al., 2014; Wood et al., 2014). With mixed results for efficacy studies of HMRE, a valuable next step would be discerning similarities and differences in contextual and implementation variables that may explain these outcomes (e.g., program design, curriculum content, participant characteristics, attendance patterns, facilitator characteristics, etc.). Our results can speak only for the experiences in these curricula, with this sample, delivered by these facilitators and do not speak for the experiences of HMRE participation as a whole.

Within our study, we note different patterns of change between the two program groups and we can only speculate as to why program effects on couple satisfaction emerged for CCM participants at the 1-year mark and not prior. While there is considerable overlap in the content of the two curricula, since each is based on the NERMEM core concepts for successful couple relationships (Futris and Adler-Baeder, 2013), CCM has a heavier emphasis on self-care, mindfulness practices and awareness, and the value of these skills in the relational dynamic. Other research on mindfulness-based interventions finds that it is the practicing of these skills

over time that leads to other benefits in physical and relational health (see meta-analysis by Parsons et al., 2017). Benefits are enhanced for individuals who develop a "habit" of mindful practices. The development of these mindful practices could explain the delayed effects on couple functioning.

It is somewhat surprising, however, that program effects on mental health were not documented at either of the follow-up timepoints for those assigned to the CCM condition given its emphasis on self-care. While these effects may emerge at a later timepoint, it would be valuable to explore in future research other contextual variables that may explain this finding up through the one-year follow-up. Understanding to what extent CCM participants continued to use and practice the skills taught in the class after program completion would be helpful. Participants in this RCT were not aware that one of the program options had a heavy emphasis on mindfulness practices and may not have been as open to these as study participants might be if they were informed of the specifics of a program before consenting to participate. While efficacy trials rely on random assignment and baseline similarities of participants, it would also be valuable to assess program effects for couples who select into a mindfulness-based HMRE program, indicating they already have "buy in" for this approach.

As an initial exploration of comparisons between the program groups, we estimated GCMs comparing the patterns of change between the CCM and ELEVATE groups. We found that slopes for the two curriculum groups did not differ significantly for mental health (B = .171, $p$ = .137) or couple satisfaction (B = .013, $p$ = .285) over 6 months or over one year (B = .055, $p$ = .071; B = .009, $p$ = .125). Further, we found that the CCM group's rate of growth in mental health was statistically significant (B=.209; $p$ = .014), while the GCM testing for group differences between the control and the CCM groups indicated no significant growth for the control group. We also tested whether the CCM and the ELEVATE groups' mean level of mental health and couple satisfaction at baseline differed since a higher start-point for the CCM group could also explain more limited growth in comparison to the control group. The two program groups did not differ at baseline on these outcomes. These results along with the finding that differences between the control group and the CCM group in mental health improvement at 6 months and one year and in couple satisfaction at 6 months approached significance (~ .10-.15) suggest both similarity between the CCM and the ELEVATE groups and that the magnitude of the differences between CCM and the control group was insufficient to reach a level of significance. As GCM relies on estimating the "average" experience of the group in comparison to the other, it is likely that there is more variation in the group experience in the CCM group. Based on current analyses, it is still too preliminary to assert that ELEVATE or CCM was "more effective" than the other. We plan in future analyses to utilize mixture modeling within both program groups to better understand profiles of participants who changed over time in distinct ways. These types of analyses can be very useful for practitioners who are interested in finding a "best match" for participant couples and program design.

## B. Testing change on change

The current study also explicitly tested the "change on change" assumption evident in the AHMREI logic model (Appendix A), that the change expected for ELEVATE and CCM participation in couple relationship skills (immediate effect) leads to the change in couple relationship satisfaction, a longer-term effect of HMRE participation. This assumption was substantiated for both ELEVATE and CCM participants. Greater immediate post-program (i.e., 2-month) improvements in couple relationship skills (explicitly linked with curriculum content) predicted greater long-term improvements in couple satisfaction 6 months after program participation. This "spillover effect" also supports the basis for the NERMEM theoretical framework proposing that improvements in the core relational skills emphasized in the NERMEM will result in better relationship quality (Futris et al., 2015). While this connection between outcomes is a common assumption in HMRE work, these outcomes are typically assessed concurrently. This study is one of the first to test the relationship between outcomes prospectively and provides another novel contribution to our understanding of specific HMRE programming and pathways of relationship growth among economically and educationally diverse couples.

## C. Sample characteristics and generalizability

The AHMREI recruited an economically and educationally diverse sample of couples, racially representative of the general population of Alabama (U.S. Census Bureau, 2019). Nearly 40% were non-White couples (i.e., 32% Black/African American couples); nearly half reported a combined household income of less than $40,000 a year; another 30% reported a combined household income of $40,000 to $75,000; over half the sample had less than a 4-year college degree; and about a third were unmarried. These characteristics render a significant portion of our sample more vulnerable couples due to economic instability and the ongoing effects of systemic racism. A coordinated system of individual, community and structural supports are critical for promoting resiliency and success in the face of risks and challenges for these individuals and families. Our findings offer some promising evidence of sustained, positive impact of HMRE participation for a diverse sample of couples, suggesting that offering these programs may be a valuable additional resource for a broad spectrum of couples that can also serve as part of a system of supports for historically underserved couples in Alabama.

## D. Factors influencing recruitment and retention success

The study can be considered successful in meeting target number goals for recruitment and retention, and we offer information that may be useful for replication. The sample of 930 couples enrolled in the study and randomly assigned to program or control groups exceeded the target number in our original evaluation plan of 750 couples for random assignment and far exceeded the number of couples suggested by initial power analyses ($N = 375$ couples). This number was based on power analysis calculations for the number required to detect small to moderate effect sizes ($d \sim .25$) when modeling for 3 groups at 5 time points, assuming 10% attrition at each time point and a .50 correlation between measures over time. Notably, this assessment is an over-

estimate for the sample number needed for analyses in the current report, since 2-group comparisons were made and a maximum of 4 time points were utilized. The post-hoc power analyses using G*Power3 demonstrated that the models were fully powered (100%; Faul et al., 2007), indicating we had more than adequate power (80%) to have conducted up to three repeated measures across two groups with samples ranging from 805 to 1,041 individuals (based on each RQ in the current report). Given experiences in previous years that approximately 20-30% of couples contacting sites with interest in attending programming will drop out for a variety of reasons prior to program start, enrollment targets for each cohort at each site were 130%. This over-enrollment strategy best explains the resulting success of enrolling and randomly assigning at baseline 124% of the target number of couples.

Retention in the study at each follow-up time point is considered successful as well, given that 10% attrition was expected at each subsequent timepoint (i.e., 90% retention at 2-month post-baseline follow-up; 80% retention at 6-month post-baseline follow-up and 70% retention at one-year post-baseline follow-up), based on guidelines for survey research with community samples (Dillman, 2007). High retention rates suggest participants felt invested in our study and that our engagement efforts were effective. These strategies included: 1) after enrollment in the study, participants received an email from the evaluation team reminding them of the time frame for receiving future survey links via email and compensation levels; 2) emailing participants the links to complete baseline and follow-up surveys electronically, as well as providing computer access for participants, if needed; 3) sending periodic reminder emails after the survey links became available until the final closing date; 4) in addition to the evaluation team's reminders, having the community educators contact their program and non-program participants to confirm they received their surveys and to answer any questions about the survey completion process; 5) sending monthly "stay involved" emails, as well as birthday and anniversary celebratory emails; and 6) providing appropriate compensation in a timely manner for completing the study surveys.

The positive program effects might be partially attributable to reasonably high attendance rates in the sample, although this hypothesis remains to be empirically tested. As noted in the implementation findings (Section V.A.), two-thirds (66%) of ELEVATE participants and almost three-quarters (73%) of CCM participants attended half or more of the program sessions. Some participants assigned to the program condition did not attend any sessions (i.e., 25% for ELEVATE, 15% for CCM), which is a much lower no-show rate than the BSF study (Dion et al., 2010; Wood et al., 2014) and comparable to other RCT studies of economically and educationally diverse participants in HMRE (Lundquist et al., 2014; Moore et al., 2018; Stanley et al., 2014). The AHMREI used a block random assignment approach to ensure equal numbers in ELEVATE, CCM, and control conditions at each partner site; however, no data were collected on reasons for non-attendance. Therefore, we have no explanation for the higher no-show rate for ELEVATE participants.

In Arnold and Beelmann's (2019) recent meta-analysis of HMRE evaluations among economically strained samples, only half of the studies reported average attendance rates that exceeded 50%. Stronger effect sizes were found for these studies. Thus, low attendance rates have implications for demonstrating program effects since the "intent to treat" design of efficacy

studies includes the data from participants assigned to the program group regardless of their attendance record. Despite the numbers of no-show participants, the AHMREI's overall average attendance rate was 62% and can be considered a better than average attendance rate for community-based CRE studies using randomly assigned program participants, particularly since a portion were more economically strained. Over 40% of the sample was at or below the poverty line for a family of four; another 30% had a combined household income of $40,000 - $75,000, which can be considered "working class" (U.S. Census Bureau, 2018). These subgroups of participants may have added barriers, to program attendance.

Anticipating that lower-income couples experience a host of barriers to attending HMRE (e.g., work/school conflicts, transportation, childcare issues; Williamson, Karney and Bradbury, 2019), the AHMREI utilized a range of strategies to remove barriers and to better ensure that the program effects analyses would reflect actual attendance experiences. These strategies included providing meals and childcare at each session and offering transportation to/from classes or vouchers to cover public transportation or gas. Further, staff invested considerable effort in engaging with and checking in with couples in between sessions to remind them of upcoming sessions and to discern if there were any specific barriers to attendance at the next session that could be addressed. Staff used thank-you emails just after attending sessions, calls to participants who did not attend to gather information, and reminder emails just before sessions. Sites also were consistent with the specific person who was the participant's point of contact so that a relationship could be developed. The relationship between facilitator and participants was further bolstered by the emphasis among AHMREI staff on the use of experienced, well-trained facilitator teams of men and women. Previous demonstration project experiences found that male/female facilitator teams were the most well-received by participants. Similarity in race to at least one facilitator can also be helpful, although the quality of facilitation skills is the most potent predictor of positive change (Bradford, Adler-Baeder, et al., 2012).

## E. Limitations

The robust evidence for treatment effects on couple functioning for both ELEVATE and CCM participants and on mental health for ELEVATE participants should be considered in light of some study limitations. Measures were limited to self-report, which represent the participant's subjective perspective of skills and individual/relationship functioning. Future studies with multiple informants and methods (e.g., observation, partner report, physiological assessments, etc.) would provide more objective assessments of change over time. In addition, approximately a quarter of program participants reported use of outside support services from baseline to 1-year follow-up. Although a minority of the participants within the study groups accessed outside resources, and the proportions are similar across program and control groups, it limits our ability to assert that the differences in outcomes found between groups is owing solely to ELEVATE and CCM program effects. This will remain a limitation in future community-based studies of HMRE since limiting access to available community resources to isolate the HMRE experience would involve ethical issues and likely not meet conditions for protection of human subjects' rights in research. Further, questions regarding accessing other resources from baseline to 1-year

follow-up were limited to therapeutic services. It is not known whether and to what extent study participants took advantage of other types of community resources or informal social support networks that may also influence the outcomes assessed. Finally, there was a proportion of program participants (25% of the ELEVATE group and 15% of the CCM group) who did not attend any classes offered to them. Future analyses utilizing a treatment-on-treated (TOT) approach conducted with this data could assess the effect of dosage to investigate the effect of attendance on changes in couple and individual skills and well-being. The TOT approach may indicate whether the magnitude of effects is larger for those who engage in the program more frequently.

## F. Future directions

The current report provides considerable fuel for a range of future research. A next step for analyzing the data collected for this study can involve more focus on comparing the two program experiences. In additional to straightforward comparisons between groups on growth trajectories of change in multiple outcome areas, it would be interesting to understand better the varying patterns of change within program groups. Mixture modeling procedures can uncover distinct patterns of growth and discriminant class analyses procedures can help identify profiles of individuals and couples who benefit more or less in each program group. Results of these types of analyses will provide program designers and facilitators with valuable information for recruitment of participants to a program experience that may best serve their needs and interests.

Additionally, with minimal prior attention to the assessment of dyadic effects of HMRE programming (for recent exceptions see Carlson et al., 2019; Liu et al., 2020), future research can utilize actor-partner interdependence models to better understand individual and partner influences on outcomes over time. It will also be valuable to create and assess couple level variables using both partners' reports in common factor models. Further, long-term studies with multiple follow-up points could use dyadic latent profile analyses to examine differential trajectories of responses to HMRE as a couple, based on demographic, social, or relationship-functioning factors (see Roberson et al., 2020 for an example).

In addition, it would be valuable to understand better the potential program effects on a broader range of individual, relational, and family outcomes. We included a measure of mental health and found benefits for ELEVATE participants. Use of a family systems ecological perspective expects that benefits to relational health have implications for many other aspects of health and development (e.g., physical, psychological, economic) for individuals in the family system. There have been limited efforts in HMRE studies to document these potential benefits, particularly in the realm of physical health using objective assessments. Decades of research find links between relational health and physical health and longevity (see reviews by Kiecolt-Glaser and Wilson, 2017; Robles, 2014).

Further, our analyses utilized a rigorous RCT design within an ITT framework, which uses all couples randomly assigned for analyses, regardless of attendance or dosage rates for those assigned to program groups, and thus the program impacts found for ELEVATE (on individual

and couple functioning outcomes) and CCM (on couple functioning outcomes) are all the more noteworthy. Future research can incorporate the use of TOT analyses and discern patterns and levels of program effects based on dosage levels (Hawkins et al., 2012). While it may be that more motivated individuals have higher attendance rates and may be more likely to report benefits, it is also plausible that more time spent in programs and services provides more opportunities to absorb information, practice skills, and observe models, as described by social learning (Bandura, 1977) and experiential learning theories (Kolb, 1984).

Future HMRE work will also benefit from more studies of implementation processes and their influence on program experiences and program effects (Ballard, 2020; Stanley et al., 2019). It is likely that program dosage, participants' engagement with the facilitator, perception of the group climate, and judgment of the facilitator's quality and competence may act as mediators or moderators of program efficacy. In addition, characteristics of the program delivery model (e.g., time of day; duration; interval between sessions; group size, etc.), content of the curriculum, and characteristics of the couples and families influence program efficacy. Further, there is minimal research directly assessing facilitator effects on treatment outcomes (for exceptions see Bradford et al., 2012; Higginbotham and Myler, 2010; Ketring et al., 2017). Related clinical research suggests that participant-provider engagement and alliance contribute substantially to improved functioning (see meta-analysis by Friedlander, Escudero, Welmers-van de Poll, and Heatherington, 2018). Indeed, interviews with low-income, disadvantaged HMRE participants reveal they highly value their relationships and connections with facilitators (Halpern-Meekin, 2019), and thus HMRE implementation sites may be a vital factor in expanding disadvantaged couples' support systems and mitigating social poverty (see Carlson et al., 2019 for related research). Research on these implementation elements will provide valuable information for the refinement of models of best practices for diverse audiences.

## G. Conclusion

The current study is situated in a foundational body of HMRE research among economically diverse couples, the development of which is credited primarily to ACF funding support over the last two decades for demonstration projects and efficacy studies. Adding to this body of research, this AHMREI efficacy trial evidenced robust, enduring program impacts, likely due in part to its extensive efforts to embed in communities, build support, and engage participants across the study period. Both the ELEVATE and CCM curriculum demonstrated significant program impacts on relationship functioning, with immediate post-program (i.e., 2-month) improvements in a collection of key couple relationship skills, and long-lasting (up to one year) positive effects on couple satisfaction. Of note, CCM's impacts on couple satisfaction were not evident in the earlier assessment (at six months) and emerged at one year post-baseline. Further, ELEVATE participants reported additional program impacts on individual functioning (i.e., improved mental health) at the 6 month and one-year marks. Assessment of the links between outcomes validates the assumption that the couple relationship skills targeted for enhancement through the ELEVATE and CCM curricula are linked to later improvements in couple satisfaction. Importantly, these treatment effects on individual and relationship functioning were evidenced in

a large, economically and educationally diverse sample of White and Black married and unmarried couples, reflecting the population of Alabama. The representative nature of our sample provides some confidence that similar populations in Alabama and other areas will benefit from these curricula in similar ways. The successful conduct of this study and the reporting of these key results are an important milestone in the evolution of the AHMREI. A decade and a half of support for establishing a strong social infrastructure between Auburn University and state and local partners, building capacity among agency staff to deliver HMRE to broad populations of couples, and fine-tuning recruitment and retention strategies provided a solid foundation for this work. The AHMREI team is dedicated to continued growth and learning and further development of the work in pursuit of the goal of equalizing and expanding the opportunity to strengthen marriages and relationships and promote individual, family, and community health and well-being.

# VII. REFERENCES


Adler-Baeder, F., Bradford, A., Skuban, E., Lucier-Greer, M., Ketring, S., and Smith, T., (2010). Demographic predictors of relationship and marriage education participants' pre- and post- program relational and individual functioning. *Journal of Relationship and Couple Therapy: Innovations in Clinical and Educational Interventions, 9*, 113-132. https://doi.org/10.1080/15332691003694885.

Adler-Baeder, F., Calligas, A., Skuban, E., Keiley, M., Ketring, S. and Smith, T. (2013). Linking changes in couple functioning and parenting among couple relationship education participants. *Family Relations, 62,* 284-297. https://doi-org.spot.lib.auburn.edu/10.1111/fare.12006.

Adler-Baeder, F., Garneau, C., Vaughn, B., McGill, J., Harcourt, K. T., Ketring, S., and Smith, T. (2018). The effects of mother participation in relationship education on coparenting, parenting, and child social competence: Modeling spillover effects for low-income minority preschool children. *Family Process*, *57*(1), 113–130. https://doi.org/10.1111/famp.12267.

Adler-Baeder, F., McGill, J., Dede Yildirim, E., Futris, T. G., and Richardson, E. W. (2019, November). *The Couple Relationship Skills Inventory (CRSI): Validating a new measure*. Poster presented at the National Council on Family Relations Annual Conference, Dallas, TX.

Adler-Baeder, F., and Shirer, K. (2011). Coparenting interventions for nonmarried parents. In J. McHale and K. Lindahl (Eds.), *Coparenting: Theory, research, and clinical applications.* New York, NY: American Psychological Association.

Amato, P. R. (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*(3), 343–355. https://doi.org/10.1111/fare.12069.

Arnold, L. S., and Beelmann, A. (2019). The effects of relationship education in low-income couples: A meta-analysis of randomized-controlled evaluation studies. *Family Relations: An Interdisciplinary Journal of Applied Family Studies*. https://doi-org.spot.lib.auburn.edu/10.1111/fare.12325.

Asparouhov, T. and Muthén, B. (2019). *Advances in Bayesian model fit evaluation for 465 structural equation models.* http://www.statmodel.com/download/BayesFit.pdf.

Ballard, S. M. (2020). The Practice of Family Life Education: Toward an Implementation Framework. *Family Relations*.

Bandura, A. M. (1977). *Social learning theory.* Englewood Cliffs, NJ: Prentice Hall.

Bradford, A. B., Adler-Baeder, F., Ketring, S. A., Bub, K. L., Pittman, T. A. (2014). Relationship quality and depressed affect among a diverse sample of relationally unstable relationship education participants. *Family Relations, 63(2),* 219-231.

Bradford, A. B., Adler-Baeder, F., Ketring, S. A., and Smith, T. A. (2012). The role of participant-facilitator demographic match in couple and relationship education. *Family Relations*, *61*(1), 51–64. https://doi.org/10.1111/j.1741-3729.2011.00679.x.

Bradford, A. B., Drean, L., Adler-Baeder, F., Ketring, S. A., and Smith, T. A. (2017). It's About time! Examining received dosage and program duration as predictors of change among non-distressed and distressed married couple and relationship education participants. *Journal of Marital and Family Therapy*, *43*(3), 391–409. https://doi.org/10.1111/jmft.12209.

Broderick, C. B. (1993). *Understanding family process: Basics of family systems theory*. Sage.

Bronfenbrenner, U. (1977). Toward an experimental ecology of human development. *American Psychologist*, *32*(7), 513–531. https://doi.org/10.1037/0003-066X.32.7.513.

Bureau of Labor Statistics (2014). Regional and state unemployment – 2014 annual averages. Retrieved from https://www.bls.gov/news.release/archives/srgune_03042015.pdf.

Burgette, L. F., and Reiter, J. P. (2010). Multiple imputation for missing data via sequential regression trees. *American Journal of Epidemiology*, *172*(9), 1070–1076. https://doi.org/10.1093/aje/kwq260.

Burke, L. K., McGill, J., and Adler-Baeder, F. (2019). Exploring the links between facets of mindfulness and parenting efficacy and stress. *Journal of Child and Family Studies*. https://doi.org/10.1007/s10826-019-01692-2.

Cain, M. K. and Zhang, Z. (2019). Fit for a Bayesian: An evaluation of PPP and DIC for structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal, 26*(1), 39-50.

Carlson, R. G., Wheeler, N. J., Lui, X., Hipp, C., and Daire, A. P. (2019). The Relationship Between Social Support and Family Relationships Among Low-Income Couples Attending Relationship Education. *Family Process*. doi:10.1111/famp.12499.

Carson, J. W., Carson, K. M., Gil, K. M., and Baucom, D. H. (2004). Mindfulness-based relationship enhancement. *Behavior Therapy, 35,* 471–494. https://doi.org/10.1016/S0005-7894(04)80028-5.

Carson, J. W., Carson, K. M., Gil, K. M., and Baucom, D. H. (2007). Self-expansion as a mediator of relationship improvements in a mindfulness intervention. *Journal of Marital and Family Therapy, 33,* 517–528. https://doi.org/10.1111/j.1752-0606.2007.00035.x.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum.

Cowan, P. A., Cowan, C. P., and Knox, V. (2010). Marriage and fatherhood programs. *The Future of Children*, 205-230. https://doi.org/10.1353/foc.2010.0000.

DeMaria, R. M. (2005). Distressed couples and marriage education. *Family Relations*, *54*(2), 242–253. https://doi.org/10.1111/j.0197-6664.2005.00019.x.

Dillman, D.A. (2007). *Mail and Internet Surveys: The Tailored Design Method, second ed.* John Wiley Co.: Hoboken, New Jersey.

Doss, B. D., Knopp, K., Roddy, M. K., Rothman, K., Hatch, S. G., and 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*(4), 283–294.

Faul, F., Erdfelder, E., Lang, A. G., and Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175-191.

FFCA Policy Brief (2015). *Improving and enhancing Fatherhood and Marriage/Relationship Education*. Retrieved from https://fathersandfamiliescoalition.org/policy-brief-improving-and-enhancing-fatherhood-and-marriage-relationship-education.

Friedlander, M. L., Escudero, V., Welmers-van de Poll, M. J., and Heatherington, L. (2018). Meta-analysis of the alliance–outcome relation in couple and family therapy. *Psychotherapy, 55*(4), 356–371. https://doi-org.spot.lib.auburn.edu/10.1037/pst0000161.

Funk, J. L., and Rogge, R. D. (2007). Testing the ruler with item response theory: increasing precision of measurement for relationship satisfaction with the Couples Satisfaction Index. *Journal of Family Psychology*, *21*(4), 572. https://doi.org/10.1037/0893-3200.21.4.572.

Futris, T. G. and Adler-Baeder, F. (2013). *The National Extension Relationship and Marriage Education Model: Core teaching concepts for relationship and marriage enrichment programming*. Athens, GA: The University of Georgia Cooperative Extension. Available from http://www.nermen.org/NERMEM.php.

Futris, T. G., Adler-Baeder, F., Ketring, S., Smith, T., Bradford, A., Cook, L., …, and Kehoe, J. (2015). *ELEVATE: Taking your relationship to the next level.* Auburn, AL: Alabama Cooperative Extension System. Publication No. FCS-2047. Available at http://www.nermen.org/ELEVATE.php.

Galovan, A. M., Drouin, M., and McDaniel, B. T. (2018). Sexting profiles in the United States and Canada: Implications for individual and relationship well-being. *Computers in Human Behavior*, *79*, 19-29.

Gambrel, L. E., and Piercy, F. P. (2014a). Mindfulness-based relationship education for couples expecting their first child—Part 1: A randomized mixed-methods program evaluation. *Journal of Marital and Family Therapy, 41(1)*, 5–24. https://doi.org/10.1111/jmft.12066.

Gambrel, L. E., and Piercy, F. P. (2014b). Mindfulness-based relationship education for couples expecting their first child—Part2: Phenomenological findings. *Journal of Marital and Family Therapy, 41(1),* 25–41. https://doi.org/10.1111/jmft.12065.

Gandek, B., Ware, J. E., Aaronson, N. K., Apolone, G., Bjorner, J. B., Brazier, J. E., et al. (1998). Cross-validation of item selection and scoring for the SF-12 Health Survey in nine countries: Results from the IQOLA Project. International Quality of Life Assessment. *Journal of Clinical Epidemiology, 51(11),* 1171–1178. https://doi.org/10.1016/S0895-4356(98)00109-7.

Garneau, C. L., Higginbotham, B., and Adler-Baeder, F. (2015). Remarriage beliefs as predictors of marital quality and positive interaction in stepcouples: An actor–partner interdependence model. *Family Process*, *54*(4), 730–745. https://doi.org/10.1111/famp.12153.

Grossman, P., Niemann, L., Schmidt, S., and Walach, H. (2004). Mindfulness-based stress reduction and health benefits: A meta-analysis. *Journal of psychosomatic research*, *57*(1), 35-43. https://doi.org/10.1016/S0022-3999(03)00573-7.

Gubits, D., Lowenstein, A. E., Harris, J., Hsueh, J. (2014). Do the effects of a relationship education program vary for different types of couples? Exploratory subgroup analysis in the supporting healthy marriage evaluation (OPRE report 2014-22). Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

Gueorguieva, R., and Krystal, J. H. (2004). Move over anova: progress in analyzing repeated-measures data andits reflection in papers published in the archives of general psychiatry. *Archives of general psychiatry*, *61*(3), 310-317. https://doi.org/10.1001/archpsyc.61.3.310.

Halford, W. K., Bodenmann, G. (2013). Effects of relationship education on maintenance of couple relationship satisfaction. *Clinical Psychology Review, 33*, 512–525. https://doi.org/10.1016/j.cpr.2013.02.001.

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

Harcourt, K. T., Adler-Baeder, F., Rauer, A., Pettit, G. S., and Erath, S. (2017). Relationship education for incarcerated adults. *Family Process*, *56*(1), 75–90. https://doi.org/10.1111/famp.12164.

Hawkins, A. J. (2019). *Are federally supported relationship education programs for lower-income individuals and couples working?* Washington, DC: American Enterprise Institute. 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., 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., and Fackrell, T. A. (2010). Does relationship and marriage education for lower-income couples work? A meta-analytic study of emerging research. *Journal of Couple and Relationship Therapy, 9*, 181–191. https://doi.org/10.1080/15332691003694927.

Hawkins, A. J., and Ooms, T. (2012). Can marriage and relationship education be an effective policy tool to help low income couples form and sustain healthy marriages and relationships? A review of lessons learned. *Marriage and Family Review, 48*, 524–554. https://doi.org/10.1080/01494929.2012.677751.

Hawkins, A. J., Stanley, S. M., Blanchard, V. L., and Albright, M. (2012). Exploring programmatic moderators of the effectiveness of marriage and relationship education programs: A meta-analytic study. *Behavior Therapy*, *43*(1), 77-87.

Hawkins, A. J., and VanDenBerghe, B. (2014). *Facilitating forever: A feasible public policy agenda to help couples form and sustain healthy relationships and enduring marriages*. Charlottesville, VA: National Marriage Project.

Higginbotham, B. J., and Myler, C. (2010). The influence of facilitator and facilitation characteristics on participants' ratings of stepfamily education. *Family Relations, 59,* 74-86.

Hoofs, H., van de Schoot, R., Jansen, N. W. H., and Kant, I. (2018). Evaluation model fit for Bayesian confirmatory factor analysis with large samples: Simulation study introducing the BRMSEA. *Educational and Psychological Measurement, 78*(4), 537-568.

Hsueh, J., Principe Alderson, D., Lundquist, E., Michalopoulos, C., Gubits, D., Fein, D., and Knox, V. (2012). The supporting healthy marriage evaluation: Early impacts on low-income families. *OPRE Report*, *11*.

Hyndman, R. J., and Athanasopoulos, G. (2018). Forecasting: principles and practice, 2nd edition, OTexts: Melbourne, Australia. OTexts.com/fpp2. Accessed on 8/26/2020.

Kabat-Zinn, J. (1990). *Full catastrophe living: The program of the stress reduction clinic at the University of Massachusetts Medical Center.* New York, NY: Delta.

Ketring, S. A., Bradford, A. B., Davis, S. Y., Adler-Baeder, F., McGill, J., and Smith, T. A. (2017). The role of the facilitator in couple relationship education. *Journal of Marital and Family Therapy, 43*(3), 374-390. https://doi.org/10.1111/jmft.12223.

Kevin, J., and Risla, N. (2020). Influence of socioeconomic status and mindset on individual's marital satisfaction during the national lockdown. *The International Journal of Indian Psychology, 8*(2), 2349-3429.

Kids Count (2014). Data book: State trends in child well-being (25th ed.). Retrieved from www.aecf.org/2014db.

Kiecolt-Glaser, J. K., and Wilson, S. J. (2017). Lovesick: How couples' relationships influence health. *Annual review of clinical psychology*, *13*, 421-443.

Kirkland, C. L., Skuban, E., Adler-Baeder, F., Ketring, S. A., Bradford, A., Smith, T., et al. (2011). Effects of relationship/marriage education on co-parenting and children's social skills: Examining rural minority parents' experiences. *Early Childhood Research and Practice, 13(2).*

Kolb, D. A. 1984. *Experiential learning: Experience as the source of learning and development.* Englewood Cliffs, New Jersey: Prentice-Hall.

Liu, X., Wheeler, N. J., Broda, M. D., Daire, A. P., Dominguez, V. N., Griffith, S. A. M., and Pease, J. C. (2020). Relationship satisfaction trajectories among low-income ethnic minority couples before and after a relationship education intervention. *Journal of Social and Personal Relationships.* doi:10.1177/0265407520916203.

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

Lundquist, E., Hsueh, J., Lowenstein, A. E., Faucetta, K., Gubits, D., Michalopoulos, C., and Knox, V. (2014). A family-strengthening program for low-income families: Final impacts from the Supporting Healthy Marriage evaluation. New York, NY: MDRC.

Markman, H. J., and Rhoades, G. K. (2012). Relationship education research: current status and future directions. *Journal of Marital and Family Therapy*, *38*(1), 169–200. https://doi.org/10.1111/j.1752-0606.2011.00247.x.

Maruish, M. (2012). *User's manual for the SF-12v2 health survey* (3rd ed.). Lincoln: Quality Metric Incorporated.

McDonald, J.H. 2014. *Handbook of Biological Statistics (3rd ed.)*. Sparky House Publishing: Baltimore, Maryland.

McGill, J., and Adler-Baeder, F. (2019). Exploring the link between mindfulness and relationship quality: Direct and indirect pathways. *Journal of Marital and Family Therapy*. https://doi.org/10.1111/jmft.12412.

McGill, J., Adler-Baeder, F. and Garneau, C. L. (in press). An evaluation of the ELEVATE program for couples: Considering vulnerabilities and relationship length. *Family Relations.*

McGill, J., Adler-Baeder, F., and Rodriguez, P. (2016). Mindfully in love: A meta-analysis of the association between mindfulness and relationship satisfaction. *Journal of Human Sciences and Extension, 4*, 89-101.

McGill, J.M., Ketring, S., Adler-Baeder, F. (2015). *Couples Connecting Mindfully.* Unpublished facilitators manual.

Moore, Q., Avellar, S., Patnaik, A., Covington, R., and Wu, A. (June, 2018). Parents and children together: Effects of two healthy marriage programs for low-income couples. Office of Planning, Research and Evaluation Administration for Children and Families. https://www.acf.hhs.gov/opre/resource/parents-and-children-together-effects-of-two-healthy-marriage-programs-for-low-income-couples?utm_source=ACF+OPRE+NEWS+6%2F7%2F18&utm_campaign=OPRE+News+6%2F7%2F18&utm_medium=email.

Parsons, C. E., Crane, C., Parsons, L. J., Fjorback, L. O., and Kuyken, W. (2017). Home practice in Mindfulness-Based Cognitive Therapy and Mindfulness-Based Stress Reduction: A systematic review and meta-analysis of participants' mindfulness practice and its association with outcomes. *Behaviour Research and Therapy*, *95*, 29-41.

Patterson, J. M. (2002). Understanding family resilience. *Journal of clinical psychology*, *58*(3), 233-246. https://doi-org.spot.lib.auburn.edu/10.1002/jclp.10019.

R Core Team (2019). R: a language and environment for statistical computing, version 3.0. 2. Vienna, Austria: R Foundation for Statistical Computing; 2013.

Rauer, A. J., Adler-Baeder, F., Lucier-Greer, M., Skuban, E., Ketring, S. A., and Smith, T. (2014). Exploring processes of change in couple relationship education: Predictors of change in relationship quality. *Journal of Family Psychology, 28(1),* 65-76. https://doi.org/10.1037/a0035502.

Roberson, P. N. E., Lenger, K. A., Gray, T., Cordova, J., and Gordon, K. C. (2020). Dyadic latent profile analyses and multilevel modeling to examine differential response to couple relationship education. *Journal of Family Psychology*. https://doi-org.spot.lib.auburn.edu/10.1037/fam0000667.

Robles, T. F. (2014). Marital quality and health: Implications for marriage in the 21st century. *Current Directions in Psychological Science*, *23*(6), 427-432.

Roddy, M. K., Knopp, K., Georgia Salivar, E., and Doss, B. D. (2020). Maintenance of relationship and individual functioning gains following online relationship programs for low-income couples. *Family Process*.

Roddy, M. K., Rhoades, G. K., and Doss, B. D. (2020b). Effects of ePREP and OurRelationship on low-income couples' mental health and health behaviors: A randomized controlled trial. *Prevention Science.* Advance online publication. https://doi.org/10.1007/s11121-020-01100-y.

Rosenthal, R. (1994). Parametric measures of effect size. In H. Cooper and L. V. Hedges (Eds.), *The Handbook of Research Synthesis* (pp. 231-244). New York, NY: Sage.

Sher, T. G. (2012). What, why, and for whom: couples interventions: A deconstruction approach. *Behavior Therapy*, *43*(1), 123-131. https://doi.org/10.1016/j.beth.2011.03.006.

Stan Development Team (2017). RStan: The R interface to Stan. R package version 2.16.2. Retrieved from http://mc-stan.org.

Stanley, S. M., Carlson, R. G., Rhoades, G. K., Markman, H. J., Ritchie, L. L., and Hawkins, A. J. (2019). Best practices in relationship education focused on intimate relationships. *Family Relations: An Interdisciplinary Journal of Applied Family Studies*. https://doi-org.spot.lib.auburn.edu/10.1111/fare.12419.

U.S. Census Bureau (2018). Income and poverty in the United States: 2017. Retrieved from https://www.census.gov/library/publications/2018/demo/p60-263.html.

U.S. Census Bureau (2019). *Quick Facts: Alabama*. Retrieved from https://www.census.gov/quickfacts/AL.

van Buuren, S., and Groothuis-Oudshoorn, K. (2011). mice: multivariate imputation by chained equations in RJ Stat. Softw. 45, 1–67. https://doi.org/10.18637/jss.v045.i03.

Wadsworth, M. E., and Markman, H. J. (2012). Where's the action? Understanding what works and why in relationship education. *Behavior Therapy, 43(1),* 99–112. https://doi.org/10.1016/j.beth.2011.01.006.

Ware, J., Kosinski, M., and Keller, S. D. (1996). A 12-Item Short-Form Health Survey: Construction of scales and preliminary tests of reliability and validity. *Medical Care*, *34*(3), 220–233. https://doi.org/10.1097/00005650-199603000-00003.

Weiss, H. B., and Jacobs, F. H. (Eds.). (2008). *Evaluating Family Programs: Current Issues in Theory and Policy*. Transaction Publishers.

Williamson, H. C., Hammett, J. F., Ross, J. M., Karney, B. R., and Bradbury, T. N. (2018). Premarital education and later relationship help-seeking. *Journal of Family Psychology*, *32*, 276–281. https://doi.org/10.1037/fam0000383.

Williamson, H. C., Karney, B. R., and Bradbury, T. N. (2019). Barriers and facilitators of relationship help-seeking among low-income couples. *Journal of Family Psychology, 33*(2), 234–239.

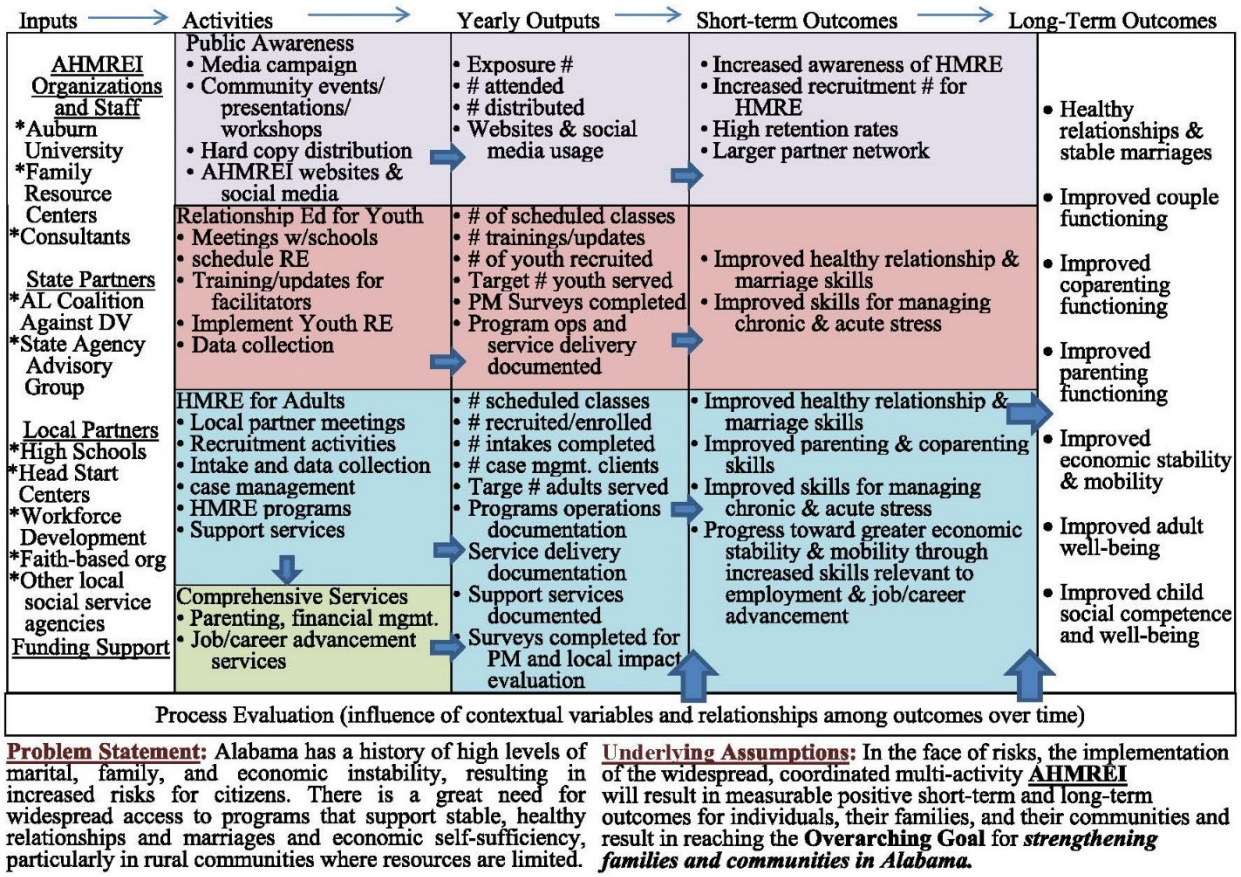
Williamson, H. C., Trail, T. E., Bradbury, T. N., and Karney, B. R. (2014). Does premarital education decrease or increase couples' later help-seeking? *Journal of Family Psychology, 28,* 112–117. http://dx.doi.org/10.1037/a0034984b

Wood, R. G., Moore, Q., Clarkwest, A., and Killewald, A. (2014). The long-term effects of Building Strong Families: A program for unmarried parents. *Journal of Marriage and Family, 76*, 446–463. https://doi.org/10.1111/jomf.12094.

# VIII. APPENDICES

## A. Logic model (or theory of change) for the AHMREI

**Figure A.1. Logic model (or theory of change) for the AHMREI**




Note: AL = Alabama; Org = organization; Ed = education; RE = relationship education; HMRE = healthy marriage and relationship education; PM = performance measurement.

## B. Data and study sample

Data collection procedures for the implementation study are reported in Section III.B.1 (see Table B.1 below). Description of the methods used to analyze the implementation data can be found in Section V.A. Description of the data sources for the impact study can be found in Section III.B.1 (see table B.2 below). Finally, the consort diagrams included here display how the final samples for each impact research question were selected.

**Table B.1. Data sources used to address the implementation research questions**

| Implementation element | Research question | Data source | Timing/frequency of data collection | Party responsible for data collection |
|---|---|---|---|---|
| **Intervention** | | | | |
| Fidelity | On average, what is the adherence to the curriculum content, as reported by (1a) ELEVATE and (1b) CCM facilitators? | Qualtrics survey | 2-month follow-up | Evaluation staff |
| Dosage | Of those assigned to each program group, (2a) ELEVATE and (2b) CCM, what proportion of individuals attend none, 1-2 sessions, 3-5 sessions, and all 6 of the sessions offered? | Workshop sessions in nFORM | At each class during 6-week program delivery | Intervention staff |
| Quality | What is the average rating of facilitator quality? | Qualtrics survey | 2-month follow-up | Evaluation staff |
| Quality | What is the average rating of the facilitator-participant alliance? | Qualtrics survey | 2-month follow-up | Evaluation staff |
| Engagement | What is the average rating of self- engagement in the program content? | Qualtrics survey | 2-month follow-up | Evaluation staff |
| Engagement | What is the average rating of perceived engagement of the other class members as a group? | Qualtrics survey | 2-month follow-up | Evaluation staff |
| Context | What proportion of the ELEVATE and CCM participants reported participating in other HMRF programming or outside services from baseline to 1-year follow-up? | Qualtrics survey | 2-month, 6-month, 1-year, and 2-year follow-up | Evaluation staff |
| **Counterfactual (Control/comparison group questions** | | | | |
| Context | What proportion of the control group reported participating in other HMRF programming or outside services from baseline to 1-year follow-up? | Qualtrics survey | 2-month, 6-month, 1-year, and 2-year follow-up | Evaluation staff |

**Table B.2. Key features of the impact analysis data collection**

| Data source | Timing of data collection | Mode of data collection | Party responsible for data collection | Start and end date of data collection |
|---|---|---|---|---|
| **Intervention** | | | | |
| Intervention group study participants | Enrollment (baseline) | Online Qualtrics survey, nFORM applicant characteristics survey, entrance survey | Evaluation staff<br>Implementation staff | September 2016–January 2018 |
| | 2-month follow-up | Online Qualtrics survey, nFORM exit survey | Evaluation staff<br>Implementation staff | October 2016–March 2018 |
| | 6-month follow-up | Online Qualtrics survey | Evaluation staff | March 2017–August 2018 |
| | 1-year follow-up | Online Qualtrics survey | Evaluation staff | September 2017–February 2019 |
| | 2-year follow-up | Online Qualtrics survey | Evaluation staff | September 2018–February 2020 |
| **Counterfactual (control/comparison)** | | | | |
| Comparison group study participants | Enrollment (baseline) | Online Qualtrics survey | Evaluation staff | September 2016–January 2018 |
| | 2-month follow-up | Online Qualtrics survey | Evaluation Staff | October 2016–March 2018 |
| | 6-month follow-up | Online Qualtrics survey | Evaluation staff | March 2017–August 2018 |
| | 1-year follow-up | Online Qualtrics survey | Evaluation Staff | September 2017–February 2019 |
| | 2-year follow-up | Online Qualtrics survey | Evaluation Staff | September 2018–February 2020 |

**Figure B.1. CONSORT diagram for Couple Relationship Skills Inventory (CRSI)**

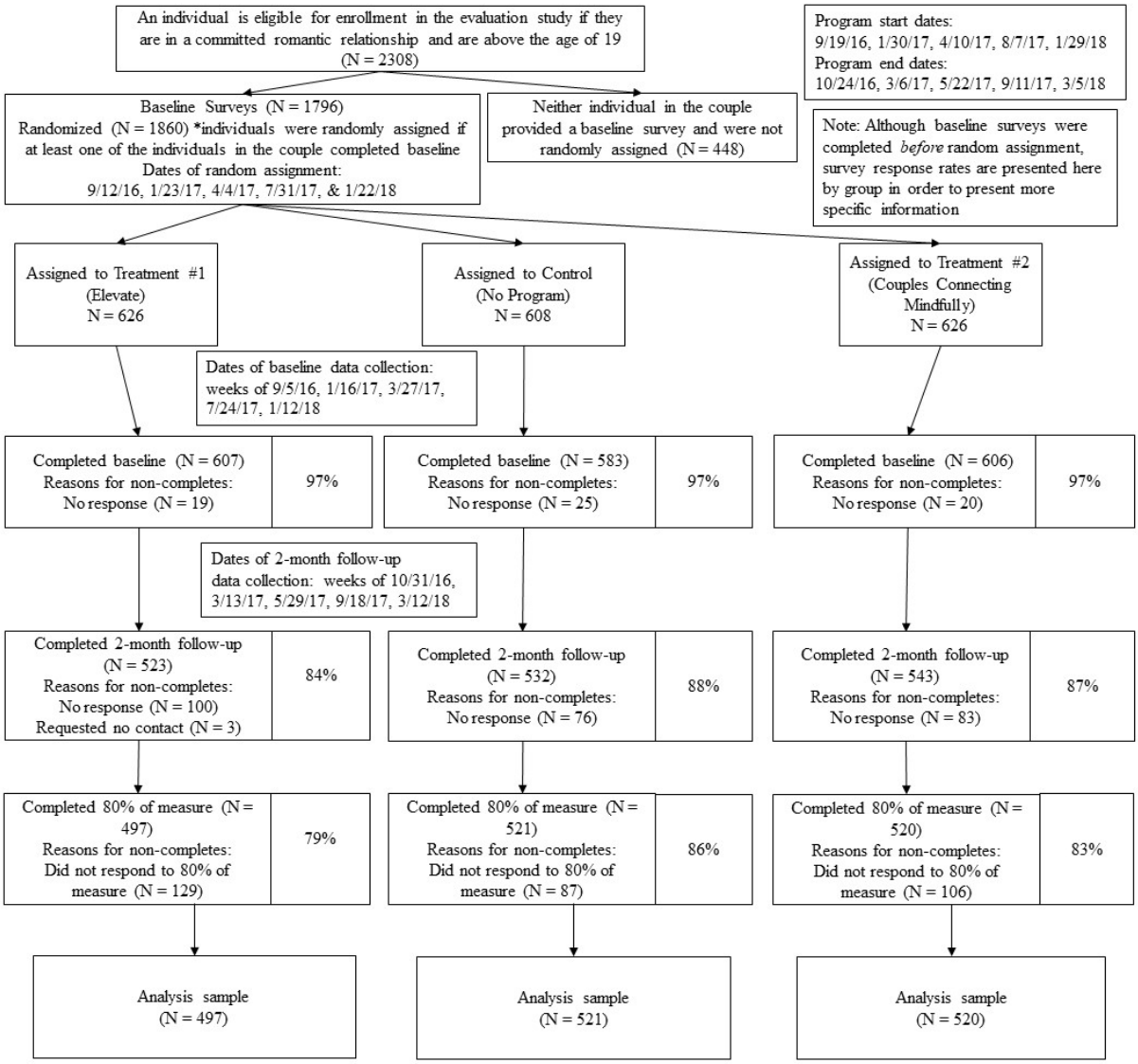


**Figure B.2. CONSORT diagram for mental health**

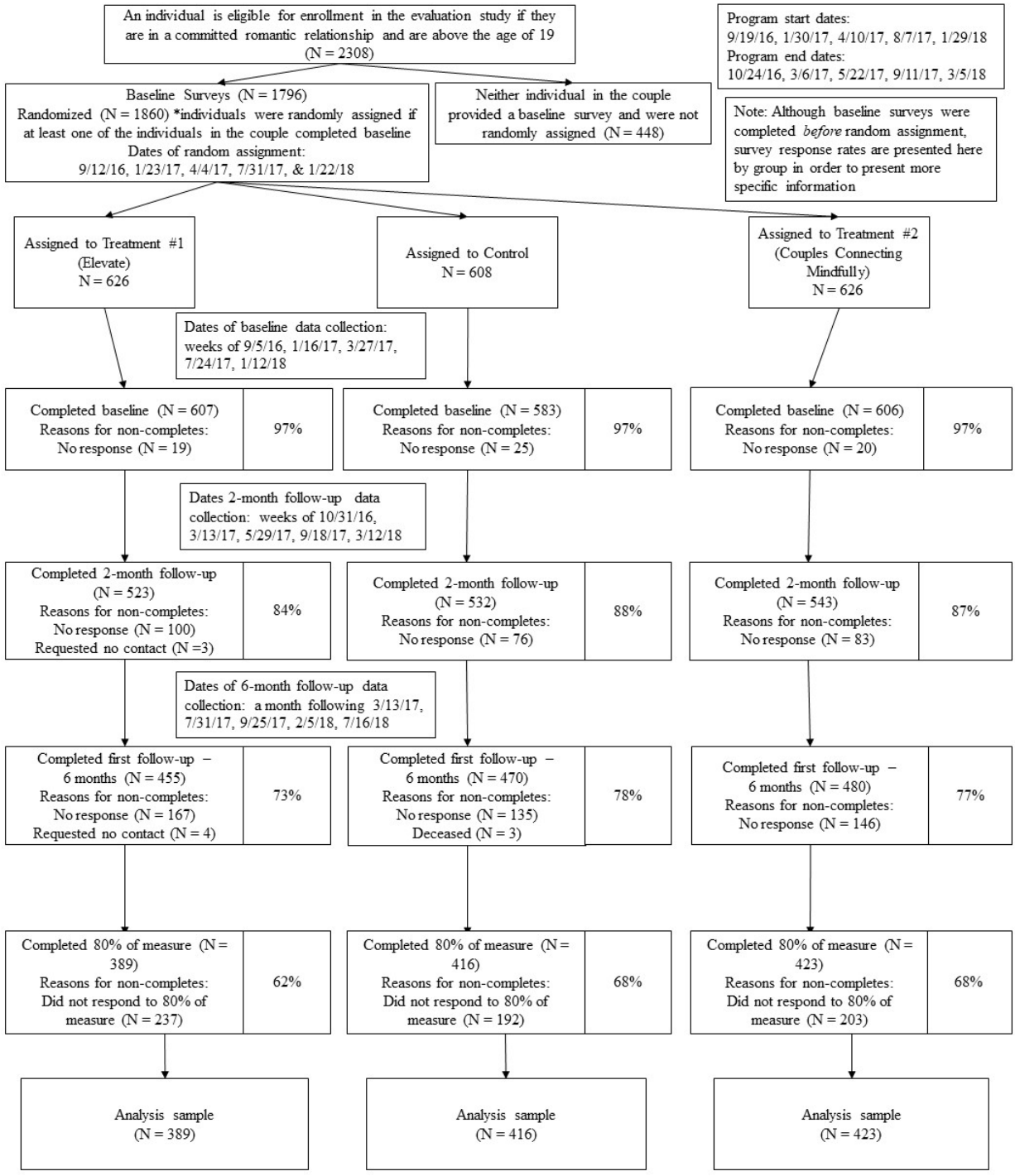
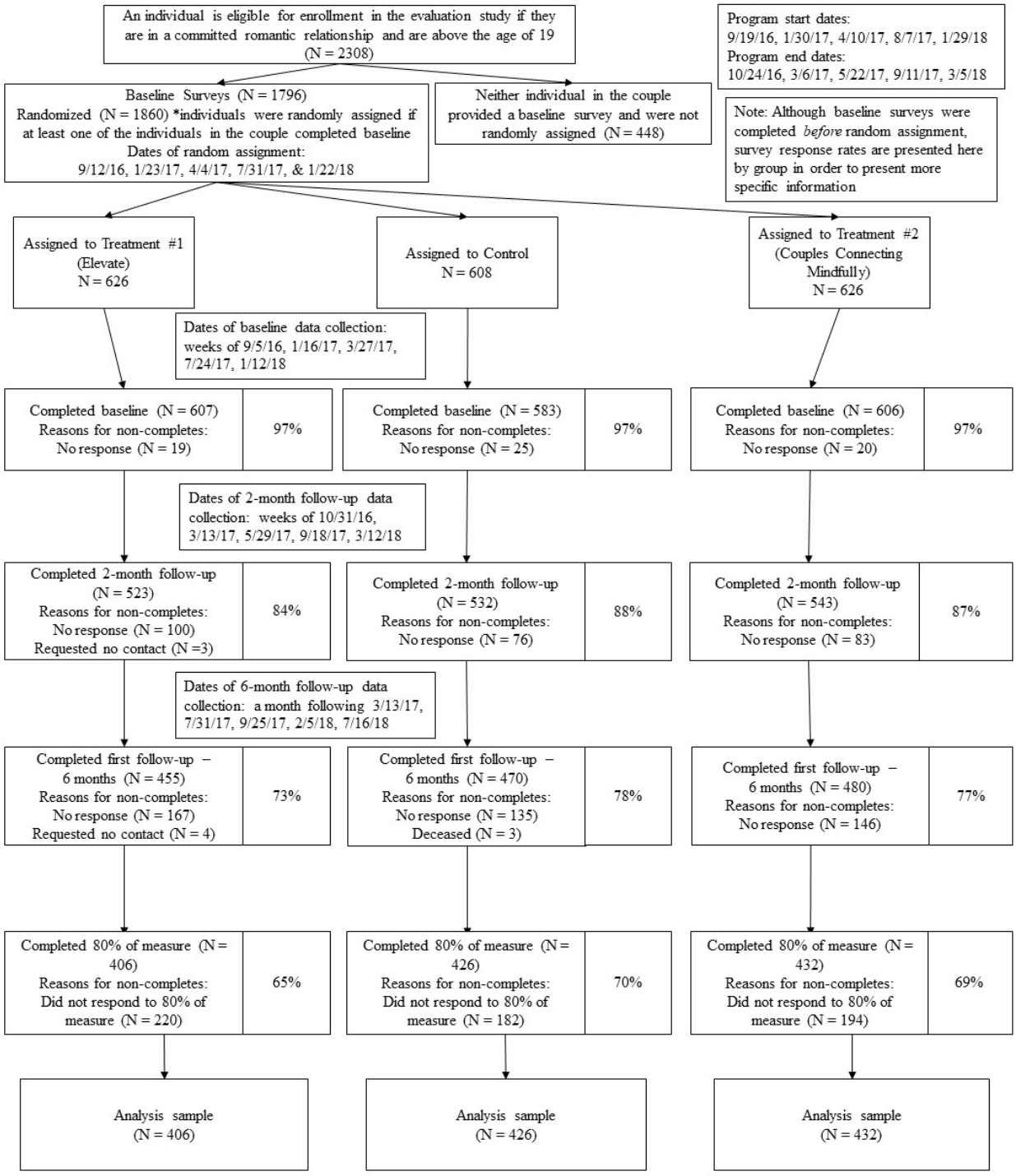


**Figure B.3. CONSORT diagram for Couple Satisfaction Index (CSI)**



## C. Attrition rates and baseline equivalence of the RCT design

Because we had very high retention standards for the analytic sample (i.e., 80+% of outcome measure completed at all time points relevant to the RQ) there were 3 out of 6 instances when allowable differential attrition rates did not meet the cautious boundary (the most stringent level of allowable attrition). Overall attrition rates for research questions focused on immediate post-program (i.e., 2-month follow-up) change ranged from 16-18%. Overall attrition rates for research questions focused on changes over 6-months ranged from 30-35%. Differential attrition rates for all research questions ranged from 0-7%. Details about attrition can be seen in the CONSORT diagrams (Figures A, B, and C) above. Below, we briefly summarize the analytic sample sizes for the respective research questions (we also note that these sample sizes are reported as individual participants, not as couples).

For primary Impact RQ #1 *(What is the impact of ELEVATE when compared to the no-program control group on change in couple relationship skills from baseline to two-month follow-up?)*, 626 individuals were randomly assigned to Elevate and 608 individuals were randomly assigned to the control group. Of those randomly assigned, 523 respondents assigned to ELEVATE and 532 control respondents completed some part of both the baseline and immediate follow-up surveys. Of those, 497 ELEVATE participants and 521 control respondents completed 80% of the Couple Relationship Skills Inventory items and were thus included in the final analytic sample for research question #1. The overall attrition rate from random assignment to inclusion in this sample was 18%. The differential attrition rate between ELEVATE and control for research question #1 was 7%. Considering the overall attrition rate of 18% owing to the high standard of factors for inclusion, the differential attrition rate of 7% did not meet the cautious boundary based on the WWC guidelines (5.7%) and can be considered high.

**Table C.1. Sample sizes of individuals by intervention status for Impact RQ #1**

| Couple relationship skills number of individuals | ELEVATE sample size | Control sample size | Total sample size | Total response rate | ELEVATE response rate | Control response rate |
|---|---|---|---|---|---|---|
| Assigned to condition | 626 | 608 | 1,234 | n.a. | n.a. | n.a. |
| Contributed a baseline survey | 607 | 583 | 1,190 | 96% | 97% | 97% |
| Contributed to baseline and 2-month follow-up | 523 | 532 | 1,055 | 85% | 84% | 88% |
| Analytic sample: Contributed to baseline and 2-month follow-up (accounts for item non-response) | 497 | 521 | 1,018 | 82% | 79% | 86% |

Note: Item non-response referred to participants who did not respond to 80+% of the 32-item measure so that mean imputation could be conducted.

n.a. = not applicable.

For primary Impact RQ #2 *(What is the impact of CCM when compared to the no-program control group on change in couple relationship skills from baseline to 2-month follow-up?)*, 626 individuals were randomly assigned to CCM and 608 individuals were randomly assigned to the control group. Of those randomly assigned, 543 respondents assigned to CCM and 532 control

respondents completed some of both the baseline and 2-month follow-up surveys. Of those, 520 CCM participants and 521 control respondents completed 80% of the Couple Relationship Skills Inventory items and were thus included in the final analytic sample for research question #2. The overall attrition from random assignment to inclusion in this sample was 16%. The differential attrition rate between Elevate and control for research question #2 was 3%. Considering the overall attrition rate of 16% owing to the high standard of factors for inclusion, the differential attrition rate of 3% met the cautious boundary based on the WWC guidelines (5.9%) and is considered acceptable.

**Table C.2. Sample sizes of individuals by intervention status for Impact RQ #2**

| Couple relationship skills number of individuals | CCM sample size | Control sample size | Total sample size | Total response rate | ELEVATE response rate | Control response rate |
|---|---|---|---|---|---|---|
| Assigned to condition | 626 | 608 | 1,234 | n.a. | n.a. | n.a. |
| Contributed a baseline survey | 606 | 583 | 1,189 | 96% | 97% | 97% |
| Contributed to baseline and 2-month follow-up survey | 543 | 532 | 1,075 | 87% | 87% | 88% |
| Analytic sample: Contributed to baseline and 2-month follow-up (accounts for item non-response) | 520 | 521 | 1,041 | 84% | 83% | 86% |

Note: Item non-response referred to participants who did not respond to 80+% of the 32-item measure so that mean imputation could be conducted.

n.a. = not applicable.

For primary Impact RQ #3 (What is the impact of ELEVATE when compared to the no-program control group on change in individual mental health from baseline to six-month follow-up?), 626 individuals were randomly assigned to ELEVATE and 608 individuals were randomly assigned to the control group. Of those randomly assigned, 455 respondents assigned to ELEVATE and 470 control respondents completed some of the baseline, 2-month follow-up, and 6-month follow-up surveys. Of those, 389 ELEVATE participants and 416 control respondents completed all of the SF-12 Mental Health items and were thus included in the final analytic sample for research question #3. Full completion was required since the SF-12 scoring procedures represents a sum rather than an average of items. The overall attrition from random assignment to inclusion in this sample was 35%. The differential attrition rate between ELEVATE and control for research question #3 was 6%. Considering the overall attrition of 35% owing to the high standard of factors for inclusion, the differential attrition of 6% did not meet the cautious boundary based on the WWC guidelines (3.3%) and can be considered high.

**Table C.3. Sample sizes of individuals by intervention status for Impact RQ #3**

| Mental health number of individuals | ELEVATE sample size | Control sample size | Total sample size | Total response rate | ELEVATE response rate | Control response rate |
|---|---|---|---|---|---|---|
| Assigned to condition | 626 | 608 | 1,234 | n.a. | n.a. | n.a. |
| Contributed a baseline survey | 607 | 583 | 1,190 | 96% | 97% | 97% |
| Contributed to baseline and 2-month follow-up survey | 523 | 532 | 1,055 | 85% | 84% | 88% |
| Contributed to baseline, 2-month follow-up, and 6-month follow-up surveys | 455 | 470 | 925 | 75% | 73% | 78% |
| Analytic sample: Contributed to baseline, 2-month follow-up, and 6-month follow-up surveys (accounts for item non-response) | 389 | 416 | 805 | 65% | 62% | 68% |

Note: Item non-response referred to participants who did not respond to all items in the measure because mean imputation could not be conducted for the F-12 measure.

n.a. = not applicable.

For primary Impact RQ #4 *(What is the impact of CCM when compared to the no-program control group on change in individual mental health from baseline to six-month follow-up?)*, 626 individuals were randomly assigned to CCM and 608 individuals were randomly assigned to the control group. Of those randomly assigned, 480 respondents assigned to CCM and 470 control respondents completed some of the baseline, 2-month follow-up, and 6-month follow-up surveys. Of those, 423 CCM participants and 416 control respondents completed all of the SF-12 Mental Health items and were thus included in the final analytic sample for research question #4. Full completion was required since the SF-12 scoring procedures represents a sum rather than an average of items. The overall attrition from random assignment to inclusion in this sample was 32%. The differential attrition rate between CCM and control for research question #4 was 0%. Considering the overall attrition rate of 32% owing to the high standard of factors for inclusion, the differential attrition rate of 0% met the cautious boundary based on the WWC guidelines (3.8%) and is considered acceptable.

**Table C.4. Sample sizes of individuals by intervention status for Impact RQ #4**

| Mental health number of individuals | CCM sample size | Control sample size | Total sample size | Total response rate | ELEVATE response rate | Control response rate |
|---|---|---|---|---|---|---|
| Assigned to condition | 626 | 608 | 1,234 | n.a. | n.a. | n.a. |
| Contributed a baseline survey | 606 | 583 | 1,189 | 96% | 97% | 97% |
| Contributed to baseline and 2-month follow-up survey | 543 | 532 | 1,075 | 87% | 87% | 88% |
| Contributed to baseline, 2-month follow-up, and 6-month follow-up surveys | 480 | 470 | 950 | 77% | 77% | 78% |

| Mental health number of individuals | CCM sample size | Control sample size | Total sample size | Total response rate | ELEVATE response rate | Control response rate |
|---|---|---|---|---|---|---|
| Analytic sample: Contributed to baseline, 2-month follow-up, and 6-month follow-up surveys (accounts for item non-response) | 423 | 416 | 839 | 68% | 68% | 68% |

Note: Item non-response referred to participants who did not respond to all items in the measure because mean imputation could not be conducted for the F-12 measure.

n.a. = not applicable.

For primary Impact RQ #5 *(What is the impact of ELEVATE when compared to the no-program control group on change in couple satisfaction from baseline to six-month follow-up?)*, 626 individuals were randomly assigned to ELEVATE and 608 individuals were randomly assigned to the control group. Of those randomly assigned, 455 respondents assigned to ELEVATE and 470 control respondents completed some of the baseline, 2-month follow-up, and 6-month follow-up surveys. Of those, 406 ELEVATE participants and 426 control respondents completed all three of the Couple Satisfaction Index items and were thus included in the final analytic sample for research question #5. Full completion was required since completion of 2 of the 3 items on the CSI did not meet the 80+% completion criterion. The overall attrition from random assignment to inclusion in this sample was 33%. The differential attrition rate between ELEVATE and control for research question #5 was 5%. Considering the overall attrition rate of 33% owing to the high standard of factors for inclusion, the differential attrition rate of 5% did not meet the cautious boundary based on the WWC guidelines (3.6%) and can be considered high.

**Table C.5. Sample sizes of individuals by intervention status for Impact RQ #5**

| Couple satisfaction number of individuals | ELEVATE sample size | Control sample size | Total sample size | Total response rate | ELEVATE response rate | Control response rate |
|---|---|---|---|---|---|---|
| Assigned to condition | 626 | 608 | 1,234 | n.a. | n.a. | n.a. |
| Contributed a baseline survey | 607 | 583 | 1,190 | 96% | 97% | 97% |
| Contributed to baseline and 2-month follow-up survey | 523 | 532 | 1,055 | 85% | 84% | 88% |
| Contributed to baseline, 2-month follow-up, and 6-month follow-up surveys | 455 | 470 | 925 | 75% | 73% | 78% |
| Analytic sample: Contributed to baseline, 2-month follow-up, and 6-month follow-up surveys (accounts for item non-response) | 406 | 426 | 832 | 67% | 65% | 70% |

Note: Item non-response referred to participants who did not respond to all 3 items in the measure because 80% of the measure was needed to conduct mean imputation.

n.a. = not applicable.

For primary Impact RQ #6 *(What is the impact of CCM when compared to the no-program control group on change in couple satisfaction from baseline to six-month follow-up?)*, 626 individuals were randomly assigned to CCM and 608 individuals were randomly assigned to the control group. Of those randomly assigned, 480 respondents assigned to CCM and 470 control respondents completed some of the baseline, 2-month follow-up, and 6-month follow-up surveys. Of those, 432 CCM participants and 426 control respondents completed all three of the Couple Satisfaction Index items and were thus included in the final analytic sample for research question #6. Full completion was required since completion of 2 of the 3 items on the CSI did not meet the 80+% completion criterion. The overall attrition from random assignment to inclusion in this sample was 30%. The differential attrition rate between CCM and control for research question #6 was 1%. Considering the overall attrition rate of 30% owing to the high standard of factors for inclusion, the differential attrition rate of 1% met the cautious boundary based on the WWC guidelines (4.1%) and is considered acceptable.

**Table C.6. Sample sizes of individuals by intervention status for Impact RQ #6**

| Couple satisfaction number of individuals | CCM sample size | Control sample size | Total sample size | Total response rate | ELEVATE response rate | Control response rate |
|---|---|---|---|---|---|---|
| Assigned to condition | 626 | 608 | 1,234 | n.a. | n.a. | n.a. |
| Contributed a baseline survey | 606 | 583 | 1,189 | 96% | 97% | 97% |
| Contributed to baseline and 2-month follow-up surveys | 543 | 532 | 1,075 | 87% | 87% | 88% |
| Contributed to baseline, 2-month follow-up, and 6-month follow-up surveys | 480 | 470 | 950 | 77% | 77% | 78% |
| Analytic sample: Contributed to baseline, 2-month follow-up, and 6-month follow-up surveys (accounts for item non-response) | 432 | 426 | 858 | 70% | 69% | 70% |

Note: Item non-response referred to participants who did not respond to all 3 items in the measure because 80% of the measure was needed to conduct mean imputation.

n.a. = not applicable.

Tables IV.4, IV.5, IV.6 summarize the baseline equivalence results for all impact research questions. Overall, the random assignment process was successful in creating nearly equivalent groups at baseline, even for RQs with high levels of attrition. Specifically, there were differences related to relationship status; however, these differences are only relevant for two of the six impact RQs. For RQ #1, female ELEVATE participants were more likely to be married compared to female control respondents (76% vs. 65%; *Chi-Square* = 8.17, $p$ = .004). Similarly, for RQ #3, female ELEVATE participants were more likely to be married compared to female control respondents (79% vs. 64%; *Chi-Square* = 11.64, $p$ = .001). Moreover, for all the impact RQs, no significant differences on baseline levels of outcomes (i.e., CRSI, mental health, couple satisfaction) were found.

## D. Data preparation

Survey responses for each cohort across each time point were downloaded from Qualtrics into separate SPSS databases. These databases were matched based on the individual participant ID in order to create a master dataset with survey responses from each of the five cohorts across each of the data collection time points. Basic descriptive statistics were run on all variables of interest to assess certain assumptions such as the normality assumptions for certain continuous measures or the presence of impossible/incorrect scores on each scale.

Furthermore, evaluation staff used a master code list containing participant IDs, demographic data, and nFORM to verify the accuracy and consistency of demographic data across participants and different timepoints. Responses that were obvious keystroke errors made by respondents based on master code list information and inconsistency with other timepoint reports were corrected by evaluation staff. We also assessed couple demographics reported between dyads for agreement. If a couple disagreed on marital status (i.e., one member reports being married and the other one reports nonmarried), we deferred to the nonmarried report. This situation occurred in a small proportion of the sample (N = 33 couples; less than 4% of the sample).

## E. Impact estimation

The following equations were used to assess the effect size of baseline equivalence estimation:

**Fisher's z**

$$z = .5[\ln(1+r) - \ln(1-r)]$$

Where:

$r$ = correlation

ln = natural logarithm

**Cohen's d**

$$d = M_1 - M_2 \,/\, S_{pooled}$$

Where:

$M_1$ = mean of group 1

$M_2$ = mean of group 2

$S_{pooled}$ = pooled standard deviations for the two groups.

Formula: $(S_1^2 + S_2^2) / 2$

$S_1$ = standard deviation of group 1

$S_2$ = standard deviation of group 2

The following equations were used in the assessment of program impacts:

**Two-level regression model: Impact RQs #1 and 2**

These models predict the value of the outcome at 2-month follow-up, as predicted by randomly assigned treatment or control condition, while accounting for baseline levels of the outcome. In other words, these models predict future behaviors based on past behaviors and the outcome can be interpreted as residual change (Hyndman and Athanasopoulos, 2018). The models include individuals (level one) as indicated by i within couples (level 2) as indicated by j.

*The level one equation is:*

$$Y_{ij} = \beta_{0j} + \beta_1 RA_{ij} + e_{ij}$$

Where:

$Y_{ij}$ = observed outcome of interest for individual i in couple j

$\beta_{0j}$ = the baseline mean of the outcome of interest for couple j in the control group

$\beta_1$ = the average treatment effect on the outcome of interest

$RA_{ij}$ =1 for the program (either Elevate or CCM), and $RA_{ij}$=0 for the control

$e_{ij}$ = residuals of each individual within couple j under the assumption that it is normally distributed with mean equal to 0, and constant variance

*The level two equations are:*

$$\beta_{0j} = \gamma_{00} + \mu_{0j}$$

Where:

$\gamma_{00}$ = the mean outcome of interest for all couples

$\mu_{0j}$ = the unique effect of couple j on the mean outcome of interest. It is a random term with mean 0 and variance

**Three-level growth model: Impact RQs #3–#6**

These equations model the outcome from baseline to 2-month follow-up to 6-month follow-up, as predicted by the treatment indicator. The models account for the three repeated measures over time within individuals (level 1) as indicated by i, individuals (level two) as indicated by j, and between couples (level 3) as indicated by k.

*The level one equation is:*

$$Y_{ijk} = \pi_{0jk} + \pi_{1jk} Time_{ijk} + \varepsilon_{ijk}$$

Where:

$Y_{ijk}$ = the outcome of interest at time i for individual j in couple k

$\pi_{0jk}$ = the mean of the outcome of interest for the control group across individuals at time 0 (baseline)

$\pi_{1jk}$ = the average growth rate of the outcome of interest

$Time_{ijk}$ = the values of 0, 2 and 6 for individuals

$\varepsilon_{ijk}$ = error at level 1 (time within individual) assuming it is normally distributed, has a mean of 0, and the variance is constant

*The level two equations are:*

$$\pi_{0jk} = \beta_{00k} + r_{0jk}$$

$$\pi_{1jk} = \beta_{10k} + r_{1jk}$$

Where:

$\pi_{0jk}$ = the random intercept of the outcome of interest for individual j in couple k at baseline as a function of the mean of the outcome of interest across individuals within couple k and the random deviation of couple level intercept

$\beta_{00k}$ = the mean of the outcome of interest for couple j at baseline

$r_{0jk}$ = the random deviation of individual j in couple k level intercept assuming it is normally distributed, has a mean of 0, and the variance is constant

$\pi_{1jk}$ = the growth rate for individual j in couple k across individuals as a function of the mean of slope of the outcome of interest across individuals and the residuals for the slope at level 2

$\beta_{10k}$ = the average growth rate of the outcome of interest for couple k

$r_{1jk}$ = error for slope at level 2 (between couples) assuming it is normally distributed, has a mean of 0, and the variance is constant

*The level three equations are:*

$$\beta_{00k} = \gamma_{000} + \gamma_{001}Z_k + \mu_{00k}$$

$$\beta_{10k} = \gamma_{010} + \gamma_{010}Z_k + \mu_{01k}$$

Where:

$\beta_{00k}$ = the mean t of the outcome of interest at baseline between random assignment groups as a function of the mean of outcome of interest across groups and the error of the intercept at level 3

$\gamma_{000}$ = the mean of the intercept of the outcome of interest for the control group at baseline between couples

$\gamma_{001}$ = the mean of the outcome of interest at baseline for the control group between couples

$Z_k$ = 1 for the program (either Elevate or CCM) and 0 for the control

$\mu_{00k}$ = error for intercept at level 3 (between couples) assuming it is normally distributed, has a mean of 0, and the variance is constant

$\beta_{10k}$ = the average growth rate of the of the outcome of interest between random assignment as a function of the average growth rate across random assignment groups and the error of the slope at level 3

$\gamma_{010}$ = the average growth rate of the outcome of interest for the control group between couples

$\mu_{01k}$ = error for slope at level 3 (between couples) assuming it is normally distributed, has a mean of 0, and the variance is constant

The following equations were used to assess the effect size (Impact RQs #1 and 2) or amount of variance explained by program participation (Impact RQs #3–#6) in the assessment of program impacts:

**Cohen's d**

$$d = M_1 - M_2 / S_{pooled}$$

Where:

$M_1$ = mean of group 1

$M_2$ = mean of group 2

$S_{pooled}$ = pooled standard deviations for the two groups.

Formula: $\sqrt{[S_1^2 + S_2^2 / 2]}$

$S_1$ = standard deviation of group 1

$S_2$ = standard deviation of group 2

**Adjusted (or change) in $R^2$**

$$R^2_{\text{marginal}} = \frac{\sigma_f^2}{\sigma_f^2 + \sum(\sigma_l^2) + \sigma_e^2}$$

Where:

$\sigma_f^2$ = the variance explained by fixed effects

$\sum(\sigma_l^2)$ = the sum of all variance

$\sigma_e^2$ = the residual variance

## F. Sensitivity analyses and alternative model specifications

Details related to sensitivity analyses are in Section V.C. The alternative model specifications are the same equations as the impact estimations above.

## G. Additional analyses

Description of the methods used for the additional analyses can be found in Section V.D. Figure 1 is an illustration of the conceptual model tested in Additional Impact RQs #5 and #6. Couple relationship skills and couple satisfaction reports at baseline were allowed to covary because they were reported at the same timepoint and are likely to be highly correlated based on the extant literature (Futris and Adler-Baeder, 2014). Two-month follow-up scores, as well as six-month follow-up scores, represent residual change in each outcome because baseline levels were accounted for in the model (Singer and Willett, 2003). Finally, residual change in couple satisfaction at 6-month follow-up was regressed on residual change in couple relationship skills at 2-month follow-up to understand whether changes in targeted skills influence longer-term changes in quality assessments of the couple relationship.

Goodness of fit indices were calculated to assess how well the data used fit the structural equation model, or how consistent the data was with the given model. The current study utilizes common tests for model fit: the chi-square test of model fit, comparative fit index (CFI), and the root mean square error of approximation (RMSEA) to examine goodness of fit. Larger values for the chi-square test of model fit indicate poorer fit. For CFI, values of .95 or higher indicate good model fit, .90-.95 indicate acceptable model fit, and values .90 or lower indicate poor model fit. An RMSEA value of .01 indicates strong model fit; .05 indicates good model fit, and .08 indicates acceptable model fit. Further, for the RMSEA, a non-significant p-value indicates an acceptable model fit because the RMSEA p-value is the probability that the RMSEA is less than or equal to .05. If the p value is greater than .05, this suggests the RMSEA value does not indicate a model rejection. For the ELEVATE group, goodness of fit indices suggested excellent fit of the data to the model (CFI = .99; TLI = .97; RMSEA = .08, $p = .147$; $\chi^2 = 8.21$, $df = 2$, $p = .016$). For the CCM group, Goodness of fit indices suggested excellent fit of the data to the model (CFI = 1.00; TLI = 1.01; RMSEA = .00, p = .996; $\chi^2 = .029$, $df = 2$, $p = .986$).

**Figure G.1. Conceptual model for secondary RQs #5 and #6**