Impact Evaluation of the Alabama Healthy Marriage and Relationship Education Project in Auburn, AL

Final Impact Evaluation Report for

Alabama Healthy Marriage and Relationship Education (AHMRE)

July 29, 2025

Prepared by

Kim Gregson, PhD, LMFT, Gregson Consulting LLC

Francesca Adler-Baeder, PhD, Auburn University

Julianne McGill, PhD, Auburn University

Menglin Wei, PhD, Auburn University

Ziying Wang, PhD, Auburn University

Recommended Citation:

Gregson, K., Adler-Baeder, F., McGill, J., Wei, M., & Wang, Z. (2025). *Impact Evaluation of the Alabama Healthy Marriage and Relationship Education Project in Auburn, AL.* Office of Planning, Research, and Evaluation; Administration for Children and Families.

Acknowledgements:

A heartfelt thank you to the staff, students, and state partners of the Alabama Healthy Marriage and Relationship Education project. Your dedication, enthusiasm, and hard work have been instrumental in the success of this impact evaluation. Most importantly, we deeply appreciate the couples who participated in the programs and the study. Your commitment to advancing research on relationship education in the community is truly exemplary, and we value the time you invested in this process.

We are also grateful to Dan Friend from Mathematica for your exceptional technical assistance and thoughtful and supportive contributions to the process throughout the project. Additionally, we extend our sincere appreciation to our Federal Program Specialists, Toya Joyner and Megan Heffron, for your unwavering leadership, dedication, and support at every stage of our project and research.

Auburn University faculty, staff, & students (listed alphabetically):
Leah Burke, Melike Coskun, Erin Cooper, Jenna Lawlor, Brian Gillis, Mallory Lucier-Greer, Lisa Moyer, Rachel Odomes, Raequon Pinkney, Tonya Rogers, Je'Kylynn Steen, Brad Thomas, Teresa Wagoner, Paityn Whittington

Implementation partner sites (listed alphabetically):
Circle of Care Center for Families, Family Guidance Center of Alabama, IMPACT Family Counseling of Birmingham, Sylacauga Alliance for Family Enhancements, The Family Success Center of Etowah County, The Gulf Coast Family Center, Tuscaloosa's One Place

Disclosure:

The lead author of this report is not affiliated with Auburn University and served as the outside evaluator in collaboration with the co-authors affiliated with Auburn University and the program team.

This publication was prepared under Grant Number 90ZB0007 from the Office of Family Assistance (OFA) within the Administration for Children and Families (ACF), U. S. Department of Health & 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 the AHMRE Project in Auburn, AL"

Objective. The Alabama Healthy Marriage and Relationship Education (AHMRE) project, a partnership between Auburn University and six family resource centers, conducted a longitudinal randomized trial of ELEVATE — a relationship education program— delivered either weekly or monthly to adult couples in communities across Alabama.

Study design. Adult couples (*n* = 1,120 couples) enrolled in the randomized trial in seven separate cohorts over three years. Couples were randomly assigned to either the monthly program condition, experiencing six 2-hour monthly sessions of ELEVATE curriculum + mini-booster retention emails between sessions (*n* = 559 couples), or the weekly program condition, experiencing six 2-hour weekly sessions of ELEVATE curriculum (*n* = 561 couples). After the opportunity to complete consent forms and baseline surveys, couples participated in their randomly assigned program condition: either weekly or monthly ELEVATE, facilitated by two trained facilitators in community settings across Alabama. Participant follow-up surveys were collected at 3-months, 6-months, and 1-year after the baseline survey. Primary impact research questions focused on comparisons of changes in the two program conditions on three outcomes (i.e., couple relationship skills, couple satisfaction, and mental health symptoms) from baseline to post-program (3-months for the weekly group and 6-months for the monthly group) and from baseline to 1-year post-baseline.

Results. The study had low attrition at both individual- and couple-levels for all analytic samples (total analytic sample: *n* = 989 couples; i.e., 490 monthly condition / 499 weekly condition). There were no significant differences in the changes between program conditions. Participants in both monthly and weekly conditions demonstrated significant increases in couple relationship skills and couple satisfaction, as well as decreases in mental health symptoms, from baseline to immediate post-program and from baseline to one-year follow-up.

Conclusion. The study highlights that ELEVATE curriculum is effective in improving relational and individual outcomes at immediate post-program and one-year follow-up regardless of delivery mode. Outcomes were comparable across weekly and monthly formats. High fidelity, participant engagement, and facilitation quality were also evident, indicating consistent implementation across the modalities. These findings underscore the curriculum's adaptability and utility in meeting varied couples' scheduling needs--broadening the implementation options for community-based HMRE.

Impact Evaluation of the Alabama Healthy Marriage and Relationship Education Project in Auburn, AL

I. Introduction

This section gives an overview of the evaluation, including motivation and goals for this study, as well as the impact and implementation research questions.

A. Study overview

Alabama has long experienced high levels of marital and family instability, making it difficult for many individuals and families to achieve economic independence and family cohesion. As a result, children across the state face considerable obstacles to their healthy development and overall well-being. In the 2024 national rankings for child well-being, Alabama ranked 39th out of 50 states (Kids Count, 2024). While this reflects some progress (improving from 45th place the previous year, Kids Count, 2024), significant challenges remain for children in the state. Over two in five (41%) children in Alabama live in single-parent households, compared to 35% nationwide, and one-third reside with caregivers who lack regular employment (Kids Count, 2024). Nearly a quarter (24%) of families with children live below the poverty line, including 10% of married couples with children and 46% of single-parent households (Kids Count, 2024). According to the U.S. Census Bureau, from 2019 to 2023, 15.6% of Alabama's population lived below the poverty line, compared to the national average of 11.1%, ranking Alabama as the seventh poorest state (Flynt, 2023). In connection with these economic and employment stressors, Alabama has consistently exhibited higher divorce rates compared with many other U.S. states. For instance, in 2022, Alabama's divorce rate was 3.2 divorces per 1,000 population, which is above the national average of 2.4 divorces per 1,000 (CDC, 2022). These statistics highlight the ongoing hardships and relationship instability many Alabama couples face in securing a better future for themselves and their children.

Since 2005, because of the Administration for Children and Families' (ACF) efforts to support community-based healthy marriage programs and services, there is evidence from the Alabama Healthy Marriage and Relationship Education program (AHMRE) and others underscoring the value of providing programs that center on healthy marriage and relationship education (HMRE) with additional support services as a key for positively influencing multiple areas of functioning, such as individual well-being, couple relationship, and coparenting functioning (e.g., Adler-Baeder et al., 2022; Hawkins et al., 2022a; 2022b; McGill & Adler-Baeder, 2020; McGill et al., 2021b; Rauer et al., 2014; Hawkins & Ooms, 2012; Markman & Rhoades, 2012; Adler-Baeder et al., 2010), as well as spillover benefits for participants' children (Adler-Baeder et al., 2025; Adler-Baeder et al., 2018). The most recent evaluation of AHMRE's 2015-2020 ACF-funded randomized control trial (RCT) impact study (n = 930 couples), found significant improvements for participants receiving the ELEVATE curriculum compared with a no-program control group on relational- and individual-functioning outcomes up to one-year post-baseline (Adler-Baeder et al., 2022). Further, this prior impact study documented the core assumption in AHMRE's logic model that changes in relationship skills predict the change in couple satisfaction at six-month follow-up (Adler-Baeder et al., 2022).

The AHMRE project historically and currently utilizes a risk and resiliency perspective, assuming that outcomes are based on the interaction of risk and protective factors. Some individuals experience more economic, social, and emotional distress and face challenges that limit the availability of important resources. In turn, according to prevention science (Coie et al., 1993), protective factors, such as

coordinated programs and support services addressing multiple areas of the family system, may serve to promote resilience and positive outcomes.

The current impact study evaluated the relative impact of two HMRE delivery modalities, comparing program participants randomly assigned to either the novel condition - monthly program delivery (i.e., six 2-hour monthly sessions of ELEVATE curriculum + mini-booster retention emails between monthly sessions) or the validated condition - weekly program delivery (i.e., six 2-hour weekly sessions of ELEVATE curriculum). The weekly program condition was the focus of the prior RCT that demonstrated effects compared with a no-program control group (Adler-Baeder et al., 2022). While no known studies have assessed HMRE delivered in monthly sessions, past AHMRE couples and facilitators have requested this delivery format, with couples emphasizing their need to spread out sessions due to demanding family/work schedules and staff speculating that the persistent exposure of multi-month delivery might increase program impacts. In the final year of the prior ACF-funded grant cycle, AHMRE offered a pilot program delivering HMRE curriculum on a monthly (versus weekly) basis, and couples expressed great enthusiasm for this modality. The current impact evaluation represents the first known HMRE evaluation using random assignment procedures to compare program delivery modality, and as such has potential implications for implementation design and policy. Specifically, this study is the first to assess the potential benefits of longer-term connections with a facilitator and class group in a primary workshop delivered over six months.

B. **Primary research questions**

We assessed comparative impact of program condition with short-term (RQs 1-3; baseline to immediate post-program) and long-term (RQs 4-6; baseline to one-year post-baseline) outcomes. Outcomes in both relational (RQs 1, 2, 4, 5) and individual domains (RQs 3, 6) were assessed, in line with the ecological family systems approach (Bronfenbrenner, 1977) and extant literature linking healthy couple functioning to other individual outcomes (e.g., Landolt et al., 2023; Kiecolt-Glaser & Wilson, 2017; Wei et al., 2024), and specifically aligned with AHMRE curriculum content and the project logic model (see Appendix A). We expected the change after program participation in the two randomly assigned program conditions on specified outcomes to differ, with the hypothesis that the monthly program condition would demonstrate stronger effects due to longer duration of exposure to HMRE content.

The primary RQs are as follows:

1. What is the impact of the monthly program condition relative to the weekly program condition on change in couple relationship skills from baseline to immediate post-program?
2. What is the impact of the monthly program condition relative to the weekly program condition on change in couple satisfaction from baseline to immediate post-program?
3. What is the impact of the monthly program condition relative to the weekly program condition on change in mental health symptoms from baseline to immediate post-program?
4. What is the impact of the monthly program condition relative to the weekly program condition on change in couple relationship skills from baseline to 1-year post-baseline?
5. What is the impact of the monthly program condition relative to the weekly program condition on change in couple satisfaction from baseline to 1-year post-baseline?
6. What is the impact of the monthly program condition relative to the weekly program condition on change in mental health symptoms from baseline to 1-year post-baseline?

II. Monthly and Weekly program conditions

This section describes the focal population, the monthly program condition being evaluated (i.e., novel delivery condition), the weekly program condition (i.e., validated delivery condition), and implementation RQs about the delivery of monthly and weekly program conditions.

A. Focal population

The AHMRE project's target population for both monthly and weekly program conditions was adult couples (ages 19 or older) who were in a romantic relationship at enrollment and primarily low-resourced. For details about sample characteristics, see Section IV.B of this report.

B. Description of monthly program condition as intended

The monthly program condition consisted of six monthly sessions of the ELEVATE curriculum, plus mini-booster retention emails between monthly sessions. See Table II.1 for a detailed description of both monthly and weekly conditions.

Components. In this intervention, couples received six monthly class sessions (e.g., first Thursday of each month, etc.), facilitated in a group setting with other couples and two trained facilitators. Once each week between the monthly class sessions (e.g., second, third, and fourth Thursdays of the month, on which monthly class sessions did not occur), all couples in the monthly condition were emailed a "mini-booster" retention email, a short informational email reminding them of the prior class session's content and curriculum-based relationship activities. The "mini-booster" emails that occurred once each week between class sessions were considered an engagement strategy to enhance retention in the monthly classes. There was no variation in the delivery or receipt of mini-boosters and no interactive component with facilitators; thus "mini-booster" emails were not considered a part of dosage. Individual case management was also provided on an as-needed basis. A list of local and national resources was provided to all participants at the first session.

Content. Couples received the *ELEVATE: Taking Your Relationship to the Next Level* curriculum (Futris et al., 2015), which was developed explicitly based on an evidence-informed model including core components of healthy marriage and relationship education programs. ELEVATE is both evidence-informed and evidence-based, with long-standing program impacts in both relationship and individual functioning domains lasting up to one year (Adler-Baeder et al., 2022; McGill et al., 2021b). The curriculum focuses on communication skills, conflict management and problem-solving, stress and anger management, affection and intimacy, self-care, and the effects of distress on relationship functioning. The scripted curriculum utilizes brief informational sessions, group activities, skills practices, videos, and discussion to facilitate healthy relationship skills acquisition and to emphasize the benefits of healthy relationships. A hard copy of the ELEVATE workbook was provided to each enrolled participant.

Dosage and implementation schedule. For this workshop series, the six sessions occurred once a month for two hours per session. All program sites followed strict calendar guidelines and completed fidelity checklists, and there was no variation in content, number, duration, and frequency of sessions.

Delivery mode. All workshop series were delivered at Auburn University (AU) as well as six different partner sites across Alabama: Family Guidance Center of Montgomery, Family Success Center of Etowah County, Circle of Care Center for Families in Valley, Impact Family Counseling of Birmingham, Sylacauga Alliance for Family Enhancement, and Tuscaloosa's One Place. Each session was facilitated by two trained facilitators. At each program site, the 6-session series was delivered either fully virtual or fully in-person. Determinations of whether to offer workshop series fully virtual or fully in-person for each separate cohort were made on a program site level, in collaboration with AU program staff. Various regions of Alabama experienced surges in COVID-19 cases and federal/local lockdown protocols throughout the grant cycle 2020-2025. Program sites followed protocols for virtual or in-person service delivery from the Center for Disease Control and Prevention (CDC), as well as their individual organization leaders and Board members, and they made decisions in partnership with AU program staff to offer AHMRE workshop series virtual or in-person. The evaluation team and AU program staff carefully tracked virtual versus in-person delivery and worked with program sites to ensure a balance of delivery options as much as possible. Except for during one cohort (Feb 2022) for which local/state CDC regulations required all classes to be offered virtually, recruited participants were given the virtual or in-person class option(s) available at their local program site, and if the only option offered at their local program site did not work for them, they were able to select if they wanted to attend a virtual or in-person

workshop series through a different program site. For each cohort, each program site offered both a virtual monthly and weekly workshop series OR an in-person monthly and weekly workshop series, so that couples who selected virtual or in-person delivery could be randomly assigned at each site within program setting they selected (either virtual or in-person) to either a monthly or weekly workshop series. See details about randomization below in Chapter III, Section III.A. Consent and random assignment procedures.

Staff characteristics, education, and training. Most facilitators had backgrounds in family services, and all had a minimum of an Associate's degree. Facilitators were both men and women, with a wide range of backgrounds (3% Asian or Asian American, 32% African American, 65% White or European American), education levels (30% bachelor's degree, 70% Master's or advanced degree), and age (range 22-74 years old). Each facilitator was given a comprehensive 2-day training on the ELEVATE curriculum and the evaluation study prior to implementation. Facilitators also participated in an annual day of refresher training in the curriculum and evaluation study methods, as well as ongoing monthly trainings on topics such as group engagement skills, effective virtual facilitation, current relationship education research, etc.

Importantly, the majority of AHMRE program site staff had prior experience enrolling couples in evaluation studies that used random assignment. They understood the purpose of a rigorous design, necessity of curriculum fidelity, and need for data collection integrity. Program staff provided weekly monitoring and regular technical assistance (virtual or in-person) to all partner staff. In addition, the local evaluation team interacted frequently with partner staff: collecting monthly reporting, conducting biannual evaluation trainings, creating thorough written reports (e.g., enrollment, program/survey completion, sample characteristics), and offering routine technical assistance via virtual meetings. Open communication lines (emails, virtual meetings, phone calls, Box file sharing) were utilized on a weekly basis. The local evaluation team and program staff regularly monitored program delivery fidelity and quality via routine site visits and online fidelity checklists (completed by each facilitator after each of the six sessions in every workshop series). Additionally, the local evaluation team and program staff provided ongoing program development opportunities via scheduled virtual meetings about program implementation, evaluation design, data collection, participant engagement, and evaluation feedback. See Table II.2 for details about staff characteristics, education, and training.

C. Description of the weekly program condition as intended

The weekly program condition consisted of six weekly sessions of the ELEVATE curriculum and has already demonstrated program efficacy in AHMRE's RCT during the previous ACF HMRE grant funding cycle, 2015-2020 (Adler-Baeder et al., 2022; McGill et al., 2021b). See Table II.1 for a detailed description of both monthly and weekly conditions.

Components. In this intervention, couples received a workshop consisting of six weekly classes, facilitated in a group setting with other couples and two trained facilitators. Note that weekly couples did not receive "mini-booster" retention emails. Individual case management was provided on an as-needed basis. A list of local and national resources was provided to all participants at the first class.

Content. The content in the weekly program condition was identical to the monthly program condition. See Section II.B for more details. *See above Section II.B Content.*

Dosage and implementation schedule. For this workshop series, the six sessions occurred once a week for two hours per session. All program sites followed strict calendar guidelines and fidelity checklists, and there was no variation in content, number, duration, and frequency of sessions.

Delivery mode. The delivery mode in the weekly program condition was identical to the monthly program condition. *See above Section II.B Delivery Mode.*

Staff characteristics, education, and training. The staff in the weekly program condition was identical to the monthly program condition. *See above Section II.B Staff characteristics, education, and training.*

Table II.1. Description of the intended monthly and weekly program conditions, including components, curriculum and content, dosage and schedule, delivery, and focal population

Component	Curriculum and content	Dosage and schedule	Delivery	Focal population
		Monthly Program Condition		
Relationship skills workshops	ELEVATE curriculum: communication skills, conflict management and problem-solving, stress and anger management, affection and intimacy, self-care, and the effects of distress on relationship functioning	Six-month workshop, with six two-hour sessions once a month (e.g., first Thursday of each month)	At 7 different program sites across Alabama, two facilitators provided the workshop series either fully virtual or fully in-person	Adult couples (both age 19+), primarily low-resource
Mini-booster retention emails	Reminders about upcoming class dates (for the purpose of retention and engagement for monthly participants), ELEVATE curriculum session summaries (from prior session), relationship skills practice suggestions	Once each week between monthly class sessions (15 mini-booster emails in total: 3 between monthly class session #1-2, 3 between monthly class session #2-3, 3 between class session #3-4, 3 between class session #4-5, and 3 between class session #5-6)	Program staff emailed these scripted emails to participants	Adult couples (both age 19+), primarily low-resource
		Weekly Program Condition		
Relationship skills workshops	ELEVATE curriculum: communication skills, conflict management and problem-solving, stress and anger management, affection and intimacy, self-care, and the effects of distress on relationship functioning	Six-week workshop, with six two-hour sessions once a week	At 7 different program sites across Alabama, two facilitators provide the workshop series either fully virtual or fully in-person	Adult couples (both age 19+), primarily low-resource

Table II.2. Staff characteristics, education, training, and development to support monthly and weekly program components

Component	Staff characteristics, education, and initial training	Ongoing staff training
Monthly Program Condition		
Relationship skills workshops	Facilitators varied in sex, race, age, and relationship status. Most of them had backgrounds in family services and held at least an associate degree. They received 2 days of initial training in the ELEVATE curriculum and the evaluation study.	Facilitators received a day of annual refresher training in the intervention's curriculum and evaluation study methods from program and local evaluation staff. They also received monthly trainings on group engagement skills, effective virtual facilitation, current relationship education research, etc.
Mini-booster retention emails	Program staff, who held at least a bachelor's degree and were licensed as Certified Family Life Educators, wrote these scripted emails based on the ELEVATE curriculum. Program staff received verbal and written guidance as a means for initial training and received regular reminders to send booster emails.	Program staff received 5-7 days of annual refresher training in healthy marriage and relationship education research and programming. They also received monthly trainings on group engagement skills, effective virtual facilitation, current relationship education research, etc.
Weekly Program Condition		
Relationship skills workshops	Facilitators varied in sex, race, age, and relationship status. Most of them had backgrounds in family services and held at least an associate degree. They received 2 days of initial training in the ELEVATE curriculum and the evaluation study.	Facilitators received a day of annual refresher training in the intervention's curriculum and evaluation study methods from program and local evaluation staff. They also received monthly trainings on group engagement skills, effective virtual facilitation, current relationship education research, etc.

D. Implementation research questions about the monthly and weekly program conditions

This section presents the research questions that were examined to understand the implementation of the AHMRE project. We assessed implementation elements of fidelity, dosage, quality, engagement, and context in the two randomly assigned conditions: monthly program (i.e., novel delivery condition) and weekly program (i.e., validated delivery condition). Implementation RQs, grouped by implementation elements, are listed below:

1. **Fidelity**: For each condition, according to facilitator reports after each class session, what average percentage of curriculum content across the six sessions was taught?
2. **Dosage**: For each condition, what percentage of individuals attended: 0 sessions, at least 1 session, 50% or more of the sessions, and all 6 of the sessions offered?
3. **Quality**: For each condition, what is the average participant post-program rating of facilitation quality?
4. **Engagement**: For each condition, what is the average participant post-program rating of self-engagement during the class series?
5. **Context**: For each condition, what percentage of participants reported participating in therapy/counseling (outside of AHMRE program services) during the study period from baseline to one-year post-baseline?

III. Study design

This section describes the sample formation, research design, and data collection for both the impact and implementation analyses. It also describes the methods used to analyze the data as part of the study.

A. Evaluation enrollment and assignment to study conditions

Recruitment and study sample enrollment targets. Recruitment and enrollment procedures were identical for both monthly and weekly conditions. Adult couples for both program conditions were recruited into the impact study in seven separate cohorts, which were named by the month and year in which their workshop series began: Oct 2021, Feb 2022, Apr 2022, Oct 2022, Feb 2023, May 2023, and Oct 2023.

Approximately a month before the beginning of programming for each cohort, recruitment across AU and the six partner sites began. These program sites (seven in total) recruited couples from 50 Alabama counties (i.e., Autauga, Baldwin, Barbour, Bibb, Bullock, Butler, Calhoun, Chambers, Cherokee, Chilton, Choctaw, Clark, Clay, Coffee, Coosa, Cullman, Dale, Dallas, Dekalb, Elmore, Etowah, Geneva, Hale, Henry, Houston, Jefferson, Lamar, Lauderdale, Lee, Limestone, Lowndes, Macon, Madison, Marengo, Marshall, Mobile, Monroe, Montgomery, Pickens, Pike, Randolph, Russell, Shelby, St. Clair, Sumter, Talladega, Tallapoosa, Tuscaloosa, Washington, and Wilcox counties). Recruitment of couples into the impact evaluation occurred through distribution of flyers (approved by the AU Institutional Review Board; AU IRB) at local community agencies (e.g., family resource centers, churches, libraries, gyms, childcare facilities) and posted online via the program sites' social media accounts. Fliers directed interested participants to call or email AHMRE recruitment staff at local program sites, who would then schedule an immediate or subsequent 15-minute recruitment phone call (using an IRB-approved recruitment script). As another option, some program sites included QR codes on recruitment fliers to a form for interested participants to fill out basic contact information and receive a return phone call or email from AHMRE recruitment staff. During the 15-min recruitment phone call, interested participants were informed about the enrollment criteria, evaluation study, surveys, and randomization to either monthly or weekly 6-session relationship education workshop series (with a 50/50 chance of being randomized into either condition). If interested participants met the criteria and wanted to enroll, the AHMRE recruitment staff filled out the nFORM Application Form with the interested participants on the phone. (nFORM [Information, Family Outcomes, Reporting, and Management] is a secure, federally managed participant management and data collection system used by ACF-funded HMRE grantees to track enrollment, service delivery, and outcome measures for program evaluation and performance reporting.) Program

sites also provided incentives (e.g., gift cards, compensation) to previous participants who referred new couples, and at the end of each workshop series, facilitators emphasized the opportunity for current participants to refer new couples to enroll in an upcoming cohort.

By October 2023, AHMRE had successfully enrolled in the program and randomly assigned more than the targeted sample (targeted sample n = 1,320 couples; enrolled and randomly assigned sample n = 1,422 couples. The final analytic sample size of randomly assigned couples in which both partners consented and both completed the baseline survey is 1,120 couples (see Appendix B, CONSORT diagram Figures B.1 and B.2), with 561 randomly assigned to the weekly program condition and 559 to the monthly condition.

Power analyses were conducted to assess the sensitivity to detect effects with the current sample size for the experimental conditions using the "powerlmm" package in R (Magnusson, 2017), which is used for power calculations for longitudinal multilevel models (see chapter IV, section IV.C Modeling approach for planned analyses). Analyses in R indicated there is more than adequate power (96%) to conduct longitudinal analyses with multilevel models over two time points with 20% estimated overall attrition, assuming a .50 correlation between measures over time, and detect a small effect (i.e., as low as f =.17), which is standard in HMRE evaluation studies, using a total sample size of 1,120 couples/2,240 individuals.

Participant eligibility criteria. Participant eligibility criteria were the same for both monthly and weekly program conditions. Participants were eligible to participate if they were adults (age 19 or older) and in a romantic couple relationship. There were no additional criteria for recruiting and selecting the sample beyond the eligibility criteria.

Consent and random assignment process. The AU IRB approved all study procedures originally on 1/18/21 with no expiration date (Protocol #21-006 EP 2101) and two subsequent revisions to the original protocol were approved on 10/17/21 and 7/12/22, with no expiration dates. During the startup period of this grant cycle, ACF also approved the following consent and random assignment procedures. These ordered steps were followed for each cohort:

1. **Recruitment Period**: During the 4-week recruitment period, trained AHMRE staff at each program site conducted intake phone calls with potential participants, using an IRB-approved recruitment script to inform them of the nature of the impact evaluation and random assignment, with an equal (50/50) chance to be placed in either of two program conditions (i.e., monthly class delivery or weekly class delivery). If during the intake call, the potential participants declined participation in the local impact evaluation, they were given information about when the classes would be available for participation without committing to the impact study (i.e., Feb 2024).
2. **Enrollment Into nFORM & Impact Study Participant Tracking System**: If willing to proceed, program site staff enrolled interested couples in nFORM, using the Application Form. Participants' contact information was then uploaded into a sophisticated, secure participant tracking system, via Ripple Science online software (https://www.ripplescience.com). The evaluation team managed the Ripple Science online databases linking identifying information (i.e., names, addresses, phone numbers) to unique IDs (individual and couple) created by nFORM. Access to Ripple databases was limited to only trained, eligible staff.
3. **Informed Consent & Local Evaluation Baseline Survey Completion**: Immediately after the recruitment period ended, the evaluation team used Ripple to email and text the informed consent letter (approved by AU IRB) and the AHMRE local evaluation baseline survey (T1) via electronic links to all enrolled couples. The local evaluation surveys were distinct from nFORM surveys. Participants were given a 10-day window to complete the informed consent and T1 survey. While completing the T1 survey, participants were blind to their random assignment.
4. **Rationale for Randomly Assigning All Enrolled Couples**: After the T1 survey window closed, the local evaluator randomly assigned all enrolled couples to one of the two program conditions (whether or not they completed the informed consent or T1 survey). This approach was in consultation with the ACF Federal Project Specialist and allowed AHMRE to provide programming to all interested couples without condition on participation in the impact study.

This approach was also in line with IRB and Government Accountability Office (GAO) guidelines for voluntary participation in research. Out of the full sample of randomly assigned couples (n = 1,422 couples), about 79% had both partners complete an informed consent and T1 survey before random assignment (n = 1,120 couples). An additional 202 individuals completed an informed consent and T1 survey before random assignment, but their partners did not complete consent and T1, so they were not included in the analytic sample for this report (see Appendix B, CONSORT diagram Figures B.1 and B.2).

5. **Random Assignment Procedures:** For each impact study cohort, the T1 survey window closed on a Sunday at midnight. The local evaluator used block random assignment procedures per best practices with multi-site randomized trials (James Bell Associates, 2007); these procedures promoted equivalence in baseline numbers and participant demographics at each program site across monthly and weekly conditions. To conduct random assignment, all enrolled couples were systematically divided into blocks by program site. Random assignment occurred at the couple level (separately in each program site block and within program setting type [virtual/in-person] if both virtual and in person classes were offered simultaneously at a program site), and each couple had a 50% chance of being assigned to either the monthly or weekly program condition. The local evaluator used true random assignment (using a random number generator in SPSS statistical software) for each unique couple ID to the two program conditions by program site. No subsampling occurred after randomization.
6. **Participants Informed of Random Assignment**: On the Monday morning after the baseline survey window closed the previous night, the evaluation team emailed and texted all enrolled couples their random assignment details and class schedule. The evaluation team also emailed encrypted class rosters for monthly/weekly program conditions to the program site staff, and the program site staff followed up with another email, text, or phone call to each enrolled couple, confirming class details. All enrolled participants (whether or not they completed an informed consent and T1 survey) were informed of their random assignment details after the T1 survey window closed. Participants were blind to their random assignment while completing the T1 survey.
7. **Workshop Series and nFORM Entrance/Exit Surveys**: After receiving random assignment details, all enrolled participants had the opportunity to attend the relevant six monthly/weekly workshop sessions. Participants in both program conditions were also asked to complete the nFORM Applicant Characteristics Form and nFORM Entrance survey immediately prior to the first class session and the nFORM Exit survey immediately following the last class session. The data from nFORM Applicant Characteristics, Entrance and Exit surveys were not included in this local impact evaluation report.
8. **Consent Opportunities at Local Evaluation Follow-Up Surveys:** Since this research study used an Intent-to-Treat design, enrolled participants who did not complete an informed consent and T1 survey were still offered the opportunity to complete an informed consent along with a local-evaluation follow-up survey at later time points (i.e., 3-month, 6-month, and 1-year surveys post-baseline). For the current evaluation report, participants who did not complete a T1 survey were listwise deleted from the analytic sample.

B. **Outcome measures**

The AHMRE local evaluation surveys collected outcome measures for all primary research questions. See Table III.1 for a description of outcome measures (i.e., citation, items, values, internal consistency), source, and timing of measure collection. See Appendix B, Table B.3 for all outcome measure items and Appendix D, Table D.1 for internal consistency of outcome measures across time points.

C. **Implementation measures**

See Table III.2 for a description of all implementation measures, and Appendix B, Table B.5 for all implementation measure items.

Table III.1. Outcome measures used to answer primary research questions of the impact analysis

RQ #	Outcome	Description of the outcome measure and its properties	Source	Timing of measure
1	Couple Relationship Skills	Couple relationship skills were measured by the Couple Relationship Skills Inventory (Adler-Baeder et al., 2022), a 32-item scale (value range 1=”very strongly disagree” to 7=”very strongly agree"), assessing conflict management, communication, commitment, care for self, knowledge of partner, etc. Prompt and example item: “Please rate how strongly you disagree or agree with each of the following: I commit effort every day to making my relationship work.” Scores were calculated as a mean composite at the individual participant level, and higher scores indicate better couple relationship skills. The published measure has excellent psychometrics, and scale reliability in the current sample was high (range across time points α = .90-.93).	Local evaluation surveys (Qualtrics)	T1 & Immediate Post-Program (i.e., T2 for weekly condition; T3 for monthly condition)
2	Couple Satisfaction	Couple satisfaction was measured by the Couple Satisfaction Index (Funk & Rogge, 2007), a 4-item scale (value range for item #1: 1=”extremely unhappy” to 7=”perfect”; value range for items #2-3: 1=”not at all true” to 6=”completely true”), assessing happiness, comfort, and satisfaction in the relationship. Prompt and example item: “Think about your CURRENT couple relationship, as you select the best answer for the following questions: Please indicate the degree of happiness, all things considered, of your relationship.” Scores were calculated as a sum composite at the individual participant level, and higher scores indicate greater couple satisfaction. The published measure has excellent psychometrics, and scale reliability in the current sample was high (range across time points α = .88-.90).	Local evaluation surveys (Qualtrics)	T1 & Immediate Post-Program (i.e., T2 for weekly condition; T3 for monthly condition)
3	Mental Health Symptoms	Mental health symptoms were measured by the Kessler Psychological Distress Scale (Kessler et al., 2002), a 10-item scale (value range 1=”none of the time” to 5=”all of the time”), assessing nervousness, hopelessness, depressed mood, feelings of worthlessness, sadness, etc. Prompt and example item: “Please select how often you have felt each of the following in the past 30 days: Did you feel hopeless?” Scores were calculated as a mean composite at the individual	Local evaluation surveys (Qualtrics)	T1 & Immediate Post-Program (i.e., T2 for weekly condition; T3 for monthly condition)

RQ #	Outcome	Description of the outcome measure and its properties	Source	Timing of measure
		participant level, and higher scores indicate greater mental health symptoms. The published measure has excellent psychometrics, and scale reliability in the current sample was high (range across time points α = .91-.93).		
4	Couple Relationship Skills	See above description of outcome measure for RQ 1	Local evaluation surveys (Qualtrics)	T1 & T4
5	Couple Satisfaction	See above description of outcome measure for RQ 2	Local evaluation surveys (Qualtrics)	T1 & T4
6	Mental Health Symptoms	See above description of outcome measure for RQ 3	Local evaluation surveys (Qualtrics)	T1 & T4

Table III.2. Measures used to address implementation research questions

Focus	Research question	Measures
Fidelity	1. For each condition, according to facilitator reports after each class session, what average percentage of curriculum content was taught?	Percentages of curriculum content taught were calculated for each class series as the average number of curriculum activities checked on the fidelity checklist (after each class session of each workshop series) out of the total number of planned curriculum activities. Next, percentages for all workshop series in each condition were averaged, to indicate how much of the intended program was actually delivered.
Dosage	2. For each condition, what percentage of individuals attended: 0 sessions, at least 1 session, 50% or more of the sessions, and all 6 of the sessions offered?	For each condition, we calculated frequencies (percentage) of individuals who attended 0 sessions, 1 session, 50% or more of the sessions, or all 6 sessions (from nFORM session attendance records).
Quality	3. For each condition, what is the average participant post-program rating of facilitation quality?	For each condition, mean composites were calculated for participant-reported facilitation quality, a 5-item subscale of the Classroom Education Environment Scale (CEES; Adler-Baeder et al., 2023) on immediate post-program surveys (T2 for weekly participants; T3 for monthly participants). Higher scores indicate better facilitation quality (e.g., explaining material, managing time, using appropriate examples). This subscale had excellent reliability in the current sample (α range = .95-.96).
Engagement	4. For each condition, what is the average participant post-program rating of self-engagement during the class series?	For each condition, mean composites were calculated for participant-reported individual engagement, a 3-item subscale of the CEES (Adler-Baeder et al., 2023) on immediate post-program surveys (T2 for weekly participants; T3 for monthly participants). Higher scores indicate greater self-engagement in the class series (e.g., sharing in group discussions, feeling connected to other participants). This subscale had excellent reliability in the current sample (α range = .84-.86).
Context	5. For each condition, what percentage of participants reported participating in therapy/counseling (outside of AHMRE program services) during the study period from baseline to one-year post-baseline?	For each condition, we calculated frequencies (percentage) of participant-reported involvement in therapy/counseling (outside of AHMRE program services) from T4 surveys (i.e., 1-year post-baseline). Scores indicate percentages of participants in each condition who reported utilizing therapy/counseling over the prior year.

D. Data collection

This section details the data collection procedures for the impact analysis (both monthly and weekly program conditions), including engagement strategies used to retain participants. Next we describe data collection procedures for the implementation analysis.

1. Impact data collection

All impact data collection procedures were approved in the AU IRB study protocol (see above Section III.A). Data collection procedures for randomly assigned participants in both monthly and weekly program conditions were identical. Participants were asked to complete AHMRE local evaluation surveys at baseline (T1) and three follow-up time points: 3-month (T2), 6-month (T3), and 1-year post-baseline (T4). AHMRE local evaluation surveys were completed on participants' own devices via electronic links to an AHMRE local evaluation survey (via Qualtrics) texted and/or emailed to them. In a few instances as needed, program staff scheduled a date for participants to complete in-person electronic surveys at program sites via an iPad or computer belonging to the program site. Each separate T1-T4 local evaluation survey took approximately one hour to complete. All participants in the study were compensated for the completion of each local evaluation survey, $40 at each time point.

Data collection was primarily conducted by the local evaluation team. The team texted and emailed electronic survey links and completion reminders to participants over each survey window period (i.e., 10 days for T1, 3-4 weeks for T2-T4). Survey completion in Qualtrics was monitored by local evaluation staff, and completion was marked in Ripple. Survey compensation lists were generated by downloading survey completion from Ripple, and AU program staff initiated and mailed $40 paper checks via the AU Accounts Payable department. For the handful of cases when surveys were completed in-person at program sites, the relevant staff had been thoroughly trained by the local evaluator to comply with proper data collection procedures.

The AHMRE local evaluation surveys collected measures of the outcomes of interest for primary RQs. For RQs 1, 2, and 3, data were utilized from the following surveys: T1 and T2 (for weekly condition participants), and T1 and T3 (for monthly condition participants). For RQs 4, 5, and 6, data were utilized from T1 and T4 surveys (for both monthly and weekly condition participants). Participants in both program conditions were also asked to complete nFORM Applicant Characteristics and Entrance surveys immediately prior to their first class session and nFORM exit surveys immediately following their last class session. The data from nFORM Applicant Characteristics, Entrance and Exit surveys were not included in this impact evaluation report. See Appendix B, Table B.1 for key features of data collection for the impact analysis.

Tracking participants and reducing attrition. A common anticipated challenge for impact studies is low response rate, particularly among low-resource participants (see Hawkins & Erickson, 2015). To promote program engagement and optimize survey response rates, a variety of modes of contact (i.e., phone, email, mailing address) were obtained from participants at enrollment. During the workshop series, reminder emails/texts about class times were sent at specified intervals (several times throughout the month for the monthly program condition, weekly for the weekly program condition).

For each of the AHMRE local evaluation surveys (T1, T2, T3, T4), the local evaluation team sent "save the date" notices (via email and/or text) one or two days prior to emailing survey links, reminding participants about the upcoming survey and asking about any changes to contact info. After emailing/texting the initial survey link, the local evaluation team sent reminder texts/emails/phone calls during the "open" window for survey completion. Additionally, during each "open" window for survey completion, the evaluation team sent encrypted lists of participants who had not yet completed the survey to partner staff, who directly contacted relevant participants to remind them to complete before the survey window closed. Several additional engagement strategies were used to retain participants up to the 1-year T4 time point: frequent social media postings about healthy relationship skills and relevant resources, birthday notes emailed to participants on their actual birthday, and anniversary cards emailed to participants on a date they signified was important to their relationship (obtained on the T1 survey).

In order to monitor survey completion and monitor attrition, we used Ripple software to organize and track participants by program site. We calculated and closely monitored overall attrition of the entire sample as well as differential attrition between the two program conditions (at both individual- and couple-

levels), and ongoing efforts were made to minimize differential attrition in accordance with the What Works Clearinghouse (WWC; 2022) cautious boundary. In tracking survey completion rates and attrition, the Local Evaluator and AHMRE program staff used Continuous Quality Improvement procedures to brainstorm and implement adjustments to participant enrollment and engagement strategies. See Appendix C, Tables C.1a and C.1b for overall and differential attrition rates at the individual- and couple-levels.

2. Implementation data collection

A variety of data sources were used to address the implementation RQs (see Appendix B, Table B.2 for an overview of data sources, timing of data collection, and responsible parties for implementation measures). To address **Fidelity RQ #1,** lead facilitators completed online fidelity checklists after each of the six class sessions, indicating which of the required curriculum elements were taught during that class session. During the duration of the impact study, AHMRE offered 112 class series (i.e., 56 monthly condition; 56 weekly condition), and facilitators completed 112 fidelity checklists, for a response rate of 100%. For **Dosage RQ #2**, program staff tracked participant attendance in nFORM for all sessions during each of the 6-session workshops for both monthly and weekly conditions. Participant local evaluation survey data was used for the final three implementation RQs. On the immediate post-program survey (i.e., T2 for weekly condition; T3 for monthly condition), participants completed two subscales of the Classroom Education Environment Scale (CEES; Adler-Baeder et al., 2023): the 5-item facilitation quality subscale was used for **Quality RQ #3**, and the 3-item individual engagement subscale was used for **Engagement RQ #4**. Finally, for **Context RQ #5**, monthly and weekly participants also responded to an item on the final T4 survey asking about therapy/counseling they may have utilized during the study period (from baseline to 1-year post-baseline) other than the AHMRE 6-session relationship education classes to which they were randomly assigned.

IV. Analytic methods

This chapter describes the analytic sample, measures, baseline equivalence, and sample characteristics. It also describes the analytic methods for estimating program impacts on the primary research questions.

A. Analytic sample

We first describe condition crossover, attrition, missing data, and determining the analytic samples for each primary RQ. For details about data preparation and cleaning, reference Appendix D.

Condition Crossover. The local evaluation team and program staff monitored participant enrollment and class attendance via nFORM and frequent communication with partner staff. Enrolled participants were tracked closely and informed that they could not switch program conditions after receiving their random assignment. Participants checked in at each class session to ensure they attended the workshop series to which they were assigned. If they choose not to participate in their assigned workshop series, they were given the opportunity to continue completing all local evaluation follow-up surveys and they were offered a chance to participate in programming at the end of the evaluation study (beginning Feb 2024). Any randomly assigned participants who attempted to re-enroll were informed that they were not eligible to participate again (until the impact study concluded in Feb 2024). Throughout the duration of the impact study, there was no condition crossover; that is, no participants randomly assigned to the monthly condition participated in the weekly condition, and vice versa.

Missing data. We examined patterns of missingness at the survey- and item-levels. First for the overall analytic sample, couples in which both partners did not have T1 survey data were listwise deleted. For each primary RQ analytic sample, we assessed missingness at the item-level for each relevant outcome. Since all RQs used multi-item composite outcomes, if a case was missing 20% or fewer items, we kept this case and imputed missing items using regression-based imputation methods in R online statistical software (based on technical assistance provided to grant recipients; see Appendix D for details about imputing missing data). If a case was missing greater than 20% of items, we listwise deleted the relevant case. We did this process separately for each Primary RQ and relevant composite outcome.

Since this study had low attrition at the couple level, the analytic approach used hierarchical linear modeling and included all couples with at least one partner responding at the relevant follow-up for Primary RQ analytic samples. We imputed missing data using regression-based imputation methods in R online statistical software (see Appendix D for details about imputing missing data). Sensitivity analyses were conducted with complete-case analytic samples to ensure robustness of results with imputed data. See Appendix D, Table D.4 for detailed numbers and rates of missingness on each outcome for each RQ analytic sample, separately by condition.

Determining analytic samples for each primary RQ. See Table IV.1a for cluster sample sizes and Table IV.1b for individual sample sizes by condition and primary outcome; see Appendix B for CONSORT diagram Figures B.1 and B.2. The full analytic sample included all couples in which both partners completed T1 and at least one partner completed T4 surveys (total analytic sample *n* = 989 couples; i.e., 490 monthly condition / 499 weekly condition). Out of that full analytic sample, we separately determined each RQ analytic sample, considering survey- and item-level missingness for relevant composite outcome:

1. The analytic sample included all couples in which both partners had >=80% items on the Couple Relationship Skills Inventory at T1 and at least one partner had >=80% items on the Couple Relationship Skills Inventory at immediate post-program (i.e., T2 for weekly participants; T3 for monthly participants).
2. The analytic sample included all couples in which both partners had >=80% items on the Couple Satisfaction Index at T1 and at least one partner had >=20% items on the Couple Satisfaction Index at immediate post-program (i.e., T2 for weekly participants; T3 for monthly participants).
3. The analytic sample included all couples in which both partners had >=80% items on the Kessler Psychological Distress Scale at T1 and at least one partner had >=80% items on the Kessler Psychological Distress Scale at immediate post-program (i.e., T2 for weekly participants; T3 for monthly participants).
4. The analytic sample included all couples in which both partners had >=80% items on the Couple Relationship Skills Inventory at T1 and at least one partner had >=80% items on the Couple Relationship Skills Inventory at T4.
5. The analytic sample included all couples in which both partners had >=80% items on the Couple Satisfaction Index at T1 and at least one partner had >=80% items on the Couple Satisfaction Index at T4.
6. The analytic sample included all couples in which both partners had >=80% items on the Kessler Psychological Distress Scale at T1 and at least one partner had >=80% items on the Kessler Psychological Distress Scale at T4.

Attrition. We used an Intent-to-Treat design, which is the most rigorous evaluation approach and yields the most reliable efficacy results for testing a program in a “real world” setting (Weiss & Jacobs, 2008; Wood et al., 2014). Group comparisons were based on random assignment to, rather than participation in, the two program conditions (monthly or weekly program delivery). Before imputing missing data, we first tracked overall and differential attrition in the overall sample at the couple- and individual-levels. In the overall sample at the couple level, our overall attrition was 12%; our differential attrition rates (ranging from 0-3.8%) were all below the WWC low attrition threshold of 6.2% (i.e. cautious boundary according to WWC Standards, 2022). In the overall sample at the individual level, our overall attrition was 21%; our differential attrition rates (ranging from 0-2.4%) were all below the WWC low attrition threshold of 5.3% (i.e., cautious boundary according to WWC Standards, 2022). We also assessed attrition in each of the final analytic samples for each RQ. **The final analytic samples for each RQ classified as low attrition (according to WWC Standards, 2022)**, with overall attrition ranging from 12.5%-18.2% and differential attrition ranging from 0.1%-3.1%, all within the corresponding WWC low attrition thresholds. See Appendix C, Table C.1a for details about attrition in the overall sample and Table C.1b for details about attrition in the final analytic samples for each RQ.

Table IV.1a. Cluster sample sizes by condition for each time point and research question

Number of:	MONTHLY sample size	WEEKLY sample size	TOTAL sample size	TOTAL response rate	MONTHLY response rate	WEEKLY response rate
Clusters						
Clusters: Contributed at least one individual at T1	559	561	1120	100.0%	100.0%	100.0%
Clusters: Contributed at least one individual at T2 (3-month)	526	537	1063	94.9%	94.1%	95.7%
Clusters: Contributed to the impact analysis of outcome at T2 (3-month), accounting for item nonresponse and any other analysis restrictions (out of T4 sample)						
RQ 1 Outcome: CRSI		472				84.1%
RQ 2 Outcome: CSI		473				84.3%
RQ 3 Outcome: K10		488				87.0%
Clusters: Contributed at least one individual at T3 (6-month)	491	514	1005	89.7%	87.8%	91.6%
Clusters: Contributed to the impact analysis of outcome at T3 (6-month), accounting for item nonresponse and any other analysis restrictions (out of T4 sample)						
RQ 1 Outcome: CRSI	455				81.4%	
RQ 2 Outcome: CSI	454				81.2%	
RQ 3 Outcome: K10	469				83.9%	
Clusters: Contributed at least one individual at T4 (1-year)	490	499	989	88.0%	87.7%	88.4%
Clusters: Contributed to the impact analysis of outcome at T4 (1-year), accounting for item nonresponse and any other analysis restrictions						
RQ 4 Outcome: CRSI	461	455	916	81.8%	82.5%	81.1%

Number of:	MONTHLY sample size	WEEKLY sample size	TOTAL sample size	TOTAL response rate	MONTHLY response rate	WEEKLY response rate
RQ 5 Outcome: CSI	458	460	918	82.0%	81.9%	82.0%
RQ 6 Outcome: K10	488	492	980	87.5%	87.3%	87.7%

Notes: n.a. = not applicable.

Table IV.1b. Individual sample sizes by condition for each time point and research question

Number of:	MONTHLY sample size	WEEKLY sample size	TOTAL sample size	TOTAL response rate	MONTHLY response rate	WEEKLY response rate
Individuals in non-attriting clusters						
Individual: Who consented	1118	1122	2240	100.0%	100.0%	100.0%
Individual: Contributed a T1 survey	1118	1122	2240	100.0%	100.0%	100.0%
Individual: Contributed to T2 (3-month)	989	1020	2009	89.7%	88.5%	90.9%
Individual: Contributed to the impact analysis of outcome at T2 (3-month), accounting for item nonresponse and any other analysis restrictions (out of T4 sample)						
RQ 1 Outcome: CRSI		944				84.1%
RQ 2 Outcome: CSI		946				84.3%
RQ 3 Outcome: K10		976				87.0%
Individual: Contributed to T3 (6-month)	924	950	1874	83.7%	82.6%	84.7%
Individual: Contributed to the impact analysis of outcome at T3 (6-month), accounting for item nonresponse and any other analysis restrictions (out of T4 sample)						
RQ 1 Outcome: CRSI	910				81.4%	
RQ 2 Outcome: CSI	908				81.2%	
RQ 3 Outcome: K10	938				83.9%	
Individual: Contributed to T4 (1-year)	882	898	1780	79.5%	78.9%	80.0%
Individual: Contributed to the impact analysis of outcome at T4 (1-year), accounting for item nonresponse and any other analysis restrictions						
RQ 4 Outcome: CRSI	922	910	1832	81.8%	82.5%	81.1%
RQ 5 Outcome: CSI	916	920	1836	82.0%	81.9%	82.0%

Number of:	MONTHLY sample size	WEEKLY sample size	TOTAL sample size	TOTAL response rate	MONTHLY response rate	WEEKLY response rate
RQ 6 Outcome: K10	976	984	1960	87.5%	87.3%	87.7%

Notes: n.a. = not applicable.

B. **Baseline equivalence and sample characteristics**

This section describes baseline equivalence and sample characteristics of the monthly and weekly program conditions.

Assessment of Baseline Equivalence. Although this classified as a low attrition study at both individual- and couple-levels according to WWC Standards (2022), for full transparency and to determine covariates in primary analyses, we assessed baseline equivalence between conditions separately in the analytic samples for each primary RQ, separately for women and men (due to non-independent couple-level data). We assessed the equivalence of demographic characteristics (i.e., race, age, education, married status, and parent status) and baseline measures of all relevant outcomes for primary RQs (i.e., couple relationship skills, couple satisfaction, and mental health symptoms) on non-imputed data. Statistical tests were conducted (i.e., *t*-tests of between-group differences and Hedge's *g* effect sizes for continuous variables, and chi-square difference tests and Cox's index effect size for binary variables) to determine whether there were differences between respondents in the two program conditions. When statistical tests of differences between conditions had absolute value effect sizes equal to or less than .05, these demographics or outcomes were considered equivalent across conditions (according to WWC Standards, 2022). When statistical tests of differences between conditions had absolute value effect sizes greater than .05 but less than .25, we adjusted for these effect size differences by including these demographics as covariates in primary analyses (in line with WWC Standards, 2022). No differences between conditions had absolute value effect sizes greater than .25.

Table IV.2a Summary statistics of key baseline measures and baseline equivalence across study conditions, for individuals in the sample for RQ1

Baseline measure	MONTHLY mean or % (women | men)	MONTHLY standard deviation (women | men)	WEEKLY mean or % (women | men)	WEEKLY standard deviation (women | men)	Monthly & Weekly difference in means or % (women | men)	*p*-value of difference in means (women | men)	Effect size abs. value (women | men)
Race							
White/European American	43.3% | 43.5%	n.a.	46.2% | 43.0%	n.a.	2.9% | 0.5%	.36 | .86	.05 | .02
Other	56.7% | 56.5%		53.8% | 57.0%				
Age	36.67 | 38.85	9.94 | 11.06	35.87 | 38.40	10.20 | 11.18	.80 | .45	.23 | .55	.08^ | .04
Education							
No 4-year college degree	48.0% | 61.3%	n.a.	49.2% | 61.2%	n.a.	1.2% | 0.1%	.71 | .97	.05 | .00
4-year college degree or higher	52.0% | 38.7%		50.8% | 38.8%				
Married Status							
Married	63.8% | 65.8%	n.a.	64.8% | 66.2%	n.a.	1.0% | 0.4%	.75 | .92	.00 | .00
Non-married	36.2% | 34.2%		35.2% | 33.8%				
Parent Status							
Parent			75.8% | 76.4%		0.2% | 0.9%	.96 | .75	

Baseline measure	MONTHLY mean or % (women | men)	MONTHLY standard deviation (women | men)	WEEKLY mean or % (women | men)	WEEKLY standard deviation (women | men)	Monthly & Weekly difference in means or % (women | men)	*p*-value of difference in means (women | men)	Effect size abs. value (women | men)
Non-parent	75.6% | 75.5% 24.4% | 24.5%	n.a.	24.2% | 23.6%	n.a.			.07^ | .07^
CRSI Relationship Skills *(mean composite; range 1 to 7)*	5.02 | 5.15	.74 |.70	5.00 | 5.08	.69 |.68	.02 | .07	.76 | .13	.02 | .10^
Sample size	480 | 461	n.a.	467 | 442	n.a.	n.a.	n.a.	n.a.

Notes: n.a. = not applicable. See Table IV.3 for description of demographic covariates. CRSI=Couple Relationship Skills Inventory. Analyses were run separately for women and men. For tests of differences, chi-square tests were used for binary variables and *t*-tests were used for continuous variables. For effect sizes, Cox's index was used for binary variables and Hedge's *g* for continuous variables. Baseline equivalence tests were computed with non-imputed data. Effect sizes of baseline differences > .05 are marked with ^.

Table IV.2b Summary statistics of key baseline measures and baseline equivalence across study conditions, for individuals in the sample for RQ2

Baseline measure	MONTHLY mean or % (women | men)	MONTHLY standard deviation (women | men)	WEEKLY mean or % (women | men)	WEEKLY standard deviation (women | men)	Monthly & Weekly difference in means or % (women | men)	*p*-value of difference in means (women | men)	Effect size abs. value (women | men)
Race							
White/European American	43.3% | 43.6%	n.a.	45.7% | 42.4%	n.a.	2.4% | 1.2%	.46 | .71	.05 | .02
Other	56.7% | 56.4%		54.3% | 57.6%				
Age	36.63 | 38.82	9.98 | 11.10	35.85 | 38.40	10.17 | 11.13	.78 | .42	.24 | .57	.08^ | .04
Education							
No 4-year college degree	48.1% | 61.5%	n.a.	49.1% | 61.0%	n.a.	1.0% | 0.5%	.76 | .90	.05 | .03
4-year college degree or higher	51.9% | 38.5%		50.9% | 39.0%				
Married Status							
Married	63.7% | 65.8%	n.a.	64.7% | 66.0%	n.a.	1.0% | 0.2%	.77 | .94	.00 | .03
Non-married	36.3% | 34.2%		35.3% | 34.0%				
Parent Status							
Parent		n.a.		n.a.	.2% | 1.0%	.93 | .72	

Baseline measure	MONTHLY mean or % (women | men)	MONTHLY standard deviation (women | men)	WEEKLY mean or % (women | men)	WEEKLY standard deviation (women | men)	Monthly & Weekly difference in means or % (women | men)	*p*-value of difference in means (women | men)	Effect size abs. value (women | men)
Non-parent	75.8% | 75.7% 24.2% | 24.3%		76.0% | 76.7% 24.0% | 23.3%				.07^ | .10^
CSI Couple Satisfaction *(sum composite; range 4 to 25)*	18.15 | 19.15	4.83 | 4.48	18.23 | 19.10	4.87 | 4.16	.07 | .06	.79 | .84	.02 | .01
Sample size	466 | 441	n.a.	481 | 462	n.a.	n.a.	n.a.	n.a.

Notes: n.a. = not applicable. CSI=Couple Satisfaction. Analyses were run separately for women and men. For tests of differences, chi-square tests were used for binary variables and *t*-tests were used for continuous variables. For effect sizes, Cox's index was used for binary variables and Hedge's *g* was used for continuous variables. Baseline equivalence tests were computed with non-imputed data. Effect sizes of baseline differences > .05 are marked with ^.

Table IV.2c Summary statistics of key baseline measures and baseline equivalence across study conditions, for individuals in the sample for RQ3

Baseline measure	MONTHLY mean or % (women | men)	MONTHLY standard deviation (women | men)	WEEKLY mean or % (women | men)	WEEKLY standard deviation (women | men)	Monthly & Weekly difference in means or % (women | men)	*p*-value of difference in means (women | men)	Effect size abs. value (women | men)
Race							
White/European American	43.1% | 43.7%	n.a.	45.7% | 42.2%	n.a.	2.6% | 1.5%	.42 | .65	.07^ | .02
Other	56.9% | 56.3%		54.3% | 57.8%				
Age	36.42 | 38.68	9.92 | 11.05	35.86 | 38.39	10.23 | 11.15	.56 | .29	.39 | .69	.06^ | .03
Education							
No 4-year college degree	48.0% | 61.9%	n.a.	50.7% | 62.4%	n.a.	2.7% | 0.5%	.40 | .88	.07^ | .00
4-year college degree or higher	52.0% | 38.1%		49.3% | 37.6%				
Married Status							
Married	61.5% | 63.9%	n.a.	63.2% | 64.7%	n.a.	1.7% | 0.8%	.59 | .79	.03 | .03
Non-married	38.5% | 36.1%		36.8% | 35.3%				
Parent Status							
Parent		n.a.		n.a.	.2% | .6%	.93 | .84	.03 | .07^

Baseline measure	MONTHLY mean or % (women | men)	MONTHLY standard deviation (women | men)	WEEKLY mean or % (women | men)	WEEKLY standard deviation (women | men)	Monthly & Weekly difference in means or % (women | men)	*p*-value of difference in means (women | men)	Effect size abs. value (women | men)
Non-parent	76.0% | 75.7% 24.0% | 24.3%		75.8% | 76.3% 24.2% | 23.7%				
Kessler Mental Health symptoms *(mean composite; range 1 to 5)*	2.19 | 2.13	.76 |.76	2.25 | 2.08	.85 | .79	.06 | .05	.21 | .30	.08^ | .07^
Sample size	483 | 454	n.a.	497 | 476	n.a.	n.a.	n.a.	n.a.

Notes: n.a. = not applicable. See Table IV.3 for description of demographic covariates. Analyses were run separately for women and men. For tests of differences, chi-square tests were used for binary variables and *t*-tests were used for continuous variables. For effect sizes, Cox's index was used for binary variables and Hedge's *g* for continuous variables. Baseline equivalence tests were computed with non-imputed data. Effect sizes of baseline differences > .05 are marked with ^.

Table IV.2d Summary statistics of key baseline measures and baseline equivalence across study conditions, for individuals in the sample for RQ4

Baseline measure	MONTHLY mean or % (women | men)	MONTHLY standard deviation (women | men)	WEEKLY mean or % (women | men)	WEEKLY standard deviation (women | men)	Monthly & Weekly difference in means or % (women | men)	*p*-value of difference in means (women | men)	Effect size abs. value (women | men)
Race							
White/European American	43.8% | 43.7%	n.a.	47.1% | 43.5%	n.a.	3.3% | 0.2%	.31 | .95	.07 | .02
Other	56.2% | 56.3%		52.9% | 56.5%				
Age	36.61 | 38.75	10.00 | 11.07	35.9 | 38.30	10.06 | 10.87	.71 | .46	.28 | .54	.07^ | .04
Education							
No 4-year college degree	48% | 62%	n.a.	48.2% | 60.5%	n.a.	0.2% | 1.5%	.96 | .66	.00 | .03
4-year college degree or higher	52% | 38%		51.8% | 39.5%				
Married Status							
Married	64.5% | 66.7%	n.a.	67.2% | 68.2%	n.a.	2.7% | 1.5%	.97 | .63	.08^ | .03
Non-married	35.5% | 33.3%		32.8% | 31.8%				
Parent Status							
Parent	76.2% | 76.7%	n.a.	76.4% | 76.3%	n.a.	0.2% | 0.4%	.96 | .89	.00 | .03
Non-parent	23.8% | 23.3%						

Baseline measure	MONTHLY mean or % (women | men)	MONTHLY standard deviation (women | men)	WEEKLY mean or % (women | men)	WEEKLY standard deviation (women | men)	Monthly & Weekly difference in means or % (women | men)	*p*-value of difference in means (women | men)	Effect size abs. value (women | men)
			23.6% | 23.7%				
CRSI Relationship Skills *(mean composite; range 1 to 7)*	5.02 | 5.15	.75 | .71	5.03 | 5.10	.68 | .67	.01 | .05	.86 | .32	.01 | .07^
Sample size	473 | 447	n.a.	461 | 446	n.a.	n.a.	n.a.	n.a.

Notes: n.a. = not applicable. See Table IV.3 for description of demographic covariates. CRSI=Couple Relationship Skills Inventory. Analyses were run separately for women and men. For tests of differences, chi-square tests were used for binary variables and *t*-tests were used for continuous variables. For effect sizes, Cox's index was used for binary variables and Hedge's *g* for continuous variables. Baseline equivalence tests were computed with non-imputed data. Effect sizes of baseline differences > .05 are marked with ^.

Table IV.2e Summary statistics of key baseline measures and baseline equivalence across study conditions, for individuals in the sample for RQ5

Baseline measure	MONTHLY mean or % (women | men)	MONTHLY standard deviation (women | men)	WEEKLY mean or % (women | men)	WEEKLY standard deviation (women | men)	Monthly & Weekly difference in means or % (women | men)	*p*-value of difference in means (women | men)	Effect size abs. value (women | men)
Race							
White/European American	44% | 44%	n.a.	46.5% | 43.1%	n.a.	2.5% | 0.9%	.46 | .78	.05 | .02
Other	56% | 56%		53.5% | 56.9%				
Age	36.56 | 38.72	9.87 | 10.95	35.80 | 38.2	10.04 | 10.87	.76 | .52	.26 | .48	.08^ | .05
Education							
No 4-year college degree	48.1% | 62.2%	n.a.	48.8% | 60.4%	n.a.	1.8% | 1.5%	.82 | .60	.02 | .05
4-year college degree or higher	51.9% | 37.8%		51.2% | 39.6%				
Married Status							
Married	64.9% | 67.1%	n.a.	66.4% | 67.6%	n.a.	1.5% | 0.5%	.63 | .90	.03 | .03
Non-married	35.1% | 32.9%		33.6% | 32.4%				
Parent Status							
Parent		n.a.		n.a.	0.1% | 0.7%	.96 | .81	.00 | .00

Baseline measure	MONTHLY mean or % (women | men)	MONTHLY standard deviation (women | men)	WEEKLY mean or % (women | men)	WEEKLY standard deviation (women | men)	Monthly & Weekly difference in means or % (women | men)	*p*-value of difference in means (women | men)	Effect size abs. value (women | men)
Non-parent	76.5% | 77.2% 23.5% | 22.8%		76.4% | 76.5% 23.6% | 23.5%				
CSI Couple Satisfaction *(sum composite; range 4 to 25)*	18.18 | 19.13	4.83 | 4.50	18.38 | 19.16	4.89 | 4.14	.21 |.03	.51 | .91	.04 | .01
Sample size	470 | 444	n.a.	467 | 450	n.a.	n.a.	n.a.	n.a.

Notes: n.a. = not applicable. See Table IV.3 for description of demographic covariates. CSI=Couple Satisfaction. Analyses were run separately for women and men. For tests of differences, chi-square tests were used for binary variables and *t*-tests were used for continuous variables. For effect sizes, Cox's index was used for binary variables and Hedge's *g* for continuous variables. Baseline equivalence tests were computed with non-imputed data. Effect sizes of baseline differences > .05 are marked with ^.

Table IV.2f Summary statistics of key baseline measures and baseline equivalence across study conditions, for individuals in the sample for RQ6

Baseline measure	MONTHLY mean or % (women | men)	MONTHLY standard deviation (women | men)	WEEKLY mean or % (women | men)	WEEKLY standard deviation (women | men)	Monthly & Weekly difference in means or % (women | men)	*p*-value of difference in means (women | men)	Effect size abs. value (women | men)
Race							
White/European American	42.8% | 43.1%	n.a.	45.7% | 42.1%	n.a.	2.9% | 1%	.36 | .75	.07^ | .02
Other	57.2% | 56.9%		54.3% | 57.9%				
Age	36.41 | 38.59	9.96 | 11.04	35.93 | 38.46	10.25 | 11.18	.48 | .13	.45 | .86	.05 | .01
Education							
No 4-year college degree	48.4% | 62.5%	n.a.	51.1% | 62.7%	n.a.	2.7% | 0.2%	.39 | .95	.07^ | .03
4-year college degree or higher	51.6% | 37.5%		48.9% | 37.3%				
Married Status							
Married	61.6% | 64.0%	n.a.	62.7% | 64.2%	n.a.	1.1% | 0.2%	.71 | .95	.03 | .00
Non-married	38.4% | 36.0%		37.3% | 35.8%				
Parent Status							
Parent	76.8% | 76.4%	n.a.	75.8% | 76.5%	n.a.	1% | 0.1%	.74 | .99	.03 | .00

Baseline measure	MONTHLY mean or % (women | men)	MONTHLY standard deviation (women | men)	WEEKLY mean or % (women | men)	WEEKLY standard deviation (women | men)	Monthly & Weekly difference in means or % (women | men)	*p*-value of difference in means (women | men)	Effect size abs. value (women | men)
Non-parent	23.2% | 23.6%		24.2% | 23.5%				
Kessler Mental Health symptoms *(mean composite; range 1 to 5)*	2.19 | 2.13	.76 |.77	2.25 | 2.08	.86 | .79	.06 | .05	.25 | .32	.07^ | .07^
Sample size	502 | 472	n.a.	501 | 480	n.a.	n.a.	n.a.	

Notes: n.a. = not applicable. See Table IV.3 for description of demographic covariates. Analyses were run separately for women and men. For tests of differences, chi-square tests were used for binary variables and *t*-tests were used for continuous variables. For effect sizes, Cox's index was used for binary variables and Hedge's *g* for continuous variables. Baseline equivalence tests were computed with non-imputed data. Effect sizes of baseline differences > .05 are marked with ^.

Sample Characteristics. The full analytic sample for this report (in which both partners completed T1 surveys and at least one partner completed a T4 survey; *n* = 1978 individuals / 989 couples) had varied relational and racial backgrounds and the majority can be considered low-resourced. Specific sample characteristics included:

- **Sex**: The sample consisted of nearly equal proportions of men and women (51% women, 49% men).
- **Race:** A large portion (43%) identified as white/European American; whereas, the majority of the sample (57%) identified in another race (i.e., African American, Asian or Asian American, Multiracial or Biracial, Native American or Alaskan Native, Native Hawaiian or Pacific-Islander).
- **Education**: Over half of the sample (56%) did not have a 4-year college degree, while the remaining 44% held a 4-year college degree or higher.
- **Total household income before taxes:** Compared with the US Census Bureau (2022) average household income in Alabama of $82,992, the majority of this sample reported a lower than average income (i.e., 16% earning less than $20K, 18% between $20K-$40K, 28% between $40K-$75K, 17% between $75K-$100K, and 21% greater than $100K).
- **Married status:** Sixty-three percent of the sample were married, with the remaining 37% unmarried.
- **Parent status:** About three quarters of the sample (76%) had children.

C. Estimation approach for primary analyses

This section details the impact estimation methods used to analyze data for primary RQs. Preliminary analyses were run in SPSS statistical software, and primary models were run in MPlus statistical software.

Preliminary Analyses. We selected covariates based on their potential influence on the outcomes (see Table IV.3 for list and description of all covariates used in impact analyses). All covariates in the couple domain were measured at the individual-level but were thoroughly reviewed during the data preparation phase, assessing for match between partners (see Appendix D for data preparation methods).

Before using the modeling approach described below, model assumptions were checked (e.g., normal distribution of primary measures); no variables needed transformation. Informed by baseline equivalence testing, the few demographics which had baseline differences between conditions larger than .05 (but smaller than .25, WWC Standards, 2022) were included as covariates in models. Pearson correlation coefficients were checked for all primary outcomes across time points, and each outcome measure had correlations across time points of greater than .50 (see Appendix D, Table D.2). Pearson correlation coefficients were also checked for all potential covariates (Table IV.3) and primary outcomes (see Appendix D, Tables D.3a & D.3b), and any additional covariates significantly correlated with relevant outcomes were included in primary RQ models below (if not already included based on baseline equivalence testing).

Table IV.3. Covariates included in the impact analyses

Covariate	Domain	Description and coding of covariates
Class Setting	Couple	Dichotomous variable with response options: (0) virtual or (1) in-person
Sex	Individual	Dichotomous variable with response options: (0) man or (1) woman
Age	Individual	Continuous variable: age in years
Race	Individual	Dichotomous variable with response options: (0) White or European American or (1) Other (i.e., African American, Asian or Asian American, Multiracial or Biracial, Native American or Alaskan Native, Native Hawaiian or Pacific-Islander)
Education	Individual	Dichotomous variable with response options: (0) less than 4-year degree or (1) 4-year degree or higher
Parent status	Individual	Dichotomous variable with response options: (0) non-parent or (1) parent (e.g., biological, step, adopted, etc.)
Married status	Couple	Dichotomous variable with response options: (0) non-married or (1) married

Notes: All measures were collected at the individual level but measures can be categorized as reflecting different domains or levels of functioning from each individual's perspective. Data cleaning involved checking for match between partners on covariates in the couple domain.

Modeling approach. For Primary RQs, we used hierarchical linear modeling to assess the impact of the monthly program condition compared with the weekly program condition at the relevant follow-up time point: immediate post-program for RQs 1-3 (i.e., T2 for weekly condition; T3 for monthly condition); 1-year post-baseline for RQs 4-6 (i.e., T4 for both conditions). These multi-level models, which accounted for nesting of individuals within couples, predicted the relevant outcome by randomly assigned condition and baseline levels of the outcome. Specifically, we first assessed within-group changes by estimating the change in the total scores of outcomes from baseline to follow-up within each condition. To do so, we initially included all relevant covariates (e.g., sex, race, education, parental status) in the model. Covariates that were not statistically significant were then removed to obtain a parsimonious model to ensure interpretability. Next, we tested for between-group differences to determine whether the changes differed significantly between the monthly and weekly conditions. The models accounted for baseline outcome levels and included program condition as a predictor. Findings were considered statistically significant based on $p < .05$, two-tailed test. Finally, we calculated Hedge's *g* effect sizes to measure the magnitude of the difference in effects on outcomes between the two conditions.

Model specification. Dependent variables for the Primary RQs included couple relationship skills (RQs 1, 4), couple satisfaction (RQs 2, 5), and mental health symptoms (RQs 3, 6). Independent variables for the Primary RQs included random assignment (monthly or weekly) and the baseline level of the relevant outcome. We included covariates in each model which showed baseline inequivalence between conditions and/or were significantly correlated with the relevant outcome. Initial model results were examined and re-run with nonsignificant covariates removed for enhanced parsimony. See Appendix E for model equations for estimating impacts.

D. Implementation analyses

See Appendix B, Table B.2 for details about the measures, items, and composite construction used for all implementation RQs. All implementation RQs were addressed using either frequencies/percentages or mean composite scores.

E. Sensitivity analyses

Sensitivity analyses were performed to test the robustness of the results with non-imputed data. Since this study had low couple-level attrition, for primary analyses we included all couples in which both partners completed T1 data and at least one partner completed relevant follow-up survey data (see above details on missing data in Section IV.A). Missing data was imputed for the cases of partners missing follow-up data, as well as cases missing <=20% of items on relevant outcomes. Thus, we conducted sensitivity analyses for each RQ with analytic samples of individuals who had complete data at all relevant timepoints.

V. Findings

This section describes the key findings of the impact and implementation studies, organized by outcome.

A. Results of the primary impact evaluation

Key findings

- **RQ 1 - Couple Relationship Skills at Immediate Post-Program:** There was no significant difference in the changes between monthly and weekly program conditions. Both monthly and weekly program participants demonstrated significant increases in couple relationship skills from baseline to immediate post-program, with monthly participants on average increasing 0.33 points (on a 7-point scale) and weekly participants on average increasing 0.32 points (on a 7-point scale).

- **RQ 2 - Couple Satisfaction at Immediate Post-Program:** There was no significant difference in the changes between monthly and weekly program conditions. Both monthly and weekly participants significantly increased in couple satisfaction from baseline to immediate post-program, with monthly participants on average increasing 1.20 points (on a 21-point scale) and weekly participants on average increasing 0.97 points (on a 21-point scale).
- **RQ 3 - Mental Health Symptoms at Immediate Post-Program:** There was no significant difference in the changes between monthly and weekly program conditions. Both monthly and weekly participants significantly decreased in mental health symptoms from baseline to immediate post-program, with monthly participants on average decreasing 0.25 points (on a 5-point scale) and weekly participants on average decreasing 0.22 points (on a 5-point scale).
- **RQ 4 - Couple Relationship Skills at 1-Year Follow-Up:** There was no significant difference in the changes between monthly and weekly program conditions. Both monthly and weekly participants significantly increased in couple relationship skills from baseline to 1-year follow-up, with monthly participants on average increasing 0.17 points (on a 7-point scale) and weekly participants on average increasing 0.22 points (on a 7-point scale).
- **RQ 5 - Couple Satisfaction at 1-Year Follow-Up:** There was no significant difference in the changes between monthly and weekly program conditions. Both monthly and weekly participants significantly increased in couple satisfaction from baseline to 1-year follow-up, with monthly participants on average increasing 0.26 points (on a 21-point scale) and weekly participants on average increasing 0.45 points (on a 21-point scale).
- **RQ 6 - Mental Health Symptoms at 1-Year Follow-Up:** There was no significant difference in the changes between monthly and weekly program conditions. Both monthly and weekly participants significantly decreased in mental health symptoms from baseline to 1-year follow-up, with monthly participants on average decreasing 0.17 points (on a 5-point scale) and weekly participants on average decreasing 0.20 points (on a 5-point scale).

Couple Relationship Skills (RQs 1 & 4): A multilevel model was conducted to examine changes in couple relationship skills from baseline to immediate post-program for participants in the monthly and weekly conditions. The model demonstrated a good fit to the data ($\chi^2(17) = 50.83$, $p < .001$, CFI = 0.960, TLI = 0.925, RMSEA = 0.047). Participants in both monthly and weekly conditions reported significant increases in couple relationship skills from baseline to immediate post-program (Monthly: M[change score]=0.33, SD=0.18, $p<.001$, n=894 individuals; Weekly: M[change score]=0.32, SD=0.21, $p<.001$, n=923 individuals; see Table V.1a). There was no significant difference in the change scores between groups (M[change score difference]=0.01, p=0.87, Hedge's g effect size=0.05; Table V.1a). Further, a multilevel model was also conducted to test changes in couple relationship skills from baseline to one-year follow-up for participants in both conditions. The model also showed a good model fit ($\chi^2(13) = 32.91$, $p = 0.002$, CFI = 0.973, TLI = 0.949, RMSEA = 0.041). Participants in both monthly and weekly conditions significantly increased in couple relationship skills from baseline to 1-year follow-up (Monthly: M[change score]=0.17, SD=0.25, $p<.001$, n=916 individuals; Weekly: M[change score]=0.22, SD=0.26, $p<.001$, n=902 individuals; see Table V.1d). There was no significant difference in the change scores between groups (M[change score difference]=-0.04, p=0.21, Hedge's g effect size=0.20; Table V.1d).

Couple Satisfaction (RQs 2 & 5): A multilevel model was carried out to investigate changes in couple satisfaction from baseline to immediate post-program for participants in both the monthly and weekly conditions. This model indicated an excellent data fit($\chi^2(13) = 86.22$, $p < .001$, CFI = 0.911, TLI = 0.836, RMSEA = 0.079). Participants in both monthly and weekly conditions reported significant increases in couple satisfaction from baseline to immediate post-program (Monthly: M[change score]=1.20, SD=0.69, $p<.001$, n=892 individuals; Weekly: M[change score]=0.97, SD=1.12, $p<.001$, n=925 individuals; see Table V.1b). There was no significant difference in the change scores between groups (M[change score difference]=0.23, p=0.24, Hedge's g effect size=0.25; Table V.1b). Further, a similar multilevel model examined changes in couple satisfaction from baseline to one-year follow-up. The model

demonstrated acceptable fit to the data($\chi^2(15) = 65.23$, $p < .001$, CFI = 0.928, TLI = 0.866, RMSEA = 0.061). Participants in both monthly and weekly conditions significantly increased in couple satisfaction from baseline to 1-year follow-up (Monthly: M[change score]=0.26, SD=1.62, $p<.10$, n=900 individuals; Weekly: M[change score]=0.45, SD=1.21, $p<.05$, n=898 individuals; see Table V.1e). There was no significant difference in the change scores between groups (M[change score difference]=-0.19, p=0.36, Hedge's g effect size=0.13; Table V.1e).

Mental Health Symptoms (RQs 3 & 6): A multilevel model was estimated to assess changes in mental health symptoms from baseline to immediate post-program. The model exhibited a strong fit to the data($\chi^2(17) = 15.15$, $p = 0.585$, with CFI = 1.000, TLI = 1.000, and RMSEA = 0.000). Participants in both monthly and weekly conditions reported significant decreases in mental health symptoms from baseline to immediate post-program (Monthly: M[change score]=-0.25, SD=0.20, $p<.001$, n=919 individuals; Weekly: M[change score]=-0.22, SD=0.18, $p<.001$, n=940 individuals; see Table V.1c). There was no significant difference in the change scores between groups (M[change score difference]=-0.02, p=0.47, Hedge's g effect size=0.16; Table V.1c). Further, to examine the long-term effects, another multilevel model was fitted to the data. The model suggested good fit($\chi^2(17) = 26.94$, $p = 0.059$, CFI = 0.988, TLI = 0.978, RMSEA = 0.025). Participants in both monthly and weekly conditions significantly decreased in mental health symptoms from baseline to 1-year follow-up (Monthly: M[change score]=-0.17, SD=0.22, $p<.001$, n=953 individuals; Weekly: M[change score]=-0.20, SD=0.21, $p<.001$, n=948 individuals; see Table V.1f). There was no significant difference in the change scores between groups (M[change score difference]=0.03, p=0.49, Hedge's g effect size=0.14; Table V.1f).

Table V.1a. Estimated effects of programs on change in couple relationship skills using data from baseline to immediate post-program to address primary RQ 1

Outcome measure	MONTHLY mean change (*p*-value)	MONTHLY standard deviation	WEEKLY mean change (*p*-value)	WEEKLY standard deviation	MONTHLY and WEEKLY mean change difference	*p*-value of test of mean change difference	Effect size of difference between groups
Couple Relationship Skills	0.33 (< .001)	0.18	0.32 (< .001)	0.21	0.01	0.87	0.05
Sample size (individuals)	894	n.a.	923	n.a.	n.a.	n.a.	n.a.

Source: T2 (i.e., 3-month post-baseline) surveys for weekly participants; T3 (i.e., 6-month post-baseline) surveys for monthly participants

Notes: Effect sizes were calculated using Hedges' *g* formula. Couple relationship skills outcome is a mean composite (i.e., 7-point scale; value range 1-7) with higher scores representing better skills. The analyses included sex, age, race, education, parent status, class setting, and married status as covariates to control for their potential influence. Sample sizes (n) are slightly lower than the full analytic sample for this RQ (Table IV.1b), because cases with missing covariate data were dropped in analyses. See Table III.1 for a more detailed description of the outcome measure and report Section IV.C in Chapter IV for a description of the impact estimation approach. n.a. = not applicable.

Table V.1b. Estimated effects of programs on change in couple satisfaction using data from baseline to immediate post-program to address primary RQ 2

Outcome measure	MONTHLY mean change (*p*-value)	MONTHLY standard deviation	WEEKLY mean change (*p*-value)	WEEKLY standard deviation	FMONTHLY and WEEKLY mean change difference	*p*-value of test of mean change difference	Effect size of difference between groups
Couple Satisfaction	1.20 (< .001)	.69	0.97 (< .001)	1.12	0.23	0.24	0.25
Sample size (individuals)	892	n.a.	925	n.a.	n.a.	n.a.	n.a.

Source: T2 (i.e., 3-month post-baseline) surveys for weekly participants; T3 (i.e., 6-month post-baseline) surveys for monthly participants

Notes: Effect sizes were calculated using Hedges' *g* formula. Couple satisfaction outcome is a sum composite (i.e., 21-point scale; value range 4-25) with higher scores representing greater satisfaction. The analyses included sex, age, race, parent status, and class setting as covariates to control for their potential influence. Sample sizes (n) are slightly lower than the full analytic sample for this RQ (Table IV.1b), because cases with missing covariate data were dropped in analyses. See Table III.1 for a more detailed description of each measure and report section IV.C in Chapter IV for a description of the impact estimation approach. n.a. = not applicable.

Table V.1c. Estimated effects of programs on change in mental health symptoms using data from baseline to immediate post-program to address primary RQ 3

Outcome measure	MONTHLY mean change (*p*-value)	MONTHLY standard deviation	WEEKLY mean change (*p*-value)	WEEKLY standard deviation	MONTHLY and WEEKLY mean change difference	*p*-value of test of mean change difference	Effect size of difference between groups
Mental Health Symptoms	-0.25 (< .001)	0.20	-0.22 (< .001)	0.18	-0.03	0.47	0.16
Sample size (individuals)	919	n.a.	940	n.a.	n.a.	n.a.	n.a.

Source: T2 (i.e., 3-month post-baseline) surveys for weekly participants; T3 (i.e., 6-month post-baseline) surveys for monthly participants

Notes: Effect sizes were calculated using Hedges' *g* formula. Mental health symptoms is a mean composite (i.e., 5-point scale; value range 1-5) with higher scores representing more mental health symptoms. The analyses included sex, age, race, education level, class setting, and married status as covariates to control for their potential influence. Sample sizes (n) are slightly lower than the full analytic sample for this RQ (Table IV.1b), because cases with missing covariate data were dropped in analyses. See Table III.1 for a more detailed description of each measure and report section IV.C in Chapter IV for a description of the impact estimation approach. n.a. = not applicable.

Table V.1d. Estimated effects of programs on change in couple relationship skills using data from baseline to 1-year follow-up to address primary RQ 4

Outcome measure	MONTHLY mean change (*p*-value)	MONTHLY standard deviation	WEEKLY mean change (*p*-value)	WEEKLY standard deviation	MONTHLY and WEEKLY mean change difference	*p*-value of test of mean change difference	Effect size of difference between groups
Couple Relationship Skills	0.17 (< .001)	0.25	0.22 (< .001)	0.26	-0.04	0.21	0.20
Sample size (individuals)	916	n.a.	902	n.a.	n.a.	n.a.	n.a.

Source: T4 (i.e., 1-year post-baseline) surveys for both weekly and monthly participants

Notes: Effect sizes were calculated using Hedges' *g* formula. Couple relationship skills outcome is a mean composite (i.e., 7-point scale; value range 1-7) with higher scores representing better skills. The analyses included sex, race, education level, parent status, and class setting as covariates to control for their potential influence. Sample sizes (n) are slightly lower than the full analytic sample for this RQ (Table IV.1b), because cases with missing covariate data were dropped in analyses. See Table III.1 for a more detailed description of each measure and report section IV.C in Chapter IV for a description of the impact estimation approach. n.a. = not applicable.

Table V.1e. Estimated effects of programs on change in couple satisfaction using data from baseline to 1-year follow-up to address primary RQ 5

Outcome measure	MONTHLY mean change (*p*-value)	MONTHLY standard deviation	WEEKLY mean change (*p*-value)	WEEKLY standard deviation	MONTHLY and WEEKLY mean change difference	*p*-value of test of mean change difference	Effect size of difference between groups
Couple Satisfaction	0.26 (.093)	1.62	0.45 (.002)	1.21	-0.19	0.36	0.13
Sample size (individuals)	900	n.a.	898	n.a.	n.a.	n.a.	n.a.

Source: T4 (i.e., 1-year post-baseline) surveys for both weekly and monthly participants

Notes: Effect sizes were calculated using Hedges' *g* formula. Couple satisfaction outcome is a sum composite (i.e., 21-point scale; value range 4-25) with higher scores representing greater satisfaction. The analyses included sex, age, race, parent status, class setting, and married status as covariates to control for their potential influence. Sample sizes (n) are slightly lower than the full analytic sample for this RQ (Table IV.1b), because cases with missing covariate data were dropped in analyses. See Table III.1 for a more detailed description of each measure and report section IV.C in Chapter IV for a description of the impact estimation approach. n.a. = not applicable.

Table V.1f. Estimated effects of programs on change in mental health symptoms using data from baseline to 1-year follow-up to address primary RQ 6

Outcome measure	MONTHLY mean change (*p*-value)	MONTHLY standard deviation	WEEKLY mean change (*p*-value)	WEEKLY standard deviation	MONTHLY and WEEKLY mean change difference	*p*-value of test of mean change difference	Effect size of difference between groups
Mental Health Symptoms	-0.17 (< .001)	0.22	-0.20 (< .001)	0.21	0.03	0.49	0.14
Sample size (individuals)	953	n.a.	948	n.a.	n.a.	n.a.	n.a.

Source: T4 (i.e., 1-year post-baseline) surveys for both weekly and monthly participants

Notes: Effect sizes were calculated using Hedges' *g* formula. Mental health symptoms is a mean composite (i.e., 5-point scale; value range 1-5) with higher scores representing more mental health symptoms. The analyses included sex, age, race, education level, class setting, and married status as covariates to control for their potential influence. Sample sizes (n) are slightly lower than the full analytic sample for this RQ (Table IV.1b), because cases with missing covariate data were dropped in analyses. See Table III.1 for a more detailed description of each measure and report section IV.C in Chapter IV for a description of the impact estimation approach. n.a. = not applicable.

B. **Results of the implementation evaluation**

Key findings

- **Fidelity:** According to facilitator reports, curriculum fidelity was very high in both groups. On average, 94% of all curriculum activities were delivered in both the weekly and the monthly program class series.
- **Dosage:** For both the monthly and weekly program conditions, attendance and completion rates were very high. AHMRE achieved its benchmark of at least 80% of all participants completing all 6 sessions (i.e., 80% of monthly participants / 90% of weekly participants completed all sessions).
- **Quality:** The average participants' rating of facilitation quality of their AHMRE class series was very high in the weekly and the monthly program conditions.
- **Engagement:** The average participant rating of their own engagement in the 6-session class series was high in the weekly and monthly program conditions.
- **Context:** Overall, at one-year post-baseline, about 1 out of 5 participants (i.e., 20% for monthly; 18% for weekly) reported having begun therapy/counseling in the prior year during which they participated in the AHMRE class series.

Fidelity. The AHMRE ELEVATE curriculum was implemented with exceptionally high fidelity in both monthly and weekly program conditions; in total there were 56 monthly and 56 weekly unique 6-session class series throughout the impact study). During each class series, a total of 79 curriculum activities were included (e.g., lessons, partner-engagement activities, group discussions), and facilitators reported teaching, on average, 94% of all total activities across the 6-session class series in both monthly and weekly conditions.

Dosage. For both monthly (n=559 couples) and weekly (n=561 couples) conditions, the overwhelming majority of couples attended at least one class via live or makeup session, with similar rates across conditions (i.e., 98% for monthly; 97% for weekly). The attendance gap slightly widened over the 6-session series, such that 87% of monthly couples attended half or more of the class sessions versus 93% of weekly couples attended half or more. Overall, by the end of the six 2-hour session workshop series, both groups had completion rates at or above AHMRE's 80% benchmark; however, 8 in 10 monthly couples (80%) had completed all 6 sessions versus 9 in 10 weekly couples (90%).

Quality: At immediate post-program, participants in both monthly and weekly conditions rated their facilitation quality (e.g., teaching effectively, sharing examples, encouraging class participation, managing time) very high. Both groups reported average facilitation quality scores of 4.5 on a scale from 1 to 5, with higher scores representing optimal facilitation; Weekly M=4.5, SD=.80 *n*=1018 individuals; Monthly M=4.5, SD=.82, *n*=924 individuals).

Engagement: At immediate post-program, participants in both monthly and weekly conditions rated their own engagement in the class series high (e.g., feeling connected to other participants, engaging in group discussions/activities, feeling comfortable sharing experiences). Both groups reported average engagement scores of 4.0 on a scale from 1 to 5, with higher scores representing greater engagement (Weekly M=4.0, SD=.80, *n*=1013 individuals; Monthly M=4.0, SD=.82; *n*=920 individuals).

Context: On the final follow-up survey at 1-year post-baseline, about 1 out of 5 participants reported having begun therapy/counseling in the prior year (during which they participated in the AHMRE class series). Specifically, in the weekly group, 18% (*n*=891 individuals) and in the monthly group, 20% (*n*=877 individuals) reported having begun therapy/counseling in the prior year.

C. **Results of the sensitivity analyses**

Key findings

- **Complete Case Sensitivity Analyses:** Results of sensitivity analyses conducted with complete-case (non-imputed) data were identical to results of primary analyses comparing the changes in outcomes between monthly and weekly program conditions. Specifically, there were no significant differences between conditions in changes in couple relationship skills, couple satisfaction, or mental health symptoms from baseline to immediate post-program or from baseline to 1-year post-baseline.

Complete Case Sensitivity Analyses. Sensitivity analyses were conducted with complete case (non-imputed) data (see Table V.2). For RQ 1, similar to primary analyses, there was no significant difference in the change between groups (M[change score difference]=0.01, p=0.88). For RQ 2, similar to primary analyses, there was no significant difference in the change between groups (M[change score difference]=0.01, p=0.49). For RQ 3, similar to primary analyses, there was no significant difference in the change between groups (M[change score difference]=.00, p=0.99). For RQ 4, similar to primary analyses, there was no significant difference in the change between groups (M[change score difference]=-.02, p=0.49). For RQ 5, similar to primary analyses, there was no significant difference in the change between groups (M[change score difference]=-.12, p=0.58). For RQ 6, similar to primary analyses, there was no significant difference in the change between groups (M[change score difference]=0.04, p=0.33).

Table V.2. Differences in mean changes between monthly and weekly program conditions estimated using alternative methods (sensitivity analyses)

Outcome	Primary approach Monthly & Weekly Mean Change Difference (p-value)	Sensitivity Approach: Complete Case Data Monthly & Weekly Mean Change Difference (p-value)
Primary RQs		
RQ 1: Couple relationship skills (baseline to immediate post-program)	0.01 (p = 0.87)	0.01 (p = 0.88)
RQ 2: Couple satisfaction (baseline to immediate post-program)	0.23 (p = 0.24)	0.01 (p = 0.49)
RQ 3: Mental health symptoms (baseline to immediate post-program)	-0.03 (p = 0.47)	-0.00 (p = 0.99)
RQ 4: Couple relationship skills (baseline to 1-year post-baseline)	-0.04 (p = 0.21)	-0.02 (p = 0.49)
RQ 5: Couple satisfaction (baseline to 1-year post-baseline)	-0.19 (p = 0.36)	-0.12 (p = 0.58)
RQ 6: Mental health symptoms (baseline to 1-year post-baseline)	0.03 (p = 0.49)	0.04 (p = 0.33)

Source: For RQs 1-3, outcome was obtained on T2 (i.e., 3-month post-baseline) surveys for weekly participants and T3 (i.e., 6-month post-baseline) surveys for monthly participants. For RQs 4-6, outcome was obtained on T4 (i.e., 1-year post-baseline) surveys for both weekly and monthly participants.

Notes: Sensitivity analyses were conducted with complete case (non-imputed) data. The analyses for RQ 1 included the following covariates to control for their potential influence: sex, age, race, education, parent status, class setting, and married status. For RQ 2, we included sex, age, race, parent status, and class setting. For RQ 3 we included sex, age, race, education level, class setting, and married status. For RQ 4 we included sex, race, education level, parent status, and class setting. For RQ 5 we included sex, age, race, parent status, class setting, and married status. For RQ 6 we included sex, age, race, education level, class setting, and married status. See Table III.1 for a more detailed description of each measure and report section IV.C in Chapter IV for a description of the impact estimation approach.

VI. Summary and conclusions

This section describes the interpretation of findings, implications, considerations, and limitations of the impact and implementation evaluation.

A. Implications

The findings of this study underscore the effectiveness of both weekly and monthly ELEVATE relationship education workshops in community-based settings with a wide range of couples. A previous efficacy trial using a no-program control group established the impact of the weekly delivery model for ELEVATE (Adler-Baeder, et al., 2022). The current study extends these findings to the monthly delivery model. Notably, there was no difference in the comparative effects of these two different delivery formats, indicating that monthly ELEVATE delivery is a viable option for HMRE community programming. While we expected enhanced effects of monthly delivery due to the longer duration of program engagement, we found similar (rather than stronger) positive effects for both program conditions in relational and individual domains at both immediate post-program and one-year follow-up. This study expands the HMRE implementation possibilities for couples and program staff, potentially meeting the needs of various participants by offering greater scheduling flexibility without compromising program impact. These findings are consistent with existing literature emphasizing the potential of relationship education programs to enhance individual and couple functioning (e.g., Adler-Baeder et al., 2022; Hawkins et al., 2022a; McGill et al., 2021b). To our knowledge this is the first test of an HMRE program delivered in six monthly sessions.

The absence of significant differences in the change in all measured outcomes (i.e., couple relationship skills, couple satisfaction, and mental health symptoms) between monthly and weekly groups highlights the versatility of the ELEVATE curriculum in yielding meaningful results regardless of delivery modality. With the impact of the ELEVATE weekly program, compared with a no-program control group, already established in a prior RCT study (Adler-Baeder et al., 2022), the similarity of results for both weekly and monthly ELEVATE conditions in this randomized study is even more profound. Further reinforcing the robustness of these findings, this study achieved low attrition at both individual- and couple-levels for all analytic samples (WWC Standards, 2022).

This parity between delivery modalities aligns with prior research suggesting that program content, facilitation quality, and participant engagement may hold greater influence over outcomes than delivery format (Markman & Rhoades, 2012; Rauer et al., 2014). We can speculate that the high implementation fidelity, participant engagement, and facilitation quality observed in both conditions likely contributed to the equivalent effects; however, this remains a testable question. We do note the variation between groups in participant retention and dosage (i.e., 90% of weekly participants, 80% of monthly participants completed all 6 sessions). This may suggest potential logistical advantages of shorter-term, more frequent sessions. Our facilitators, in fact, reflected on comparatively more effort needed to retain the monthly participants. Notably, however, due to the random assignment design of the study, participants did not select their program condition. It is likely that retention efforts could be similar across weekly and monthly offerings if all participants self-select into the type of delivery model that best meets their needs, expectations, and schedules.

The study's results also reinforce the theoretical framework of the transactional ecological family systems approach (Bronfenbrenner, 1977), demonstrating the interconnectedness of relational and individual wellbeing. Improved relationship skills and satisfaction were mirrored by reductions in mental health symptomology, in line with (although not directly evaluated in this study) the notion of spillover effects reported in prior evaluations of HMRE interventions (Adler-Baeder et al., 2025; Adler-Baeder et al., 2018; Kiecolt-Glaser & Wilson, 2017). These findings align with prevention science principles, highlighting the potential of coordinated, multilevel interventions to promote resilience and positive outcomes among varying populations (Coie et al., 1993; Landolt et al., 2023). Moreover, the study's novel

focus on extended delivery and sustained participant-facilitator connections provides valuable insights for policymakers and practitioners. While long-term engagement with facilitators did not yield superior outcomes in this study, the comparable efficacy of the monthly program delivery model suggests that it remains a viable option for participants facing scheduling constraints or those requiring more flexible program structures. It is also possible that the benefits of the longer duration of the monthly program yields greater benefits in maintenance of positive effects beyond one year.

B. **Limitations and future directions**

While the findings of this study provide valuable insights into the equivalence of change in the two HMRE delivery formats, several limitations warrant consideration. First, the sample population primarily represents a wide range of families in different economic situations in Alabama, limiting the generalizability of results to broader or different populations. While this focus aligns with the study's goals of addressing Alabama's unique relational and economic challenges, further research is needed to evaluate the program's effectiveness in different contexts. Studies examining the interplay of demographic variables, such as parent or employment status, and community settings or among various couple types, such as married or dating relationships, could yield a more comprehensive understanding of HMRE impacts in the two program delivery conditions (Hawkins et al., 2022b; McGill et al., 2021b). Additionally, the study's one-year follow-up interval provides valuable insights into the short-term sustainability of program effects; however, longer-term follow-ups are necessary to determine whether these outcomes persist or evolve over time and whether differences in program effects between the two groups emerge over time. Extended longitudinal studies could also examine potential downstream benefits for children, as previous research has highlighted the intergenerational effects of improved parental relationship functioning (Adler-Baeder et al., 2025; Adler-Baeder et al., 2018). In summary, future research should prioritize incorporating extended longitudinal data and family members' reports and exploring these and other novel delivery mechanisms to optimize HMRE interventions.

C. **Other lessons learned**

This study offers several critical lessons for the design, implementation, and evaluation of HMRE programs. The findings emphasize the flexibility of the ELEVATE curriculum in producing significant relational and individual benefits, regardless of whether it is delivered in a weekly or monthly format. The study also highlights the importance of maintaining high implementation fidelity to ensure consistent outcomes across varying program models. Both conditions in this study achieved exemplary fidelity, underscoring the value of thorough facilitator training and robust program monitoring protocols. This aligns with prior research advocating for the standardization of curriculum delivery to enhance program reliability (Adler-Baeder et al., 2022). Another notable lesson is the logistical advantage of weekly sessions in achieving slightly higher dosage (i.e., completion rates). While both formats demonstrated high overall participant retention, the weekly delivery model appeared to better accommodate participants' ability to complete all program hours, which could inform future program scheduling, particularly when targeting populations with challenges to long-term participation. However, having open choice of format would also inform program staff about who would choose which method of program delivery. Since this is the first study of a monthly delivery model of HMRE and a first experience for our program staff, there remains a need to explore strategies for improving participant retention in sustained delivery models. Leveraging technology, such as virtual sessions or app-based boosters, may offer innovative solutions to further enhance convenience and adherence when schedules shift in unanticipated ways over the six months. Lastly, this study underscores the need for ongoing innovation in program delivery, particularly for participants with demanding schedules or limited mobility. By integrating these lessons, HMRE programs can refine their approaches to provide more choice and broader services to communities and further enhance outreach and results in positive relational and individual outcomes.

VII. References

Adler-Baeder, F., Bradford, A., Skuban, E., Lucier-Greer, M., Ketring, S., & 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., Futris, T., Higginbotham, B., Cooper, E., Gregson, K., Bailey, Z, & Turner, J. (2023, Nov 8-11). Developing the Class Environment Scale for Couples in Community Education [Paper presentation]. National Council on Family Relations annual conference, Orlando, FL.

Adler-Baeder, F., Garneau, C., Vaughn, B., McGill, J., Harcourt, K. T., Ketring, S., & 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., Gregson, K., Cooper, E., Burke, L., ... & Jackel, R. (2022). Concurrent randomized control trials of the 1-year efficacy of two couple relationship education programs: ELEVATE and Couples Connecting Mindfully. *Family Process*, *61*(3), 986-1004. https://doi.org/10.1111/famp.12750

Adler-Baeder, F., McGill, J., Frosch, C. A., Wei, M., & Coskun, M. (2025). The impact of Couple Relationship Education (CRE) on parents of young children's couple functioning, family harmony, parenting and their children's aggressive behaviour. *Early Child Development and Care*. Advance online publication. https://doi.org/10.1080/03004430.2025.2452578

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

Centers for Disease Control and Prevention. (2022). *Divorce rates by state: 2019-2022*. U.S. Department of Health & Human Services. https://www.cdc.gov/nchs/pressroom/sosmap/divorce_states/divorce_rates.htm

Coie, J. D., Watt, N. F., West, S. G., Hawkins, J. D., Asarnow, J. R., Markman, H. J., ... & Long, B. (1993). The science of prevention: a conceptual framework and some directions for a national research program. *American Psychologist*, *48*(10), 1013-1022. https://doi.org/10.1037/0003-066X.48.10.1013

Flynt, J. W. (2023). *Poverty in Alabama.* Encyclopedia of Alabama. https://encyclopediaofalabama.org/article/poverty-in-alabama/

Futris, T. G., Adler-Baeder, F., Ketring, S., Smith, T., Bradford, A., Cook, L., ..., & 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

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

Hawkins, A. J., Hill, M. S., Eliason, S. A., Simpson, D. M., & Hokanson, S. (2022a). Do couple relationship education programs affect coparenting, parenting, and child outcomes? A meta-analytic study. *Journal of Child and Family Studies*, *31*(2), 588-598. https://doi.org/10.1007/s10826-022-02229-w

Hawkins, A. J., Hokanson, S., Loveridge, E., Milius, E., Duncan, M., Booth, M., & Pollard, B. (2022b). How effective are ACF-funded couple relationship education programs? A meta-analytic study. *Family Process*, *61*(3), 970-985. https://doi.org/10.1111/famp.12739

Hawkins, A. J., & 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 & Family Review, 48*, 524-554. https://doi.org/10.1080/01494929.2012.677751

Kids Count. (2024). *2024 National KIDS COUNT Data Book*. VOICES for Alabama's Children. https://alavoices.org/national-kids-count/

Kiecolt-Glaser, J. K., & Wilson, S. J. (2017). Lovesick: How couples' relationships influence health. *Annual Review of Clinical Psychology*, *13*(1), 421-443. https://doi.org/10.1146/annurev-clinpsy-032816-045111

Landolt, S. A., Weitkamp, K., Roth, M., Sisson, N. M., & Bodenmann, G. (2023). Dyadic coping and mental health in couples: A systematic review. *Clinical Psychology Review*, *106*, Article 102344. https://doi.org/10.1016/j.cpr.2023.102344

Markman, H. J., & 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

McGill, J., & Adler-Baeder, F. (2020). Exploring the link between mindfulness and relationship quality: Direct and indirect pathways. *Journal of Marital and Family Therapy*, *46*(3), 523-540. https://doi.org/10.1111/jmft.12412

McGill, J., Adler-Baeder, F., & Garneau-Rosner, C. (2021a). An evaluation of the ELEVATE program for couples: Considering vulnerabilities and relationship length. *Family Relations*, *70*(1), 327-351. https://doi.org/10.1111/fare.12502

McGill, J., Adler-Baeder, F., & Gregson, K. (2021b). *Impact evaluation of ELEVATE and Couples Connecting Mindfully in Alabama: Final impact evaluation report for Auburn University.* Office of Planning, Research, and Evaluation; Administration for Children and Families. Available from https://acf.gov/sites/default/files/documents/ofa/Auburn_U_Impact_Report.pdf

Rauer, A. J., Adler-Baeder, F., Lucier-Greer, M., Skuban, E., Ketring, S. A., & 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

Wei, M., Adler-Baeder, F., McGill, J. M., & Novak, J. R. (2024). Healthy couple, better sleep: exploring connections and changes in couple relationship education participants. *Family Process*, *63*(4), 1753-1774. https://doi.org/10.1111/famp.12987

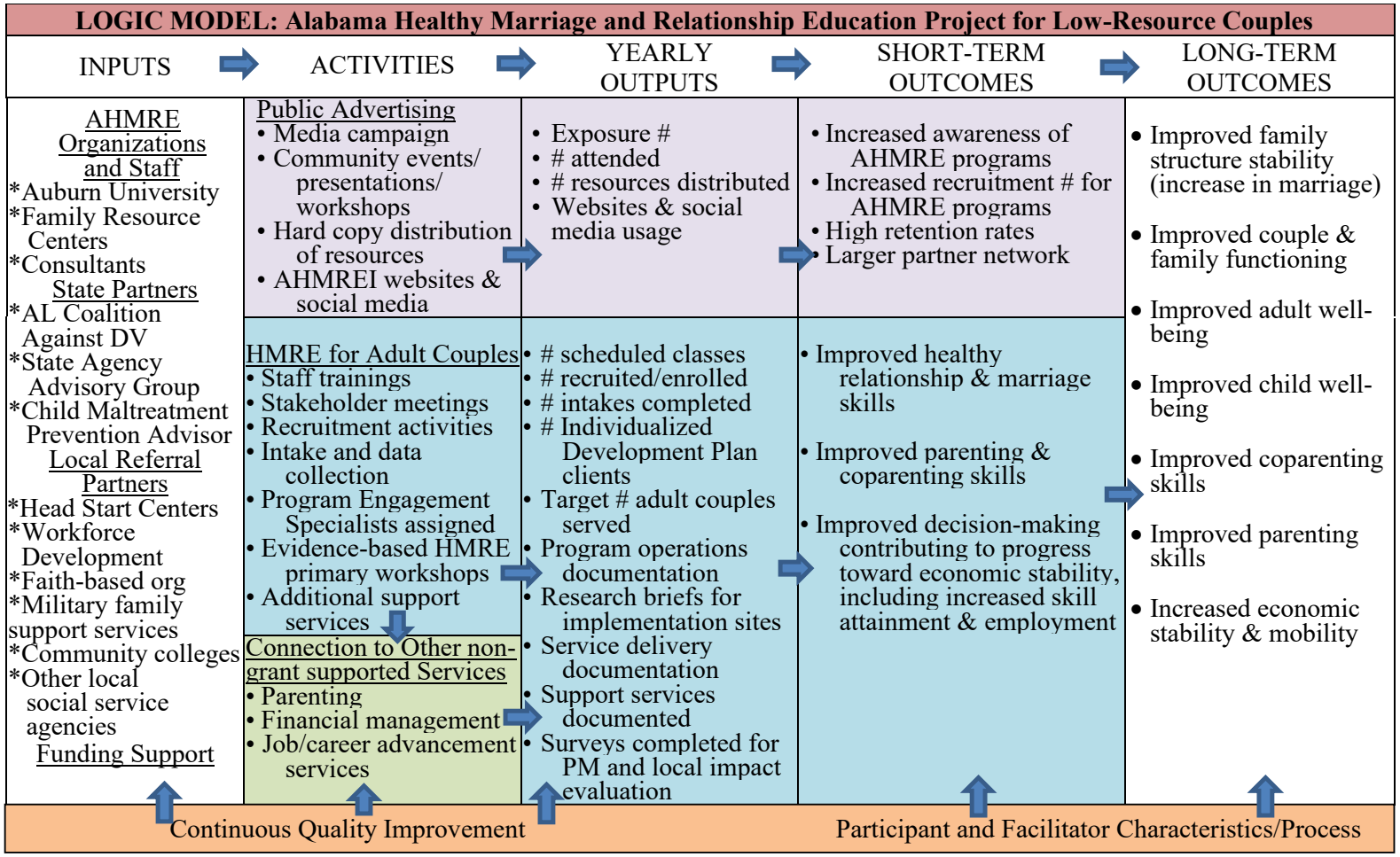
Weiss, H. B., & Jacobs, F. H. (Eds.). (2008). *Evaluating Family Programs: Current Issues in Theory and Policy*. Transaction Publishers.

What Works Clearinghouse (WWC) Standards (2022). *What Works Clearinghouse procedures and standards handbook* (Version 5.0). U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance. Available from https://ies.ed.gov/ncee/wwc/Handbooks

Wood, R. G., Moore, Q., Clarkwest, A., & 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

Appendix A: Logic Model



Problem Statement: Alabama has a history of high levels of marital, family, and economic instability, resulting in increased risks for citizens. There is a great need for widespread access to programs that support stable, healthy relationships and marriages and economic self-sufficiency, particularly in the targeted low-resource communities where supports are limited.

Underlying Assumptions: In the face of risks, the implementation of the widespread, coordinated multi-site **AHMRE project** will result in measurable positive short-term and long-term outcomes for couples and their families and result in reaching the **overarching goal** for ***strengthening families and communities in Alabama.***

Appendix B: Data and Study Sample

Table B.1. Key features of data collection for the impact analysis

Study group	Data source	Timing of data collection	Mode of data collection	Party responsible for data collection	Start and end date of data collection
Monthly Program Condition	Local evaluation surveys	T1: Baseline (i.e., pre-program) T2: 3-months post-baseline T3: 6-months post-baseline (i.e., immediate post-program) T4: 1-year post-baseline	Online survey (completed on participants' own devices)	Evaluation staff	October 2021 through November 2024
Weekly Program Condition	Local evaluation surveys	T1: Baseline (i.e., pre-program) T2: 3-months post-baseline (i.e., immediate post-program) T3: 6-months post-baseline T4: 1-year post-baseline	Online survey (completed on participants' own devices)	Evaluation staff	October 2021 through November 2024

Table B.2. Key features of data collection for the implementation analysis

Focus	Research question	Data source	Timing and frequency of data collection	Party responsible for data collection
Fidelity	1. For each condition, according to facilitator reports after each class session, what average percentage of curriculum content was taught?	Online facilitator fidelity checklists	Immediate post-program, for all workshop series	Program staff, evaluation staff
Dosage	2. For each condition, what percentage of individuals attended: 0 sessions, at least 1 session, 50% or more of the sessions, and all 6 of the sessions offered?	Workshop session attendance in nFORM	Immediately after all class sessions delivered	Program staff, evaluation staff
Quality	3. For each condition, what is the average participant post-program rating of facilitation quality?	Local evaluation participant surveys: 5-item facilitation quality subscale of CEES (Adler-Baeder et al., 2023)	Immediate post-program survey (i.e., T2 for weekly participants; T3 for monthly participants)	Evaluation staff
Engagement	4. For each condition, what is the average participant post-program rating of self-engagement during the class series?	Local evaluation participant surveys: 3-item individual engagement subscale of CEES (Adler-Baeder et al., 2023)	Immediate post-program survey (i.e., T2 for weekly participants; T3 for monthly participants)	Evaluation staff
Context	5. For each condition, what percentage of participants reported participating in therapy/counseling (outside of AHMRE program services) during the study period from baseline to one-year follow-up?	Local evaluation participant surveys	T4 survey (i.e., 1-year post-baseline) survey	Evaluation staff

Table B.3. Outcome measures, items, value scoring, and description of composite calculation for all primary outcomes

RQs	Outcome Measure	Items	Values	Description of Composite
1, 4	Couple Relationship Skills Inventory	**Care for Self Subscale:** 1. I have the power to manage the challenges in my life. 2. I ask for help from others when needed. 3. I recognize my strengths. 4. I manage the stress in my life. 5. I eat healthy meals every day. 6. I exercise at least 3 or more times a week. 7. I get 7-8 quality hours of sleep every night. 8. I have quiet time for myself every day. **Choose Subscale:** 1. I want this relationship to stay strong no matter what rough times we encounter. 2. I commit effort every day to making my relationship work. 3. I always think about how my choices could affect my relationship. 4. I always make an effort to focus on my partner's strengths. **Know Subscale:** 1. I know my partner's current life stresses. 2. I know some of my partner's major aspirations and hopes in life. 3. I know my partner's current major worries. 4. I know my partner pretty well. **Share Subscale:** 1. Had a stimulating exchange of ideas. 2. Engage in and/or talk about outside interests together. 3. Made time to touch base with each other. **Care Subscale:** 1. Say "I love you" to your partner. 2. Initiate physical affection with your partner (e.g., kiss, hug) 3. Share emotions, feelings, or problems with your partner. 4. Tell your partner things you appreciate about them and how much you care for them. **Manage Subscale:** 1. I am able to see my partner's point of view and really understand it, even if I don't agree. 2. When things "get heated" I suggest we take a break to calm down. 3. I can easily forgive my partner. 4. I shout or yell at my partner. (RC) 5. I blame, accuse, or criticize my partner. (RC)	**Care for Self Subscale:** 1 = Never to 7 = More than once a day **Choose/Know/ Share/Care/ Manage/Connect Subscales:** 1 = Very Strongly Disagree to 7 = Very Strongly Agree	Mean composite of all 32 items across all subscales (value range 1-7); higher scores represent better skills

RQs	Outcome Measure	Items	Values	Description of Composite
		Connect Subscale: 1. Many of our friends are friends of both of us. 2. We know people who care about us and our relationship. 3. If we were to need help getting by or encountered a crisis, we would have friends and family to rely on. 4. As a couple, we try to help others in need.		
2, 5	Couple Satisfaction Index	1. Please indicate the degree of happiness, all things considered, of your relationship. 2. I have a warm and comfortable relationship with my partner. 3. How rewarding is your relationship with your partner? 4. In general, how satisfied are you with your relationship?	**Item 1:** 1 = Extremely Unhappy to 7 = Perfect **Items 2-4:** 1 = Not at all to 6 = Completely	Sum composite of all 4 items (value range 4 - 25); higher scores represent higher satisfaction
3, 6	Kessler Psychological Distress Scale	Please select how often you have felt each of the following in the past 30 days. 1. Did you feel tired out for no good reason? 2. Did you feel nervous? 3. Did you feel so nervous that nothing could calm you down? 4. Did you feel hopeless? 5. Did you feel restless or fidgety? 6. Did you feel so restless that you could not sit still? 7. Did you feel depressed? 8. Did you feel that everything was an effort? 9. Did you feel so sad that nothing could cheer you up? 10. Did you feel worthless?	1 = None of the time to 5 = All of the time	Mean composite of all 10 items (value range 1-5); higher scores represent higher mental health difficulties

Note: Reverse-scored items are marked with (RC) and were recoded before creating the composite score.

Table B.5. Implementation RQ measure items and values/scoring

RQ	Implementation Measure Items & Values/Scoring
Fidelity #1	**Session 1:** Welcome & Icebreaker; Overview; Introductions & Ground Rules; Expectations; From Research to Relationships; Program Goals; Heart-Brain Connection; At the Heart of the Matter; Heart and Health; Elevate Your Love; Pulse Check **Session 2:** Individual Wellness Affects Relationships; Empowering Your Relationship; Stressors Crowd Us; Recognizing Signs of Stress; Managing Stress; Mindfulness Practice: Partner Connection; Take Your Pulse; Physical Health; Spiritual Health; Sexual Health; Emotional & Social Health; Summary & Action Plan; Laying the Foundation; Intentionality; Make the Relationship a Priority **Session 3:** Putting the Relationship First; Do not Let Other Interests Interfere; Working on Building the Relationship; Focus on Each Other's Strengths; Envision a Healthy Future Together; Laying the Foundation Summary & Action Plan; Enlighten Introduction; Remembering Our Past; The Story of Us; What is Intimacy; Love Maps; Let's Test Your Love Maps; Clear Expectations; Expectations About Money; Enlighten Summary & Action Plan **Session 4:** Mindfulness Practice: Loving Kindness; Take Your Pulse; Value, The Many Ways You Care; Remembering the Good; Appreciating the Positive; When We Feel Crabby; Caring (Not Crabby) Actions; Value Summary & Self-Evaluation; Building Friendship; Spending Meaningful Time Together; Making & Protecting Time **Session 5:** Fostering Your Couple Identity; Nurturing Positive Interactions; Filtered Messages; Unhelpful Messages; Helpful Messages; Attach Summary & Action Plan; Conflict is Normal; How We Respond to Anger; Fight or Flight Response; Anger [Mis]Management; Apply the Breaks; Mindfulness Practice: Attitude Change; Pulse Check; Find Your Calm, Then Carry On **Session 6:** Go Slow and Both W.I.N.; Practice W.I.N.; Repair Attempts; Tame Summary & Action Plan; Recap: ELEVATE Your Relationship; Engaging a Positive Support Network; Support from Family & Friends; Community Connections; Giving to Others; Connection to Sources of Meaning; Engage Summary & Action Plan; Wrap: ELEVATE Your Relationship **Values/Scoring:** All 79 session activities were marked for completion (1=Yes, 0=No)
Dosage #2	Attendance for each individual in the couple was marked after each of the six 2-hour sessions **Values/Scoring:** 0=0 sessions, 1=at least 1 session, 2=50% or more of sessions, 3=all 6 sessions
Quality #3	Facilitation Quality subscale of the CEES (Adler-Baeder et al., 2023) 1. The facilitators explained the course material clearly. 2. The facilitators effectively encouraged class participation. 3. The facilitators managed classroom comments and behavior appropriately. 4. The facilitators managed the class time well. 5. The facilitators shared appropriate examples and experiences to help us understand the course material. **Values/Scoring:** 1=strongly disagree, 2=disagree, 3=neither/mixed, 4=agree, 5=strongly agree
Engagement #4	Individual Engagement subscale of the CEES (Adler-Baeder et al., 2023) 1. I felt connected to other participants in the class. 2. I felt comfortable sharing experiences with the group. 3. I was actively engaged in group discussions and activities. **Values/Scoring:** 1=strongly disagree, 2=disagree, 3=neither/mixed, 4=agree, 5=strongly agree

RQ	Implementation Measure Items & Values/Scoring
Context #5	Are you currently or have you previously participated in therapy/counseling? If you have participated in therapy/counseling more than once, please answer by thinking about your most recent experience. **Values/Scoring:** 1=no; 2=yes it began < 3 months ago; 3=yes it began 4-6 months ago; 4=yes it began 7-12 months ago; 5=yes it began 13+ months ago

Figure B.1. CONSORT diagram for individuals, for studies in which consent occurred after assignment

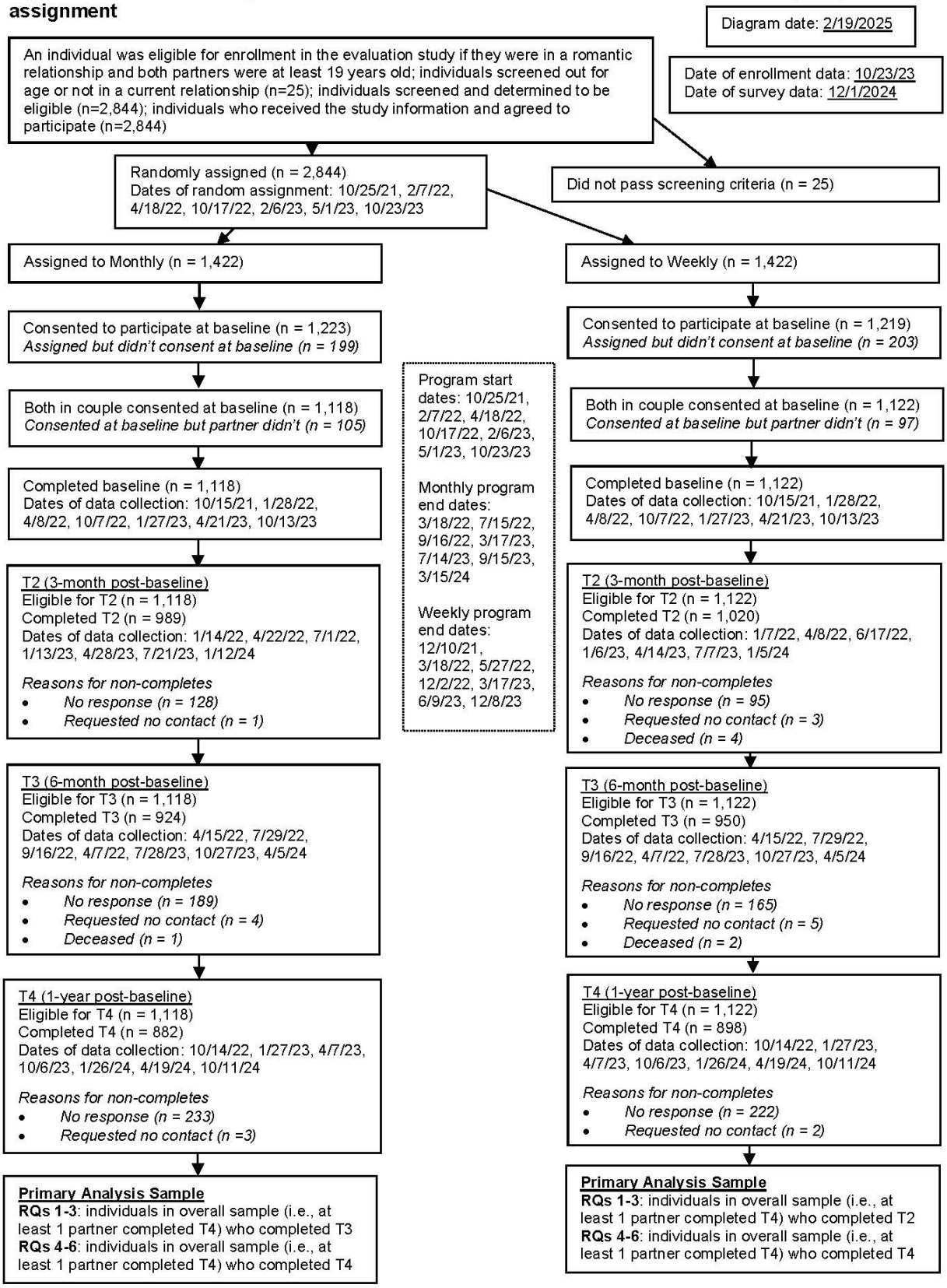


Table B.6 Flow of individual participants through stages of the impact study, by condition

Stage of impact study	Monthly Program Condition	Weekly Program Condition
Randomly assigned to condition	1,422	1,422
Consented to participate at T1	1,223	1,219
Assigned but didn't consent	*199*	*203*
Both in couple consented at T1	1,118	1,122
Consented at T1 but partner didn't	*105*	*97*
T1 Survey		
Completed T1	1,118	1,122
Dates of data collection	10/15/21, 1/28/22, 4/8/22, 10/7/22, 1/27/23, 4/21/23, 10/13/23	10/15/21, 1/28/22, 4/8/22, 10/7/22, 1/27/23, 4/21/23, 10/13/23
T2 Survey		
Eligible for T2	1,118	1,122
Completed T2	989	1,020
Dates of data collection	1/14/22, 4/22/22, 7/1/22, 1/13/23, 4/28/23, 7/21/23, 1/12/24	1/7/22, 4/8/22, 6/17/22, 1/6/23, 4/14/23, 7/7/23, 1/5/24
No response	*128*	*95*
Requested no contact	*1*	*3*
Deceased		*4*
T3 Survey		
Eligible for T3	1,118	1,122
Completed T3	924	950
Dates of data collection	4/15/22, 7/29/22, 9/16/22, 4/7/22, 7/28/23, 10/27/23, 4/5/24	4/15/22, 7/29/22, 9/16/22, 4/7/22, 7/28/23, 10/27/23, 4/5/24
No response	*189*	*165*
Requested no contact	*4*	*5*
Deceased	*1*	*2*
T4 Survey		
Eligible for T4	1,118	1,122
Completed T4	882	898
Dates of data collection	10/14/22, 1/27/23, 4/7/23, 10/6/23, 1/26/24, 4/19/24, 10/11/24	10/14/22, 1/27/23, 4/7/23, 10/6/23, 1/26/24, 4/19/24, 10/11/24
No response	*233*	*222*
Requested no contact	*3*	*2*
Primary Analysis Sample for:		
RQs 1-3	Individuals in a couple in which at least 1 partner completed T4 surveys; and who completed T3 survey	Individuals in a couple in which at least 1 partner completed T4 surveys; and who completed T3 survey
RQs 4-6	Individuals who completed T4 survey	Individuals who completed T4 survey

Figure B.2. CONSORT diagram for couples, for studies in which consent occurred after assignment

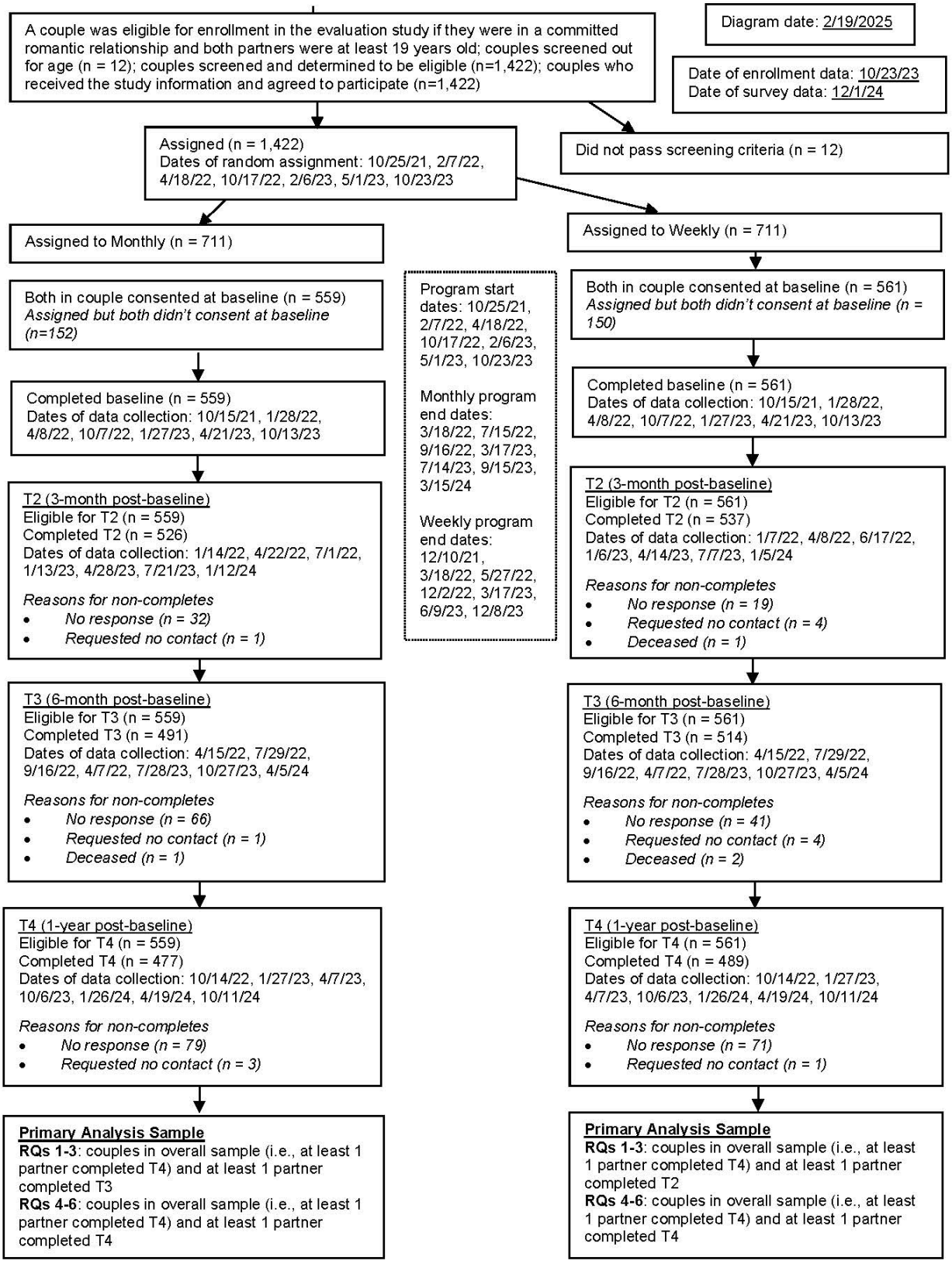


Table B.7 Flow of couples through stages of the impact study, by condition

Stage of impact study	Monthly Program Condition	Weekly Program Condition
Randomly assigned to condition	711	711
Both in couple consented at T1	559	561
Consented at T1 but partner didn't	*152*	*150*
T1 Survey		
Completed T1	559	561
Dates of data collection	10/15/21, 1/28/22, 4/8/22, 10/7/22, 1/27/23, 4/21/23, 10/13/23	10/15/21, 1/28/22, 4/8/22, 10/7/22, 1/27/23, 4/21/23, 10/13/23
T2 Survey		
Eligible for T2	559	561
Completed T2	526	537
Dates of data collection	1/14/22, 4/22/22, 7/1/22, 1/13/23, 4/28/23, 7/21/23, 1/12/24	1/7/22, 4/8/22, 6/17/22, 1/6/23, 4/14/23, 7/7/23, 1/5/24
No response	*32*	*19*
Requested no contact	*1*	*4*
Deceased		*1*
T3 Survey		
Eligible for T3	559	561
Completed T3	491	514
Dates of data collection	4/15/22, 7/29/22, 9/16/22, 4/7/22, 7/28/23, 10/27/23, 4/5/24	4/15/22, 7/29/22, 9/16/22, 4/7/22, 7/28/23, 10/27/23, 4/5/24
No response	*66*	*41*
Requested no contact	*1*	*4*
Deceased	*1*	*2*
T4 Survey		
Eligible for T4	559	561
Completed T4	477	489
Dates of data collection	10/14/22, 1/27/23, 4/7/23, 10/6/23, 1/26/24, 4/19/24, 10/11/24	10/14/22, 1/27/23, 4/7/23, 10/6/23, 1/26/24, 4/19/24, 10/11/24
No response	*79*	*71*
Requested no contact	*3*	*1*
Primary Analysis Sample for:		
RQs 1-3	Couples in which at least 1 partner completed T4 surveys; and at least 1 partner completed T3 survey	Couples in which at least 1 partner completed T4 surveys; and at least 1 partner completed T3 survey
RQs 4-6	Couples in which at least 1 partner completed T4 survey	Couples in which at least 1 partner completed T4 survey

Appendix C: Attrition and Baseline Equivalence

C.1. Attrition Rates

Table C.1a. Overall and differential attrition rates by time point in the overall sample at individual- and couple-levels

	Individual-Level Attrition				Couple-Level Attrition			
Time point	MONTHLY Condition	WEEKLY Condition	Differential Between conditions	Overall Attrition	MONTHLY Condition	WEEKLY Condition	Differential Between conditions	Overall Attrition
T1	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
T2	11.5%	9.1%	2.4%	10.3%	5.9%	4.3%	1.6%	5.1%
T3	17.4%	15.3%	2.1%	16.3%	12.2%	8.4%	3.8%	10.3%
T4	21.1%	20.0%	1.1%	20.5%	12.3%	11.6%	0.7%	12.0%

Notes: For numbers of surveys at each time point, see Table IV.1b. Individual-level attrition rates were calculated as number of individuals with survey data / number of total individuals in full sample (n=2,240). Differential between conditions was calculated as the absolute value of the difference between Monthly and Weekly program condition attrition rates. Couple-level attrition rates were calculated as number of couples in which at least one partner had survey data / number of total couples in full sample (n=1,120). According to WWC Standards (2022), the cautious boundary for differential attrition between conditions of a study with overall attrition of 21% (i.e., this study's overall attrition at the individual-level) is 5.3%; all individual-level differential attrition rates are below that cautious boundary. According to WWC Standards (2022), the cautious boundary for differential attrition between conditions of a study with overall attrition of 12% (i.e., this study's overall attrition at the couple-level) is 6.2%; all couple-level differential attrition rates are below that cautious boundary. Thus, this study classifies as a low attrition study at both the individual- and couple-levels.

Table C.1b. Overall and differential attrition rates at the final time point in the analytic samples for each RQ at individual- and couple-levels

	Individual-Level Attrition				Couple-Level Attrition			
RQ	MONTHLY Condition	WEEKLY Condition	Differential Between conditions	Overall Attrition	MONTHLY Condition	WEEKLY Condition	Differential Between conditions	Overall Attrition
RQ 1	18.6%	15.9%	2.7%	17.2%	18.6%	15.9%	2.7%	17.2%
RQ 2	18.8%	15.7%	3.1%	17.2%	18.8%	15.7%	3.1%	17.2%
RQ 3	16.1%	13.0%	3.1%	14.6%	16.1%	13.0%	3.1%	14.6%
RQ 4	17.5%	18.9%	1.2%	18.2%	17.5%	18.9%	1.2%	18.2%
RQ 5	18.1%	18.0%	0.1%	18.0%	18.1%	18.0%	0.1%	18.0%
RQ 6	12.7%	12.3%	0.4%	12.5%	12.7%	12.3%	0.4%	12.5%

Notes: The final time point for RQs 1-3 was immediate post-program (i.e., T2 for weekly condition; T3 for monthly condition). The final time point for RQs 4-6 was 1-year follow-up (i.e., T4 for both conditions). Since individuals were included in each RQ sample if one or both partners had 80% of outcome items at the relevant time points, the individual- and couple-level samples and attrition are identical. According to WWC Standards (2022), all individual- and couple-level differential attrition rates are below the cautious boundary for the overall attrition in each RQ (RQ 1: overall attrition=17%; cautious boundary=5.8%; differential attrition=2.7%; RQ 2: overall=17%; boundary=5.8%; differential=3.1%; RQ 3: overall=15%; boundary=5.9%; differential=3.1%; RQ 4: overall=18%; boundary=5.7%; differential=1.2%; RQ 5: overall=18%; boundary=5.7%; differential=0.1%; RQ 6: overall=13%; boundary=6.1%; differential=0.4%). Thus, this study classifies as a low attrition study at both the individual- and couple-levels for each RQ sample.

Appendix D: Data preparation

Proactive measures during data collection. During the startup period of this grant cycle, the local evaluation team piloted the Qualtrics local evaluation surveys (multiple times and with a wide array of respondents and possible participant scenarios). We accordingly edited surveys for readability and for accuracy in data downloads (e.g., item names, values, etc.). The evaluation team also downloaded, merged, and checked survey data frequently throughout the impact study period. Additionally, the evaluation and program teams actively worked to troubleshoot any challenges participants experienced during the data collection process. During each survey window period, local evaluation staff were available in real time to respond to participants' questions or challenges as they occurred. Immediate efforts were made to validate and/or correct data entries (e.g., if partners accidentally took each other's surveys), as well as to anticipate and address any other factors that could have hindered accurate data collection.

Data cleaning (after data collection was complete). Local evaluation survey responses for each cohort across each time point were downloaded from Qualtrics into separate SPSS databases. These databases were merged based on participant individual ID to create a master dataset with survey responses from each of the seven cohorts across each of the four time points. Basic descriptive statistics were run on all variables of interest to assess for outliers or impossible scores on each scale. Data plots (i.e., bar charts, scatter plot graphs) were also used to look for outliers or improbable values. We also spot-checked data by randomly selecting cases and checking their raw scores against the overall dataset. After all error-checking, impossible scores were removed, and outliers were assessed for accuracy on a case-by-case basis. All adjustments for data errors were documented.

Consistency in demographic data was assessed across participants and time points. Evaluation staff used a master database containing participant IDs and demographic data (obtained from both the nFORM Applicant Characteristics Survey and local evaluation T1 survey) to verify accuracy in demographic data within individuals across time points. Responses that were obvious keystroke errors based on master database information and inconsistency with other time points were corrected. We also assess couple-level covariates between partners for consistency. If a couple disagreed on marital status (i.e., one reports married and one reports unmarried), we deferred to the unmarried report and made adjustments to data to ensure consistency between partners on couple-level covariates. To ensure that all study measures performed as expected, we assessed the internal consistency of each measure (via Cronbach's alpha), as well as the associations among the measures across time points (via Pearson correlation coefficients).

Missing Data Imputation. Missing data in this study was due to either of these scenarios: (1) cases missing <=20% of items on outcome measures at baseline or relevant follow-up for each analytic sample, or (2) cases in which both partners had baseline data and only one partner had follow-up outcome data (since this was a low attrition study and we were able to retain these partners missing follow-up data in the analytic sample). Missing data were addressed through multiple imputation using the 'mice' package in RStudio (van Buuren & Groothuis-Oudshoorn, 2011; R Core Team, 2021). The mice method, which stands for Multivariate Imputation by Chained Equations, replaces missing values by leveraging other observed data points within the dataset as covariates, generating multiple possible values based on the responses of all participants. To estimate missing values, the classification and regression trees (CART) method was implemented using available raw data. The imputation process considered various covariates, including sex, race, age, education, employment, individual and household income, married status, relationship length, parental status, and other demographic factors. Ultimately, multiple imputation produced 10 imputed datasets, and the mean values from these datasets were pooled into a single dataset for analysis.

Table D.1. Internal consistency of outcome measures across time points

Outcome Measure	T1 Baseline (female | male)	T2 3-Month Follow-Up (female | male)	T3 6-Month Follow-Up (female | male)	T4 1-Year Follow-Up (female | male)
Couple Relationship Skills	.90 | .90	.93 | .92	.93 | .93	.93 | .93
Couple Satisfaction	.92 | .88	.90 | .86	.90 | .90	.90 | .85
Mental Health Symptoms	.91 | .91	.93 | .93	.93 | .94	.93 | .93

Note: Internal consistency was calculated using Cronbach's alpha of all items in each measure, after reverse-scoring necessary items. Analyses were run separately for women and men due to dependence of couple-level data. Internal consistencies indicated good to excellent reliabilities.

Table D.2. Correlations among outcome measures across time points

	1	2	3	4	5	6	7	8	9	10	11	12
1. T1 CRSI	--	.71***	.66***	.69***	.70***	.54***	.53***	.56***	-.28***	-.23***	-.22***	-.26***
2. T2 CRSI	.72***	--	.65**	.68***	.53***	.68***	.51***	.48***	-.25***	-.33***	-.24***	-.26***
3. T3 CRSI	.66***	.72***	--	.73***	.52***	.49***	.72***	.52***	-.21***	-.25***	-.30***	-.27***
4. T4 CRSI	.65***	.73***	.75***	--	.53***	.55***	.59***	.72***	-.26***	-.29***	-.27***	-.39***
5. T1 CSI	.72***	.59***	.54***	.55***	--	.65***	.62***	.64***	-.26***	-.18***	-.19***	-.19***
6. T2 CSI	.58***	.74***	.56***	.57***	.69***	--	.64***	.64***	-.24***	-.27***	-.21***	-.23***
7. T3 CSI	.51***	.57***	.73***	.59***	.64***	.65***	--	.65***	-.21***	-.23***	-.30***	-.27***
8. T4 CSI	.54***	.60***	.60***	.77***	.63***	.65***	.70***	--	-.22***	-.21***	-.22***	-.31***
9. T1 Kessler	-.32***	-.24***	-.21***	-.19***	-.32***	-.21***	-.18***	-.17***	--	.68***	.60***	.63***
10. T2 Kessler	-.23***	-.35***	-.24***	-.21***	-.23***	-.31***	-.21***	-.20***	.68***	--	.69***	.70***
11. T3 Kessler	-.16***	-.21***	-.30***	-.22***	-.16***	-.15***	-.27***	-.17***	.68***	.68***	--	.70***
12. T4 Kessler	-.14***	-.17***	-.18***	-.28***	-.16***	-.14***	-.17***	-.24***	.62***	.64***	.68***	--

Notes: CRSI=Couple Relationship Skills Inventory outcome (mean composite; value range 1-7); CSI=Couple Satisfaction Index outcome (sum composite; value range 4-25). Kessler = Mental Health symptoms (mean composite; value range 1-5). Analyses were run separately for women and men; women's results are below the diagonal, and men's results are above the diagonal. Pearson correlation coefficients were utilized to check for associations between outcome measures across time points. ***p<.001.

Table D.3a. Correlations between covariates and outcome measures for women

	T1 CRSI	T2 CRSI	T3 CRSI	T4 CRSI	T1 CSI	T2 CSI	T3 CSI	T4 CSI	T1 Kessler	T2 Kessler	T3 Kessler	T4 Kessler
Class setting	.03	.00	.01	.03	.08*	.01	.05	.05	-.09**	-.04	-.07*	-.09**
Age	-.06	-.08*	-.10**	-.09*	-.12***	-.09**	-.11***	-.10**	-.18***	-.16***	-.18***	-.20***
Race	-.21***	-.17***	-.13***	-.16***	-.17***	-.16***	-.11***	-.12***	-.03	-.02	-.04	-.10**
Education	.08**	.10**	.07*	.09**	.10***	.11***	.10**	.11***	-.19***	-.15***	-.19***	-.15***
Parent status	-.20***	-.17***	-.15***	-.16***	-.19***	-.16***	-.16***	-.16***	-.07*	-.04	-.08*	-.10**
Married status	.09**	.04	.01	.01	.10***	.07*	.05	.07*	-.16***	-.13***	-.12***	-.16***

Notes: CRSI=Couple Relationship Skills Inventory outcome (mean composite; value range 1-7); CSI=Couple Satisfaction Index outcome (sum composite; value range 4-25). Kessler = Mental Health symptoms (mean composite; value range 1-5). Analyses were run separately for women and men, due to dependence of couple-level data. Pearson correlation coefficients were utilized to check for associations between outcome measures and covariates. See Table IV.3 for description of demographic covariates. *p<.05; **p<.01; ***p<.001.

Table D.3b. Correlations between covariates and outcome measures for men

	T1 CRSI	T2 CRSI	T3 CRSI	T4 CRSI	T1 CSI	T2 CSI	T3 CSI	T4 CSI	T1 Kessler	T2 Kessler	T3 Kessler	T4 Kessler
Class setting	.14***	.03	.06	.13***	.11***	.07*	.05	.12***	-.09**	-.05	-.03	-.05
Age	-.09**	-.10**	-.15***	-.12**	-.07*	-.04	-.06	-.08*	-.17***	-.18***	-.15***	-.16***
Race	-.16***	-.12***	-.12***	-.17***	-.14***	-.09**	-.08*	-.11**	-.07*	-.10**	-.08*	-.09*
Education	.20***	.20***	.14***	.23***	.12***	.11***	.08*	.11**	-.17***	-.12***	-.13***	-.13***
Parent status	-.17***	-.15***	-.10**	-.15***	-.15***	-.09**	-.10**	-.10**	-.06	-.06	-.05	-.08*
Married status	.08*	.01	.07	.04	.06	.02	.02	.05	-.11***	-.08*	-.07*	-.06

Notes: CRSI=Couple Relationship Skills Inventory outcome (mean composite; value range 1-7); CSI=Couple Satisfaction Index outcome (sum composite; value range 4-25). Kessler = Mental Health symptoms (mean composite; value range 1-5). Analyses were run separately for women and men, due to dependence of couple-level data. Pearson correlation coefficients were utilized to check for associations between outcome measures and covariates. See Table IV.3 for description of demographic covariates. *p<.05; **p<.01; ***p<.001.

Table D.4. Missing data on each outcome for each RQ analytic sample, separately by condition

		MONTHLY Program Condition					WEEKLY Program Condition				
RQ	Outcome Measure	# Total cases in full HMRF sample	# Cases with 80% of outcome items on T1 & relevant follow-up	# Cases dropped	# Cases imputed	Missingness Rate	# Total cases in full HMRF sample	# Cases with 80% of outcome items on T1 & relevant follow-up	# Cases dropped	# Cases imputed	Missingness Rate
1	CRSI	1118	956	162	91	14.5%	1122	1040	82	76	7.3%
2	CSI	1118	956	162	93	14.5%	1122	1046	76	86	6.8%
3	Kessler	1118	982	136	77	12.2%	1122	1070	52	65	4.6%
4	CRSI	1118	922	196	112	17.5%	1122	910	212	97	18.9%
5	CSI	1118	916	202	112	18.1%	1122	920	202	106	18.0%
6	Kessler	1118	976	142	115	12.7%	1122	984	138	108	12.3%

Notes: CRSI=Couple Relationship Skills Inventory (i.e., 32 total items; cases missing >20% of items [i.e., missing 7-32 items] on either T1 or relevant follow-up survey were dropped; cases missing <=20% of items [i.e., missing 1-6 items] on both T1 and relevant follow-up survey were imputed). CSI=Couple Satisfaction Index (i.e., 4 total items; cases missing >20% of items [i.e., missing 2-4 items] on either T1 or relevant follow-up survey were dropped; cases missing <=20% of items [i.e., missing 1 item] on both T1 and relevant follow-up survey were imputed). Kessler=Kessler mental health symptoms (i.e., 10 total items; cases missing >20% of items [i.e., missing 3-10 items] on either T1 or relevant follow-up survey were dropped; cases missing <=20% of items [i.e., missing 1-2 items] on both T1 and relevant follow-up survey were imputed). Missingness rate is a percentage, calculated as # of cases dropped / #total cases in full HMRF sample. See Table IV.1b for overview of individual sample sizes by condition and outcome.

Appendix E: Impact estimation

Baseline Equivalence. Hedge's *g* effect size was used to assess baseline equivalence on continuous demographics/outcomes, and Cox's index effect size was used to assess baseline equivalence on binary demographics.

The equation for **Hedge's g effect size** calculation is:

$$g = \frac{\overline{X_1} - \overline{X_2}}{S^*}$$

where

- $\overline{X_1}$ = Mean of group 1
- $\overline{X_2}$ = Mean of group 2
- S^* = Pooled standard deviation with correction for small sample bias

The pooled standard deviation is calculated as:

$$S^* = \sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}}$$

where

- S_1^2 = Variance of group 1
- S_2^2 = Variance of group 2
- n_1 = Sample size of group 1
- n_2 = Sample size of group 2

To correct for small sample bias, Hedge's g is adjusted using J correction:

$$g_{\text{corrected}} = g \times \left(1 - \frac{3}{4(n_1 + n_2) - 9}\right)$$

where

- n_1 = Sample size of group 1
- n_2 = Sample size of group 2

The equation for **Cox's index effect size** calculation is:
g = [log(odds intervention) −log(odds control)]/1.65

Primary impact estimation (RQs #1-6). We used multilevel models to predict the value of the outcome at follow-up by randomly assigned condition (monthly or weekly), while accounting for baseline levels of the outcome and any significantly correlated covariates. In other words, these models predict future behaviors based on past behaviors and the outcome can be interpreted as residual change (Hyndman & 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 for individual i in couple j
- β_{0j} = the baseline mean of the outcome for couple j in the weekly group
- β_1 = the average treatment effect on the outcome
- RA_{ij}=1 for monthly condition, and RA_{ij}=0 for weekly condition
- 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:

γ_{00} = mean outcome for all couples
μ_{0j} = unique effect of couple j on the mean outcome.

Multiple Imputation (regression-based):

Summary of Equations:

- Overall parameter estimate:

$$\bar{\theta} = \frac{1}{m} \sum_{k=1}^{m} \hat{\theta}^{(k)}$$

- Within-imputation variance:

$$\bar{U} = \frac{1}{m} \sum_{k=1}^{m} \mathrm{Var}(\hat{\theta}^{(k)})$$

- Between-imputation variance:

$$B = \frac{1}{m-1} \sum_{k=1}^{m} (\hat{\theta}^{(k)} - \bar{\theta})^2$$

- Total variance:

$$T = \bar{U} + \left(1 + \frac{1}{m}\right) B$$

- Standard error:

$$\mathrm{SE}(\bar{\theta}) = \sqrt{T}$$