

Six Month Impact Evaluation of Prevention and Relationship Education
Program's (PREP's) *Within Our Reach* in Orlando, Florida
Final Impact Evaluation Report for
University of Central Florida

December 21, 2020

Prepared by

Ryan G. Carlson, Tammiee Dickenson, Bradley D. Rogers, Christopher J.
Hipp, & Genee Glascoe

University of South Carolina

According to the Paperwork Reduction Act of 1995, an agency may not conduct or sponsor, and a person is not required to respond to, a collection of information unless it displays a valid OMB control number. The valid OMB control number for this collection is 0970-0356; this number is valid through 6/30/2021. Public reporting burden for this collection of information is estimated to average 30 hours, including the time for reviewing instructions, gathering and maintaining the data needed, reviewing the collection of information, and revising it. This collection of information is voluntary for individuals, but the information is required from Grantees.

Recommended Citation:

Carlson, R. G., Dickenson, T., Rogers, B. D., Hipp, C. J., & Glascoe, G. (2020). Six Month Impact Evaluation of Prevention and Relationship Education (PREP's) Within Our Reach in Orlando, Florida - Final Impact Evaluation Report for University of Central Florida. Submitted to the Administration for Children and Families, Office of Family Assistance. Washington, DC.

Acknowledgements:

We would like to acknowledge several people who contributed countless hours to this evaluation. Specifically, Dr. Shannon Romagnolo who helped develop and establish the initial data protocol; Dr. Rufus Wofford who assisted with quality control efforts and running regular data reports; Dr. Jenn Olejarczyk for her statistical consultation; and Anthony Carr, Hayden Rothschild, and Madeline Moran for their efforts with entering data and conducting quality control. We would also like to thank Dr. Derekh Cornwell, our technical assistance provider from Mathematica, for his detailed feedback and guidance throughout the evaluation.

We must also acknowledge the tireless efforts of the University of Central Florida Project Harmony implementation team. We conducted several site visits and held numerous meetings during the past five years. This data and report would not be possible without the dedication, drive, and determination of the entire staff. Drs. Sejal Barden and Dalena Dillman Taylor, along with Project Directors Bridgette Toussaint and Nicole Silverio, led the implementation teams and worked with our evaluation team on the delicate balance between service provision and research integrity. Additionally, recruitment leader Maria Riccardi, family services lead Nakita Carroll, and operations manager, Nataly Larenas, worked tirelessly to meet project benchmarks, and use data to drive their decisions. We also appreciate Angel Ramos who was the data coordinator on site at UCF and responded to our numerous requests over the years. The fruits of all their efforts cannot truly be measured in this evaluation.

Finally, we would like to acknowledge the couples who participated in Project Harmony over the five-year study. Many couples dedicated their time and made sacrifices to attend the project regularly. They made these sacrifices out of dedication to their relationships and families. We could not advance the field and provide data that will help future families without their participation. We are deeply grateful for their participation and for entrusting us with their responses.

Disclosure:

We have no conflicts to report.

This publication was prepared under Grant Number 90FM0078-02-00 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: “The Evaluation of PREP’s *Within Our Reach* in Orlando, Florida”

Objective. The primary study objective was to evaluate the impact of the *Within Our Reach* curriculum on a low-income sample of couples. We sought to assess the curriculum’s effects on relationship outcomes (relationship satisfaction and dyadic coping); individual outcomes (emotional regulation and individual distress); and parenting and family adjustment. We also conducted an implementation analysis in order to understand participants’ satisfaction with services, their relationship with the implementation staff, and how services were delivered.

Study design. We randomly assigned 1,418 couples to receive either PREP’s *Within Our Reach* 12-hour curriculum, or to a six-month wait-list control group. Couples assigned to the immediate intervention attended four consecutive week-night workshops onsite. We collected data on individual, couple, and parenting outcomes at baseline, 30, 90, and 180-day follow-ups. Couples assigned to the wait-list received a brief, eight-hour, version of the curriculum 180-days after random assignment and after all data collection for the evaluation ended.

Results. The impact evaluation presented in the current report focuses on outcomes at the six-month (180-day) study period. Results using multi-level model comparisons between treatment and wait-list control group couples revealed non-significant, intervention effects at the six-month follow-up on most study outcomes. However, there was one small, positive, statistically significant intervention effect on dyadic coping at the six-month follow-up. All couples enrolled in the study demonstrated positive rates of change over the six-month study period, but differences in the rates of change between intervention and comparison group couples were not statistically significant.

Conclusion. Results of the six-month impact evaluation demonstrated one statistically significant intervention effect between treatment and control group couples. The limited statistically significant effects at six-months may be a result of high rates of overall and differential attrition. Treatment group couples were less likely to complete the six month follow-up survey than wait-list control group couples. Additionally, this study included an active control group in order to keep wait-list control couples engaged during the study period. Wait-list couples may have benefited from some of the engagement workshops provided, as well as from the knowledge that they would eventually receive the relationship education intervention.

Contents

I.	INTRODUCTION	1
A.	Introduction and study overview	1
B.	Primary research question(s): Primary outcome measures were analyzed for couples at the six-month follow-up period.....	3
C.	Secondary research question(s): Secondary research questions addressed outcomes for sub-populations (e.g., parents) within the primary sample. Additionally, secondary questions (2-4) included examination of the rate of change, or the growth curve, for major intervention target areas over the six-month study period. Thus, the growth curve models estimated the rate of change over three measurement points after random assignment (one month, three months, and six month follow-up).....	3
II.	INTERVENTION AND COUNTERFACTUAL CONDITIONS	4
A.	Description of program as intended.....	4
B.	Description of counterfactual condition as intended.....	6
C.	Research Questions about the intervention and counterfactual conditions as implemented	7
III.	STUDY DESIGN	9
A.	Sample formation and research design	9
B.	Data collection	10
IV.	ANALYSIS METHODS.....	12
A.	Analytic sample.....	12
B.	Outcome measures.....	14
C.	Baseline equivalence and sample characteristics.....	17
V.	FINDINGS AND ESTIMATION APPROACH.....	19
A.	Implementation evaluation	19
B.	Primary impact evaluation	22
C.	Sensitivity analyses.....	27
VI.	DISCUSSION.....	29
VII.	REFERENCES	33
VIII.	APPENDICES	36
A.	Logic Model (or Theory of CHANGE) for the program	36

B. Data and study sample.....37

C. CONSORT diagram for individual clients, for studies in which consent occurred before assignment.....41

D. Data preparation42

E. Impact estimation43

F. Sensitivity analyses and alternative model specifications48

Tables

II.1.	Description of intended intervention and counterfactual components and target populations	5
II.2.	Staff training and development to support intervention and counterfactual components.....	5
II.3.	Examples of research questions for each implementation element and study group	7
IV.1a.	Individual sample sizes by intervention status	13
IV.2.	Outcome measures used for primary impact analyses research questions.....	14
IV.3.	Outcome measures used for secondary impact analyses research questions	16
IV.4.	Summary statistics of key baseline measures and baseline equivalence across study groups, for individuals/couples completing the 180-day follow up data collection	18
V.A.1.	“Quality” items percentages.....	21
V.A.2.	PREP Alliance percentages.....	21
V.1.	Covariates included in impact analyses	26
V.2.	Post-intervention estimated effects using data from the RDAS, DCI, DERS, and OQ at 180-day follow up data collection to address the primary research questions	26
V.4.	Post-intervention estimated effects using data from the parenting measure at the 180-day follow up and from the RDAS, DCI, DERS, and OQ over the four data collection time points to address the secondary research questions	27
V.3.	Differences in parameter estimates between intervention and comparison groups estimated using alternative methods for secondary research questions 2-5	28
B.1.	Data used to address implementation research questions.....	39
B.2.	Key features of the impact analysis data collection.....	40

Figures

V.A.1. Percentage of responses to “How did you hear about the program”	20
C.1. CONSORT diagram.....	41
F.1. Growth patterns on relationship satisfaction over the six-month study period. This was modeled for secondary research question two.	48
F.2. Growth patterns on dyadic coping over the six-month study period. This was modeled for secondary research question three.	49
F.3. Growth patterns on emotion regulation over the six-month study period. This was modeled for secondary research question four.	49
F.4. Growth patterns on individual distress over the six-month study period. This was modeled for secondary research question five.....	50

Impact Evaluation of PREP's Within Our Reach In Orlando, Florida

I. INTRODUCTION

A. Introduction and study overview

Relationship satisfaction for couples is associated with increased life expectancy, mental/physical health, financial security, and overall quality of life (Waite & Gallagher, 2000). Yet, low-income couples experience more chronic stressors and are at greater risk for lower quality relationships compared to middle-class couples (Karney & Bradbury, 2005; Neff & Karney, 2004). Given that low-income couples are at greater risk for single parenthood, family instability, tend to have lower levels of family support, less education, and are less likely than middle-class individuals to get married and/or stay married, relationship education (RE) programs must address protective factors to alleviate this heightened level of stress (Johnson et al., 2002). Therefore, although all couples may benefit from learning relationship skills, RE programs may produce greater effects for couples who are economically strained.

Researchers have found financial discord within a couple to be a strong predictor of relationship quality (Hardie & Lucas 2010; Schramm & William Harris, 2011) as couples experience hardships and uncertainty with financial instability. Married couples are at greater risk for marital dissolution when reporting economic distress; thus, it is not surprising that unmarried low-income couples find that the biggest obstacle to maintaining a long-term relationship is financial security (Charles, Orthner, Jones, & Mancini, 2006).

In addition to financial stress, research findings identify a strong correlation between relationship discord within a couple and poor parenting (Bodenmann & Cina, 2005; Clavarino et al., 2011; Hair et al., 2011). Parental relationship quality has been linked to (a) poor adult relationships for children later in life, (b) child behavior problems, (c) child social competence, (d) child school engagement, and (e) childhood depression (Amato & Booth, 1997; Moore, Kinghorn, & Bandy, 2011). Therefore, strengthening parents' relationships may support more positive parenting and positive child outcomes. Thus, RE programs that decrease marital strain may be expected to have positive effects on parenting quality, and subsequently, children (Carlson, Barden, Daire, & Swartz, 2014; Barden et al., 2015).

To help couples navigate the relational strain that low-income couples experience, Project Harmony (located at the University of Central Florida in Orlando, FL) was designed to implement and test the *Within Our Reach* (WOR) relationship education curriculum, which was developed by the Prevention and Relationship Education Program (PREP). PREP is a well-known relationship education curriculum and has been implemented throughout the United States (US) and internationally. WOR is an adapted version of the behaviorally-based PREP curriculum, and is designed specifically for low-income couples. WOR incorporates behavioral principles that

include teaching couples about effective communication strategies and healthy conflict resolution. WOR is a unique RE program in that it includes an emphasis on intentional decision-making, as well as establishing and maintaining relationship safety. As couples learn to implement the relationship skills taught in WOR, they may be able to better regulate their emotions, experience improved stress communication, and improve their overall relationship quality. Moreover, couples may demonstrate better overall parenting and improved family relationships.

Although considerable research supports the positive effects of RE on relationship satisfaction and communication skills, the majority of existing research is based on pre/post or quasi-experimental designs. Furthermore, gaps in the empirical bases for RE remain regarding programmatic moderators that influence change (see Stanley et al., 2020). Consequently, there has been less emphasis on understanding which program elements contribute to positive change within couple relationships. To this end, this impact evaluation uses a rigorous randomized control trial (RCT) design that aims to understand the overall effects of WOR on low-income couples in the Central Florida community.

The remainder of this report is divided into several sections, and is organized and presented as follows: Section I.B and I.C outline the primary and secondary research questions; section II provides a description of the intervention and counterfactual conditions; section III presents an overview of the study design; section IV describes the analysis and methods; section V discusses the findings and estimation approach; section VI discusses the findings; section VII provides the references cited; finally, section VIII presents appendices, including the project logic model, information on the data and study sample, CONSORT diagram, data preparation procedures, and details on the impact estimation methods.

Research questions:

This section presents the primary and secondary research questions that were assessed in the impact evaluation. The proposed research questions were developed to assess goals as stated in the project logic model (see Appendix A). The project goals included improving overall relationship quality for the low-income couples enrolled into the project. Specifically, we aimed to (a) increase relationship satisfaction; (b) increase parenting quality; (c) increase couples' dyadic coping; and (d) improve communication skill-building. Although we did not measure communication skill-building directly, we operated on the premise that by teaching communication skills to couples, we would be able to demonstrate positive effects outlined in goals a-c because enhanced communication is one likely mechanism through which these outcomes would be impacted.

B. Primary research question(s): Primary outcome measures were analyzed for couples at the six-month follow-up period.

1. What is the effect of relationship education on couples' perceptions of overall relationship satisfaction?
2. What is the effect of relationship education on couples' perceptions of their dyadic coping?
3. What is the effect of relationship education on couples' perceptions of their ability to regulate emotions?
4. What is the effect of relationship education on couples' perceptions of their overall individual distress levels?

C. Secondary research question(s): Secondary research questions addressed outcomes for sub-populations (e.g., parents) within the primary sample. Additionally, secondary questions (2-4) included examination of the rate of change, or the growth curve, for major intervention target areas over the six-month study period. Thus, the growth curve models estimated the rate of change over three measurement points after random assignment (one month, three months, and six month follow-up).

1. What is the effect of relationship education on couples' perceptions of parenting and family adjustment six months after random assignment for participating couples who have children living at home under the age of 18?
2. How does relationship education affect the rate of change over the six-month study period in couples' perceptions of relationship satisfaction?
3. How does relationship education affect the rate of change over the six-month study period in couples' perceptions of dyadic coping?
4. How does relationship education affect the rate of change over the six-month study period in couples' perceptions of emotion regulation?
5. How does relationship education affect the rate of change over the six-month study period in couples' perceptions of individual distress?

II. INTERVENTION AND COUNTERFACTUAL CONDITIONS

This section describes the intervention as intended for delivery, as well as the counterfactual condition. We outline the program services received by all couples enrolled, as well as those received specifically by couples assigned to the treatment condition versus those assigned to the wait-list control condition.

A. Description of program as intended

The intervention tested in Project Harmony is 12-hours of relationship education utilizing the Prevention and Relationship Education Program's (PREP) *Within Our Reach* (WOR) curriculum. The relationship education workshops were delivered to intervention couples one night per week, three-hours each night, for a total of four week-nights over a 30-day period. Intervention couples began the relationship education workshops no more than two weeks after study enrollment and random-assignment. WOR teaches couples about effective communication, and healthy conflict resolution skills, while maintaining a focus on intentional decisions and relationship safety. Couples also discuss relationship commitment, and topics such as ways to have fun together. A description of the curriculum content can be found in Table II.1.

The relationship education workshops took place on the main campus of the University of Central Florida in east Orlando, Florida. The workshops were co-facilitated by educators trained in the *Within Our Reach* curriculum. The educators were either family case managers employed by the University through project funding, or were contracted relationship educators. Educators delivered the curriculum using a combination of didactic presentation with power point slides, videos, and group discussion. All educators participated in regularly scheduled supervision and fidelity meetings. In addition to the 12-hour relationship education curriculum, couples received case-management services, and two booster workshops (one at 90-days post enrollment and one at 180-days post enrollment) on career pathways and planning. The booster workshops were facilitated by the family case management staff and by an outside facilitator from non-profit organizations (Career Source of Central Florida, and Debt Management Credit Counseling Corp) with expertise in the career development and financial literacy. See Table II.1 for content description of the career pathways program (i.e., economic stability workshops).

Education and training of staff facilitating the educational workshops included a minimum of eight hours of training on the *Within Our Reach* curriculum, in addition to a minimum of 12 hours of shadowing a live-workshop prior to teaching. Staff whose roles included facilitation of educational workshops were required to attend an annual four-hour refresher training on the *Within Our Reach* curriculum. In addition, all Relationship Educators attended a monthly supervision meeting, led by the Assistant Director of Family Services, to review curriculum content and address content fidelity. Education facilitators also conducted peer-review observations whereby they completed a formal checklist to help monitor curriculum fidelity.

Table II.1. Description of intended intervention and counterfactual components and target populations

Component	Curriculum and content	Dosage and schedule	Delivery	Target Population
Intervention				
Relationship skills workshops	Healthy relationships curriculum: Fun, Communication Danger Signs, Speaker-Listener Technique, Stress and Relaxation, Issues and Events, Love Styles Personality and Expectations, Commitment	12 hours, with 3-hour sessions occurring once per week for four weeks	Group lessons provided at the intervention's facilities by two trained facilitators in every session	Low-income couples
Economic stability workshops	My Story, My Career: Exploration of vocational history, identifying transferable skills, determining future goals, and learning about job-search resources; Taking Control of Your Personal Finances: Learning about Budgeting,	two 90-minute workshops at 3-month and 6-months post enrollment (2 and 4 months after the relationship education workshops)	Career Development workshops are provided by two facilitators. Financial Literacy workshops are provided by one facilitator.	Low-income couples
Counterfactual				
Economic stability workshops	My Story, My Career: Exploration of vocational history, identifying transferable skills, determining future goals, and learning about job-search resources Taking Control of Your Personal Finances: Learning about Budgeting,	two 90-minute workshops at 1-month and 3-months post enrollment	Career Development workshops are provided by two facilitators. Financial Literacy workshops are provided by one facilitator. Administered only after the six-month follow-up surveys.	Low-income couples

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

Component	Education and initial training of staff	Ongoing training of staff
Intervention		
Relationship skills workshops	<u>Facilitators are male and female, hold at least an Associate's Degree and received a minimum of one day of initial training, in addition to a minimum of 12 hours of shadowing prior to teaching.</u>	Facilitators <u>receive a half-day of annual refresher training</u> in the intervention's curricula from study staff. Facilitators also attend a monthly supervision meeting to address content fidelity.
Economic stability workshops	<u>Facilitators are male and female, hold at least an Associate's Degree and received a minimum of one day of initial training, in addition to a minimum of 12 hours of shadowing prior to teaching.</u>	Facilitators <u>receive a half-day of annual refresher training</u> in the intervention's curricula from study staff.
Counterfactual		
Economic stability workshops	<u>Facilitators are male and female, hold at least an Associate's Degree and received a minimum of one day of initial training, in addition to a minimum of 12 hours of shadowing prior to teaching.</u>	Facilitators <u>receive a half-day of annual refresher training</u> in the intervention's curricula from study staff.

B. Description of counterfactual condition as intended

The study included a wait-list control group who were eligible to receive the WOR curriculum after the final data collection was completed 180 days post-random assignment. To keep wait-list control group couples engaged and to encourage participation in the study, couples were invited to participate in a booster workshop on the career pathways program at both 30 days and 90 days post-random assignment. As such, the wait-list control group received no services for the first 30-days post random assignment. Aside from the booster workshops, family case coordinators provided reminder contacts with wait-list control couples, and some referrals to community resources in extenuating circumstances. Upon initial phone contact with enrolled participants, Family Case Managers (FCMs) inquired about any current unmet needs the couple may be facing. Some of these needs may have been related to finding employment, securing childcare, difficulties with housing or finances, finding additional mental health services, etc.. Based on the participant's self-report, FCMs provided contact information for community organizations that could help address the stated needs. FCMs followed-up on the participant's use of these referral sources in all future communications with the participant. Additionally, the FCMs inquired about potential needs throughout the participant's enrollment in the program.

The intent of providing the booster workshops to wait-list control couples was to promote retention and collect follow-up outcome data. No *Within Our Reach* content was included in these booster workshops sessions for any couples. The content strictly focused on economic stability material.

Six-months after random assignment and after data collection ended, wait-list control group couples were provided an abbreviated eight-hour relationship education workshop utilizing the WOR curriculum. The workshop took place at a hotel, where couples spent the night on a Friday evening and attended the workshop on Saturday. All study participants received a resource guide that included a list of community resources that may be appropriate for study participants. Additionally, there were services provided in the local community that participants could access on their own. The project team did not formally collect information on whether control group members accessed these other services available in the community. A description of the types of services available in the community is provided below. None of the available services included relationship education, although some may have included more traditional couples' therapy. See Table II.1 for a description of the counterfactual groups.

Each participant (both in the treatment and wait-list control groups) was provided with a resource sheet at intake that included resources for childcare, education, job and career, and various other needs. Each Family Services team member also asked participants about needed resources during their initial phone call and provided relevant resources as requested. The initial phone call occurred after couples were enrolled in the study. The call provided an opportunity to build rapport with the couple, and to determine if they had any immediate resource needs that might preclude study participation. There were a number of community agencies that participants may have accessed for resources in the Central Florida area. Participants' needs varied and they may have taken advantage of many outside resources. Some of those available in the surrounding area

include those aimed at assisting residents with health and medical needs as is the goal of the Orange and Seminole County health departments, and Primary Care Access Network. Some resources assist with housing, such as Housing and Neighborhood Development Services of Central Florida and the Orlando Housing Authority. Specific to career and job advancement, local agencies that participants may have encountered are Career Source Central Florida, Christian Help, and Goodwill Job Connection. Mental Health agencies in the area include UCF Community Counseling and Research Center and Aspire Health Partners. There are also a number of agencies available in the area that provide crisis support to participants such as the Second Harvest Food Bank of Central Florida, Jewish Family Services of Orlando, and the Low Income Home Energy Assistance Program. Program staff did not track which referrals participants may have received assistance from while enrolled in the study.

C. Research Questions about the intervention and counterfactual conditions as implemented

This section presents the research questions that we examined in order to understand the implementation of PREP's *Within Our Reach*.

For the implementation analysis, we focused on the quality of the services provided, recruitment, engagement, and retention. We emphasized these areas due to retention challenges experienced during the six-month study period. The primary aim of the implementation analysis was to understand participants' experiences with the program, from recruitment to study completion. Thus, we examined the following overarching questions:

Recruitment: How were recruitment practices implemented?

Quality: Were participants satisfied with the services they received?

Engagement/Retention: What were contributing factors to participants' decision to enroll and complete the study?

Table II.3 summarizes each implementation element, and the research questions associated with each element. The implementation elements, and research questions, are presented for both the intervention and counterfactual groups. However, they are the same for both groups.

Table II.3. Examples of research questions for each implementation element and study group

Implementation element	Research question
Intervention Group Questions	
Recruitment	<ul style="list-style-type: none"> • What were the primary recruitment sources for enrolled participants? • Additionally, were recruitment strategies (i.e., passive versus active) related to study completion?
Quality	<ul style="list-style-type: none"> • How satisfied were participants with the overall program model? • What relationship did participants have with program staff, including relationship educators?

Implementation element	Research question
Engagement/Retention	<ul style="list-style-type: none"> • What were the primary motivators for enrolling in the program? • What were the main reasons for missing scheduled workshops? • Did participants plan to continue participating in the study when asked about their future intentions to return to their next scheduled workshop? • What proportion of participants actually continued to participate in the study when they indicated that they would?
Control / Comparison Group Questions	
Recruitment	<ul style="list-style-type: none"> • What were the primary recruitment sources for enrolled participants? • Additionally, were recruitment strategies (i.e., passive versus active) related to study completion?
Quality	<ul style="list-style-type: none"> • How satisfied were participants with the overall program model? • What relationship did participants have with program staff?
Engagement/Retention	<ul style="list-style-type: none"> • What were the primary motivators for enrolling in the program? • What were the main reasons for missing scheduled workshops? • How did participants respond to their intention to return to their next scheduled workshop? • What proportion of participants actually continued to participate in the study when they indicated that they would?

III. STUDY DESIGN

The current impact evaluation tested the *Within Our Reach* curriculum using an RCT design that included a six-month wait-list control group. All study participants were recruited from the greater Orlando, Florida area. In this section, we describe the primary elements of the study design, including sample formation, and data collection for both the implementation and impact analyses.

A. Sample formation and research design

The target population comprised low-income couples over the age of 18. Couples could be married or not-married, and may or may not have had children. In order to be considered for the impact evaluation, participants must have enrolled together as a couple, and planned to attend the WOR classes and booster sessions together. In the event that one partner chose not to participate after enrollment, the other partner was still invited to participate. The data for couples that did not attend the program classes was still included in the analysis.

Project Harmony employed a team of recruiters who regularly visited local county health departments, libraries, women infants and children (WIC) waiting rooms, back to school and other community events (i.e., active strategies) that traditionally targeted ethnically diverse and low-income participants. Project staff did not verify participant income, nor was income used as an exclusion criteria. Thus, we use the term ‘low-income’ to refer to a population who qualified for income-based services, such as the federal WIC program. When engaged in such active recruitment activities, the recruitment team very briefly presented Project Harmony services (including the impact evaluation of WOR) through a scripted pitch, then collected basic information from interested participants. The recruitment team then followed-up by phone with each participant to provide additional project information, and scheduled the enrollment appointment.

In addition to active recruitment strategies, many couples learned about Project Harmony from previous participants, the project website and other social media websites (i.e. Facebook and Instagram), or fliers (i.e., passive strategies) posted throughout the community. When prospective participants called to find out more information about Project Harmony services, the recruitment team conducted a brief eligibility screening over the phone to determine which service’s inclusion criteria fit for the couple. For example, if the couple was over the age of 18, and planned to participate together, they were then referred to the WOR impact study and scheduled for an enrollment appointment.

After providing consent during the study enrollment appointment, couples were randomly assigned in real-time to either program A (the intervention group) or program B (the wait-list control group). The enrollment appointment was conducted in a group with other couples representing a cohort. Cohorts were formed once per month over two group intake meetings, with approximately 20-30 couples per cohort. Cohorts were tracked throughout the study so that appropriate timing for data collection could be implemented. Childcare for children under 12 was provided for those who needed it, as well as a hot meal as the meeting was held in the evenings.

The intake/enrollment meeting was facilitated by family case managers, and the data coordinator. Each couple was checked-in, provided with a name tag, then began eating their dinner. While eating and after checking-in all couples, the meeting facilitators began with an ice-breaker activity. They then provided a brief overview of the six-month study, including the random assignment process. Next, all participants were provided with the informed consent form and after consenting, were given an Apple iPad and began by completing their nFORM Entrance Survey and Applicant Characteristics Survey. These surveys collected descriptive information from couples, such as their age, gender, race, ethnicity, as well as information about their relationship, and the data was entered into a software program that provided real-time information to the federal funding agency for reporting purposes. After completing the nFORM questions, participants next completed outcome evaluation questions on an app-based supplemental management information system (SMIS) developed by the research team. Once participants hit the 'submit' button on the final question, the system automatically assigned the couple (not individuals) to program A or B. The system utilized a unique couple ID to conduct the assignment and each member of the couple was notified of their group assignment via a notification on the iPad. The enrollment team documented the assignment and then notified the couple of their next scheduled visit (e.g., the following week or two weeks on the same weeknight for treatment couples; or 30-days later for wait-list control couples to attend the first booster workshop). The SMIS randomized couples at a 1:1 ratio of treatment to wait-list control, which ensured all couples had an equal assignment probability to the treatment or wait-list control group. The use of electronic and web-based data platforms allowed for real-time randomization, as well as real-time data access by the evaluation team.

The University of Central Florida's institutional review board (IRB) approved the study (IRB number: SBE-15-11828). All study processes were approved prior to the beginning of the study and all data collection activities. Any federally-approved changes to the study, or marketing material, were also approved by the UCF IRB.

B. Data collection

There were two primary data-sources for the WOR impact evaluation. The first was the nFORM platform, which included the pre-, and post-surveys, as well as baseline characteristics of study participants, and program-related information such as attendance records. The second data-source was the supplemental management information system (SMIS), which was a web and app-based platform used to collect outcome measure responses from participants. Of note, because some couples may have had children and others may not, the sample was sub-divided based on child status. Those with children completed the parenting measures, while those without children did not. The SMIS administered the correct batch of assessments according to the enrolled couples' child status.

Data was collected at four time-points across the study: 1) time 1 - initial enrollment/random assignment; 2) time 2 - 30-days post-random assignment; 3) time 3 - 90-days post random assignment; and 4) time 4 - 180-days post random assignment. As noted above, couples were enrolled in a group workshop format, or cohort. Each cohort represented couples in both the

intervention and wait-list control groups, and they were assessed at equivalent time points throughout the duration of the study.

1. Implementation analysis

This section provides a description of the data sources used to address the implementation study research questions.

Data sources for the research questions in Table II.3 stem from three primary sources: (a) the nFORM applicant characteristics form; (b) the Project Harmony participant satisfaction survey; and (c) the researcher-developed supplemental information management system (SMIS). Participants were administered all questions at either enrollment/random assignment (data for everyone); 30-day follow-up; or 180-day follow-up. All data was collected electronically and in either English or Spanish. See table B.1 in the appendix for a description of data sources for the implementation analysis

2. Impact analysis

The data collection process described above applies for both the implementation and impact data analysis. For example, the four time points remained the same, and the primary data sources also remained the same. For specific constructs in the impact analysis, and their respective data sources, please see Table B.2 in the appendix.

IV. ANALYSIS METHODS

In this section, we describe the enrolled sample, as well as how the analysis sample was formed. We also provide a description of the outcome measures, and baseline equivalence between the treatment and wait-list control groups.

A. Analytic sample

Couples were initially screened during recruitment ($n = 7,093$), and then assessed again during a follow-up phone call. Couples who were eligible and still interested in study participation after the second follow-up were then scheduled for their initial enrollment and random assignment meeting. The CONSORT diagram in Appendix C summarizes the number of couples who were either screened out for eligibility reasons, or who declined to participate ($n = 5,675$). The diagram also depicts both overall and differential attrition of study participants at each assessment time point. Overall attrition varied by outcome, but ranged from 67% to 41%, and differential attrition ranged from 12% to 15%. Couples assigned to the wait-list control group demonstrated higher retention than treatment couples at all time points. See Appendix C for the CONSORT diagram that depicts the flow of participants through the study.

The primary analysis sample ($n = 1,355$ individuals) represents the individuals who completed the 180-day (i.e., six month) follow-up. The data from individuals who completed the 180-day follow-up was utilized when analyzing the primary research questions and examining treatment effects six-months after random assignment. However, we also estimated longitudinal growth curve models to address the secondary research questions. The same sample was used for both the primary and secondary research questions. We did not impute any outcome data in the computation of the growth curve models. However, for the secondary research question pertaining to parenting, we only used the subset of participants who had children under 18 and completed the 180-day follow-up ($n = 913$). We also report the total number of individuals as opposed to couples because at times, one member of the couple did not complete follow-up assessments. This occurred at 30-day post-enrollment ($n = 29$ couples), 90-day follow-up ($n = 33$ couples), and 180-day follow-up ($n = 33$ couples). For each of these identified couples, one person in the couple completed the outcome measures at the respective time point and one did not. This occurred because the partner did not attend that particular event. However, we still attempted to collect data from the person who did not attend, and that person was invited to subsequent events per the Intent-To-Treat (ITT) framework that we used for this study.

The analytic samples were based on the specific outcome measures being analyzed, as well as attrition at each time point (i.e., based on available data). There are two separate samples that we analyzed. The primary sample was the total sample, or everyone with available data, regardless of child status, who responded to relationship and individual distress constructs (relationship satisfaction, dyadic coping, emotion regulation, individual distress). These couples reflect the 2,836 individuals randomly assigned ($n = 1,418$ couples), and the 1,355 individuals who completed the six month follow-up survey at the end of data collection in April 2020.

The sample used for the secondary analyses included only couples who had children under the age of 18. Having children was not a study inclusion criterion for enrollment. As a result, the sample size for research questions that address parenting outcomes is a sub-sample of the larger primary analysis sample. There were 2,013 individuals who reported having children at the time of enrollment. Of those, 913 completed the 180-day follow-up survey. Therefore, the analysis sample for the secondary research questions that assess parenting outcomes is 913.

Program staff tracked and recorded attendance at every relationship education workshop. Only couples assigned to the treatment group were included on the attendance roster. Any couple not on the attendance roster was not allowed to participate in the relationship education workshop. However, crossover of control group members to the primary intervention occurred once during the study. Specifically, one wait-list couple was attending a make-up meeting with their family case manager to address content they missed during a 30-day booster session. The family case manager confused this couple with a treatment couple and began providing them with content from the intervention. After 20 minutes, the family case manager realized the mistake, immediately stopped, and notified the evaluation team. We retained this couple's data for the impact analysis in accordance with the ITT framework. There were no other instances of crossover.

Family case managers documented all referrals provided to both treatment and wait-list control participants. However, there was no tracking mechanism to determine if participants actually followed through with the referral. Additionally, there was no mechanism to track participants who sought services on their own, outside the assistance of program staff. The randomized component of the RCT design should account for any potential confounding variable related to motivation to seek additional assistance, as participants in both the treatment and wait-list control had access to family case management as well as the same resources in the Orlando community.

Table IV.1a. Individual sample sizes by intervention status

Number of individuals	Intervention sample size	Comparison sample size	Total sample size	Total response rate	Intervention response rate	Comparison response rate
Assigned to condition	1,418	1,418	2,836	n.a.	n.a.	n.a.
Contributed a baseline survey	1,418	1,418	2,836	100%	100%	100%
Contributed to immediate post survey (30-days)	956	1,130	2,086	74%	67%	80%
Contributed to first follow-up survey (90-days)	656	887	1,543	54%	46%	63%
Contributed to second follow-up survey (180-days)	578	777	1,355	48%	41%	55%
Contributed to second follow-up (Outcome: Relationship)	578	775	1,353	48%	41%	55%
Contributed to second follow-up (Outcome: Dyadic Adjustment)	576	775	1,353	48%	41%	55%
Contributed to second follow-up (Outcome: Emotion Regulation)	576	777	1,353	48%	41%	55%

Number of individuals	Intervention sample size	Comparison sample size	Total sample size	Total response rate	Intervention response rate	Comparison response rate
Contributed to second follow-up (Outcome: Individual Distress)	570	775	1,345	47%	40%	55%

Note: The variance in samples sizes by outcome at the second follow-up is a result of incomplete data due to item non-response. We did not impute any values, and only analyzed available data.

B. Outcome measures

The primary outcome measures for the WOR impact evaluation were chosen based on their alignment with the primary constructs identified in the logic model. The measures are listed below in Table IV.2, and include (a) Revised Dyadic Adjustment Scale (Busby, Christensen, Crane, & Larson, 1995); (b) Dyadic Coping Inventory (Ledermann et al., 2010); (c) Difficulties in Emotion Regulation Scale (Gratz & Roemer, 2004); (d) Parenting and Family Adjustment Scale (Sanders, Morawska, Haslam, Filus, & Fletcher, 2014); and (e) the Outcomes Questionnaire 45.2 (Lambert et al., 2004).

Given concerns about the reduction in statistical power resulting from attrition, we only examined total scores for each of the measures, as opposed to using several sub-scale scores from each of the measures. Using total scores is supported by previous factor analyses and does not have any adverse implications for the psychometric properties of the measures. We calculated the reliability of each measure using responses from individuals who participated in this study. Each of the chosen assessments has a demonstrated history of sound psychometric properties and are provided in the table below.

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

Outcome measure	Description of the outcome measure (Example questions in italics)	Source	Timing of measure
Relationship Satisfaction	<i>The Revised Dyadic Adjustment Scale (RDAS) is a 14-item measure of overall satisfaction, and includes three sub-scales: dyadic consensus; dyadic satisfaction; and dyadic cohesion. The RDAS is a self-report measure and can therefore be interpreted when only one member of the couple completes the measure. Response options range from 0 (Always Disagree) to 5 (Always Agree). Scores are summed to calculate the total score, with higher scores indicating better adjustment (satisfaction). A sample question prompt for the first six items is as follows: Most persons have disagreements in their relationships. Please indicate below the approximate extent of agreement or disagreement between you and your partner for each item on the follow list: Religious matters; Demonstration of affection; Making major decision; Sex relations; Conventionality (correct or proper behavior); and Career Decisions. Construct validity is .97 with the original Dyadic Adjustment Scale, and .66 with the Marital Adjustment Test. Alpha reliability is .90, Guttman split-half reliability is .94, and Spearman-Brown split half reliability is .05 (Busby, Christensen, Crane, & Larson, 1995). Alpha reliabilities for each time point in the current evaluation are as follows: .91 – enrollment and .91 – 180-day follow-up.</i>	Local survey	Baseline, 30-, 90-, and 180-day follow-up

Outcome measure	Description of the outcome measure (Example questions in italics)	Source	Timing of measure
Stress and coping	<p><i>The Dyadic Coping Inventory (DCI) is a 37-item measure of relationship quality that emphasizes stress and coping.</i> The DCI assesses Self (15 items) and Partner's (15 items) coping behavior, common dyadic coping (5 items), overall satisfaction with coping (2 items), and total dyadic coping. The total dyadic coping is calculated by summing scores on all items. Higher scores indicate better dyadic coping. Responses range on a 5-point scale from <i>never/very rarely</i> to <i>very often</i>. Example items include stress communication by self - <i>I let my partner know that I appreciate his/her practical support, advice, or help</i>, or partner - <i>My partner lets me know that he/she appreciates my practical support, advice, or help; I show empathy and understanding to my partner</i>; two types of stress dyadic coping, emotion focused by self and partner – <i>I (my partner) show empathy and understanding</i>, and problem-focused by self – <i>I tell my partner that his/her stress is not that bad and help him/her to see the situation in a different light</i>, or partner – <i>My partner helps me to see stressful situations in a different light</i>; delegated dyadic coping by self – <i>I take on things that my partner would normally do in order to help him/her out</i>, or partner – <i>My partner takes on things that I normally do in order to help me out</i>; and negative dyadic coping by self – <i>I blame my partner for not coping well enough with stress</i>, or partner – <i>My partner blames me for not coping well enough with stress</i>. The scales can be combined for an aggregate measure of dyadic coping by self and partner. Alpha reliability for total dyadic coping was .95 for men and .94 for women in a recent study of US English-speaking couples (Randall, Hilpert, Jimenez-Arista, Walsh, & Bodenmann, 2016), and .94 for men and women Spanish-speaking US couples (Falconier, Nussbeck, & Bodenmann, 2013). The DCI subscales and total scores correlated with relationship satisfaction showing convergent validity, and were less correlated with active coping, showing discriminant validity (Randall et al., 2016). Alpha reliabilities in the current evaluation at each time point are .94-enrollment; and .95 at 180-day follow-up.</p>	Local survey	Baseline, 30-, 90-, and 180-day follow-up
Emotion Regulation	<p><i>The Difficulties in Emotion Regulation Scale (DERS) is a 36-item measure of emotional awareness, acceptance, goal-directed behavior, and emotion regulation strategies.</i> Respondents indicate the frequency with which items apply. Responses range from <i>almost never</i> (1) to <i>almost always</i> (5). Total scores are calculated by summing scores from all items. Higher scores indicate greater difficulty with regulating emotion. Example items include: <i>I am clear about my feelings; I pay attention to how I feel; I have difficulty making sense out of my feelings</i>; or, <i>When I'm upset, I become out of control</i>. The DERS demonstrates high internal consistency (.93), construct validity with the Negative Mood Regulation Scale, and test-retest reliability (.88) (Gratz & Roemer, 2004). Alpha reliabilities for the current evaluation are .94 at enrollment and .94 at 180-day follow-up.</p>	Local survey	Baseline, 30-, 90-, and 180-day follow-up

Outcome measure	Description of the outcome measure (Example questions in italics)	Source	Timing of measure
Individual Distress	<i>The Outcomes Questionnaire (OQ) 45.2 is a 45-item measure of individual psychological distress symptoms, such as anxiety and depression. The OQ includes three sub-scales (symptom distress, interpersonal relationships, and social role), as well as a total distress score. Responses range from never to almost always. The total score is calculated by summing the scores from all responses. Higher scores indicated greater overall distress. Example items include: I get along well with others; I tire quickly; I feel no interest in things; I feel fearful; I am a happy person; or I feel my love relationships are full and complete. The OQ demonstrates high internal consistency (.93), rest-retest reliability (.84), and concurrent validity with the Zun anxiety scale (.81) and the Social Adjustment Scale (.65) (Lambert et al., 2004). Alpha reliabilities in the current evaluation are .94 at enrollment and .94 at 180-day follow-up.</i>	Local survey	Baseline, 30-, 90-, and 180-day follow-up

All of the outcome measures used to test the primary research questions were also used in the secondary impact analysis research questions. The table below only reflects the additional measure added for the secondary impact analysis.

Table IV.3. Outcome measures used for secondary impact analyses research questions

Outcome measure	Description of outcome measure (Examples questions in italics)	Source	Timing of measure
Parenting and Family adjustment	<i>The Parenting and Family Adjustment Scale (PAFAS) is a 30-item measure that examines parenting practices, parent-child relationships, parental adjustment, family relationships, and parental teamwork. The PAFAS includes two overall scales: Parenting and Family Adjustment, with four parenting scales (parental consistency, coercive parenting, positive encouragement, parent-child relationship) and three family adjustment scales (parental adjustment, family relationships, and parental teamwork). Participants respond to items on a four-point scale ranging from Not at all to Very much. There is no total score for this measure. Scale scores are calculated by summing the total responses for respective scale items. Higher scores indicate more negative parenting or family adjustment. Example family adjustment items include: Parental adjustment – I feel stressed or worried; family relationships – Our family members get on well with each other; and parental teamwork – I work as a team with my partner in parenting. Reliability ranges from .70-.85 for parenting and .84-.87 for family adjustment (Sanders, Morawska, Haslam, Filus, & Fletcher, 2014). Concurrent and discriminant validity established with the Child Adjustment and Parental Efficacy Scale for PAFAS parenting and family adjustment (Sanders et al., 2014). Alpha reliabilities in the current evaluation for the scales we used are as follows: (a) family adjustment - .93 at enrollment and .92 at 180-day; (b) parenting relationships - .87 at enrollment and .86 at 180-day; and (c) parental teamwork - .82 at enrollment and .80 at 180-day.</i>	Local Survey	Baseline, 30, 90, and 180-day follow-up

C. Baseline equivalence and sample characteristics

We examined pre-program outcome measures and demographic characteristics for baseline equivalence between the intervention and wait-list comparison groups to determine if statistical adjustment was necessary. We assessed baseline equivalence using the pre intervention total scores for all outcome measures (RDAS, DCI, DERS, OQ, and PAFAS scales – Parental Adjustment, Family Relationship, and Parental Teamwork) and selected baseline characteristics (Age, child status, ethnicity, employment, educational attainment, and income). Effect sizes were calculated using Cohen’s *d* for continuous variables (outcome measures) and the phi coefficient for categorical variables (demographic data). We computed unadjusted means and standard deviations of the outcome measures at baseline to assess equivalence.

Equivalence of baseline demographic variables for participants who completed the 180-day follow up data collection ($N = 1,355$ individuals) for intervention and comparison groups was assessed using chi-square tests of independence. The null hypothesis assumes no association between treatment group (intervention or comparison) and the demographic variable of interest. Age was divided into two groups representing young adults (34 years of age or younger) and older adults (35 years of age or older). Child status was defined by whether the individual/couple had children or not. Ethnicity was defined as those who identified as Hispanic or Latino and those who did not. This distinction is salient given the large Latinx population in the Central Florida community. Race was classified into those who identified as white and those who identified as being in other, non-white racial groups. Education level was divided into those who had no post-secondary education, those who had some college or an Associate’s degree, and those who completed a Bachelor’s degree or higher. Employment status was divided into those who were not employed, those who were employed in part-time or variable work, and those who were employed full-time. Income level was divided into those who made \$1,000 or less, \$1,001 to \$3,000, and \$3,001 or more per month. None of the chi-square tests showed evidence of a statistically significant relationship between the treatment or control group and the demographic variable of interest. In addition, all phi coefficients were small. Thus, we concluded that there was no association between treatment condition and any of the demographic variables at baseline. No statistical adjustments were needed for demographic variables.

We also examined equivalence based on baseline measures of the outcome variables. Two sample independent t-tests were conducted between intervention and comparison groups for each outcome measure (RDAS, DCI, DERS, OQ, and PAFAS scales – Parental Adjustment, Family Relationship, and Parental Teamwork) at baseline. In addition, Cohen’s *d* was computed as an effect size measure. None of the t-tests were statistically significant, thus we concluded that there were no differences between intervention and comparison groups in the true mean on any of the outcome measures. The effect size measures were below 0.05 for RDAS, OQ, and the PAFAS Parental Adjustment scale. The effect size measure was above 0.05 for DCI (.06), DERS (.05), the PAFAS Family Relationship scale (.08), and the PAFAS Parental Teamwork scale (.08). Using the What Works Clearinghouse guidelines, we included statistical controls for these minor baseline differences in the analysis for the primary research questions.

Table IV.4. Summary statistics of key baseline measures and baseline equivalence across study groups, for individuals/couples completing the 180-day follow up data collection

Baseline measure	Intervention % or mean (standard deviation)	Comparison % or mean (standard deviation)	Intervention versus comparison χ^2 statistic or mean difference (p-value)	Effect size
Age group (%)			$\chi^2=0.0220$ (p=.8821)	$\phi = -.0040$
34 and younger	35.01%	34.62%		
35 and older	64.99%	65.38%		
Child Status (%)			$\chi^2=0.0291$ (p=.8645)	$\phi = -.0046$
No children	32.87%	32.43%		
Has children	67.13%	67.57%		
Ethnicity (%)			$\chi^2=0.0553$ (p=.8141)	$\phi = .0064$
Hispanic or Latino	61.57%	62.19%		
Not Hispanic or Latino	38.43%	37.81%		
Race (%)			$\chi^2=0.6431$ (p=.4226)	$\phi = -.0218$
White	51.30%	49.09%		
Other	48.70%	50.91%		
Education Level (%)			$\chi^2=2.4170$ (p=.2986)	$\phi = .0454$
No post-secondary education	24.75%	27.89%		
Some college or Associates degree	28.34%	29.53%		
Bachelor's degree or higher	46.91%	42.58%		
Employment Status (%)			$\chi^2=1.1403$ (p=.5654)	$\phi = .0292$
Not employed	31.22%	28.59%		
Employed part-time/variable	25.22%	26.76%		
Employed full-time	43.56%	44.65%		
Income Level (%)			$\chi^2=1.1447$ (p=.5642)	$\phi = .0293$
\$1000 or less per month	41.90%	44.81%		
\$1001 to \$3000 per month	40.85%	38.50%		
\$3001 or more per month	17.25%	16.69%		
RDAS	50.39 (11.96)	50.21 (11.36)	0.17 (p=.7802)	d = 0.0153
DCI	138.98 (22.58)	137.63 (22.41)	1.35 (p=.2735)	d = 0.0602
DERS	69.69 (21.12)	70.76 (21.28)	-1.07 (p=.3594)	d = 0.0504
OQ	40.00 (22.32)	41.08 (22.85)	-1.09 (p=.3813)	d = 0.0481
PAFAS parenting scales				
Parental Adjustment	5.69 (4.20)	5.50 (3.97)	-0.19 (p=.4812)	d = .0472
Family Relationship	4.66 (3.42)	4.38 (3.34)	0.28 (p=.2141)	d = .0833
Parental Teamwork	3.04 (2.72)	3.25 (2.69)	-0.22 (p=.2288)	d = .0807
Sample size	578	777	n.a.	n.a.

Notes: p-values are included in parentheses. Effect sizes for demographic variable associations with group membership are calculated by taking the square root of the chi-square statistic divided by the sample size. Effect sizes for baseline outcome measures are calculated by dividing the differences in means by the pooled standard deviation.

n.a. = not applicable.

V. FINDINGS AND ESTIMATION APPROACH

In this section, we present key findings from both the implementation and impact analyses. We begin by identifying key implementation findings. We separate the implementation findings into the major categories we examined. We also present a few figures to graphically present the descriptive results. Next, we present the key impact analysis findings. We first present findings for the primary research questions followed by the secondary research questions. Finally, we also present results of sensitivity analyses.

A. Implementation evaluation

Implementation analyses focused on three primary areas: study recruitment; quality of services delivered; and engagement/retention of study participants. These three areas provide an overall indication about participants' experiences with the study. We structure our presentation of key implementation findings for each primary area.

1. Key findings

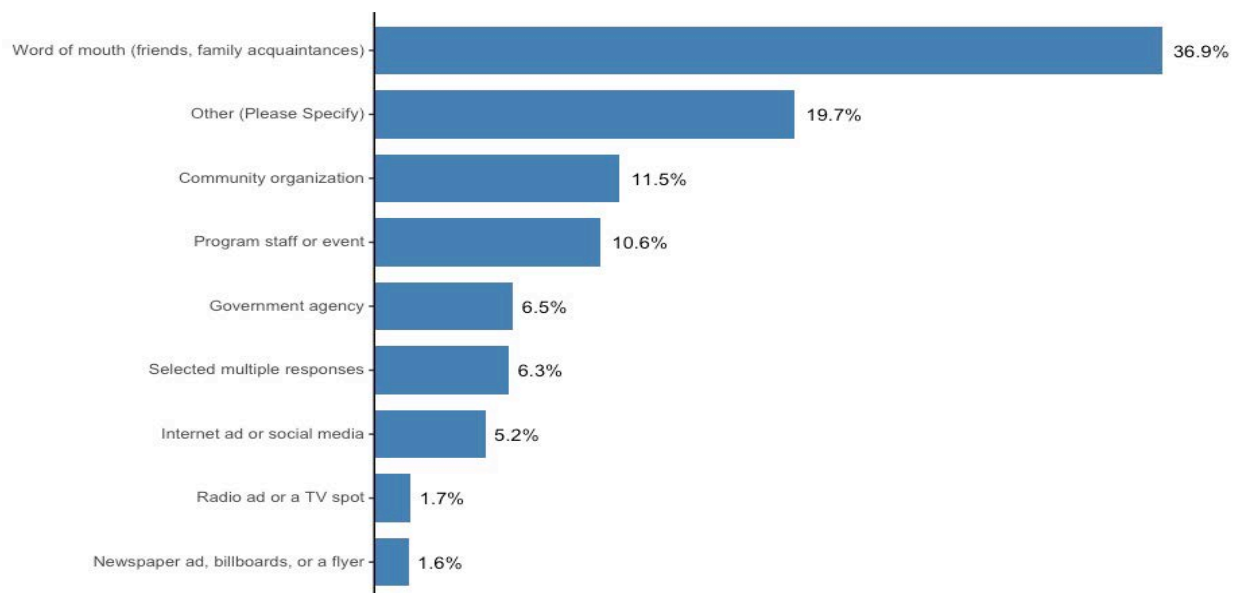
Participant recruitment serves as the first point of contact for participants with the study team. We implemented two primary recruitment strategies, active (that is, recruited by study team members at partnering organizations), and passive (that is, participants inquired on their own about the services available). Overall, the plurality of couples who enrolled were recruited via passive methods (45%). Among those couples recruited by passive methods, learning about the study through word of mouth (35%) was the most common passive strategy. We found no relationship between participants who were recruited actively or passively and the rate of overall study completion. Sixty percent of couples reported enrolling in the study because they wanted to learn about how to improve their personal relationships.

Quality implementation is a key factor in retaining couples. We found that 94% of couples were satisfied with the content of the relationship education that was offered. Most couples (99%) said they had a positive experience participating in the workshops; 98% reported having enough information to participate in the program (e.g., date/time of workshops, understanding how random assignment worked); 92% would recommend (or did recommend) the program to family or friends; and 98% thought the program was important or worthwhile. When considering relationship education workshops, 97% of couples were satisfied with the person who taught the workshop, and 91% were satisfied with the services provided by their family case manager. We also found that most couples felt a strong working alliance with the workshop facilitators. Most participants (95%) said they were extremely likely, or very likely, to return to their next scheduled visit (intent to attend) after their first visit to the center, and more than 80% of participants who said they were 'very likely' or 'extremely likely' to return to their next visit, did return.

Overall, couples enrolled because they heard about the study from other people, and wanted to learn strategies to improve their relationships. They were happy with the quality of the services delivered, and reported a strong likelihood of returning. Below, we expand on the findings in each implementation area.

Recruitment. Program recruitment staff used numerous methods to identify study participants. These methods ranged from advertisements, to attending community events, partnering with community organizations, or promoting word of mouth among current study participants. Through these methods, we met, or exceeded, yearly enrollment targets. We asked participants how they heard about the program, and graphed the distribution of their responses in Figure V.A.1. Of note is that the ‘Other’ category was the second most selected category. We examined specific responses from participants in this category and noticed that most, if not all, would have fit into one of the other pre-existing categories. For some reason, participants chose to identify them as ‘Other’. Examples of specified responses included: a Google search or Facebook (internet or social media category); learning about the project from wife, lawyer, or partner (word or mouth); and Goodwill, the library, or a job fair (community organizations).

Figure V.A.1. Percentage of responses to “How did you hear about the program”



Next, we categorized recruitment responses as either active (initiated by program staff) or passive (heard about the program on their own). We coded recruitment category as 1 = active and 0 = passive for each person. We conducted a logistic regression to determine if recruitment method was associated with study completion (1 = completed, 0 = did not complete). Results indicated no statistically significant associations among recruitment method and study completion at six-months.

Quality. We assessed the quality of services received by asking all study participants a series of program satisfaction questions at the completion of the study. The following table provides a distribution of participant responses by quality indicator.

Table V.A.1. “Quality” items percentages

	Strongly Disagree	Disagree	Undecided	Agree	Strongly Agree
I am satisfied with the content of the workshops.	0.6(7)	1.9(21)	3.7(40)	25.2(275)	68.6(749)
I am satisfied with the person(s) that taught my workshops.	0.2(2)	0.6(7)	2.6(28)	21.5(235)	75.1(820)
I am satisfied with the services provided by my Family Case Manager	1.0(11)	1.9(21)	5.6(61)	20.5(224)	70.9(774)

Note: Percentages are outside of brackets; sample sizes are presented within the brackets

We administered the PREP Alliance Measure (Owen, Antle, & Barbee, 2013) to treatment group participants after completing the four-week relationship education intervention. The following table presents the distribution of responses for each question. Overall, the average score was 6.1 out of a 7-point scale ($SD = 1.4$), indicating that participants were relatively satisfied with the relationship educators.

Table V.A.2. PREP Alliance percentages

	Strongly Disagree	2	3	Neither agree nor disagree	5	6	Strongly Agree
I believe the educators liked me.	1.5(14)	.7(6)	0.4(4)	12.1(110)	7.6(69)	21.3(193)	56.3(510)

Note: Percentages are outside of brackets; sample sizes are presented within brackets

Engagement/Retention. To understand program retention, we first assessed why people enrolled in the study. Sixty percent said they wanted to ‘improve personal relationships’, followed by 18% who said they attended because their ‘spouse/partner asked me to come’, and 11% said they wanted to ‘learn about being a better parent’. We asked about some other motivations for enrolling that were measured on a scale of one to six, with one indicating the most important reason and six indicating the least important reason. Nearly 65% indicated that the content/learning was the most important reason. See Table B.1 in the appendix for the questions asked to participants about engagement/retention.

The program services were delivered as intended, and retention in the relationship education intervention was high with 83% attending the first workshop, 74% attending the second workshop, 71% attending the third workshop, and 70% attending the fourth workshop. However, study retention for the second (90-day) and third (180-day) follow-up was low (54% and 47%, respectively). Thus, we asked people about some of the primary reasons for missing a workshop and the most frequent response was ‘the day of the workshops did not work with my schedule’ (30%).

We also assessed people’s intention to return to their next scheduled workshops. We asked people to respond to one question: “How likely is it that you will attend your next scheduled workshop?”. Participants responded on a scale of 1 (not at all likely) to 5 (extremely likely). We assessed responses at three time points (baseline, 30-days, and 90-days). The majority of

participants indicated being at least very likely to return each time they were asked: 95% at baseline; 95% at 30-day follow-up; and 96% at 90-day follow-up. We then examined the probability of responses of at least ‘very likely’ and attendance at the next assessment time point. Results indicated that 88% of people at baseline who indicated being at least ‘very likely’ to return attended the 30-day follow-up; 83% of people who responded to this question at the 30-day follow-up attended the 90-day; and 121% of people at the 90-day attended the 180-day follow-up, which means that more people attended than actually indicated they were at least ‘very likely’ to do so.

Overall, recruitment methods led to meeting or exceeding enrollment targets. Participants were satisfied with the services they received, and most indicated that they planned to return to their next visit. In fact, examining their intent-to-attend responses confirmed that many participants who said they would return, did come back. However, many couples experienced unexpected events that interfered with scheduled study events and contributed to missing sessions. Treatment couples were asked to attend the center more frequently than wait-list control couples, which likely contributed to higher attrition because they experienced more opportunities for unexpected events to interfere with scheduled attendance. However, we did not measure specific reasons for couples’ missed sessions. Another limitation to the implementation analysis is that the satisfaction surveys and the alliance questionnaire were completed by couples who attended, which might skew the results because we do not know how couples who did not attend would have responded. For example, they may have been less satisfied, which might have contributed to their attrition.

B. Primary impact evaluation

1. Key findings

To address the primary research questions and secondary research question 1, we estimated multilevel models that accounted for the dyadic structure of the data using the outcome measure of interest at the 180-day follow up period as the dependent variable. We used SAS PROC MIXED with maximum likelihood estimation for the multilevel analysis (Kenny, Kashy, & Cook, 2006). A treatment group indicator was included as an independent variable to determine if differences existed between the intervention and comparison groups. For secondary research question 1, three models were estimated using each of the three PAFAS scales as the dependent variable: Parental Adjustment, Family Relationship, and Parental Teamwork. The outcome measure score at baseline was included as a covariate to control for a small baseline differences between intervention and comparison groups for the analyses for primary research questions 2 and 3 (DCI and DERS as outcomes, respectively) and for secondary research question 1 for the Family Relationship and Parental Teamwork outcomes on PAFAS.

The sample for the primary research questions on RDAS, DCI, DERS, and OQ included 1,355 individuals. The sample for secondary research question 1 on PAFAS included 909 individuals.

Multilevel growth models were used to address secondary research questions 2-5 using the outcome measure of interest as the dependent variable. SAS PROC MIXED with maximum likelihood estimation was used for the multilevel growth analysis (Singer, 1998; Kashy et al., 2008). The means for each measure at each time point were calculated and plotted for both the treatment and control groups.

Observed patterns revealed initial changes from baseline to the 1-month post-assessment, with a leveling off for the 3-month and 6-month follow up assessments. Thus, a curvilinear growth model that included both a linear and a quadratic term to account for change in growth rate over time was selected. Examination of the difference in fit statistics, $-2 \log$ likelihood, between linear and quadratic models indicated that inclusion of the quadratic growth term significantly improved model fit for all outcome measures. Further, the linear growth parameters were statistically significant in all models and the quadratic growth parameters were statistically significant in all models except for RDAS. This indicates that the outcome measures changed from baseline and the rate of growth tended to slow over time. Time was coded to reflect elapsed time relative to the baseline measure. Specifically, time was coded as 0, 1, 3, and 6 representing data collection at baseline (pre-program), at the end of the initial program (1-month elapsed from baseline), 3-month follow-up from baseline, and 6-month follow-up from baseline, respectively. The intercept is thus interpreted as the baseline and the spacing accounts for the number of months from the baseline data collection. The linear growth parameter represents the initial growth rate for the comparison group and the quadratic growth parameter represents the change in growth rate for the comparison group.

An indicator variable for group membership (0=comparison, 1=intervention group) and an interaction term between group membership and time were also included in the models. These provide estimates of the differential effect of receiving the treatment for the baseline mean and growth rates, respectively. Separate models using the relevant outcome measure as the dependent variable were estimated for the four research questions. The analytic samples included only those who had complete data and completed the six-month follow-up survey. Thus, no imputation methods were utilized in the models¹. The sample for RDAS, DCI, DERS, and OQ included 1,355 individuals who had all completed baseline and 6-month follow-up data collection, 1,241 completed the 1-month data collection, and 1,094 completed the 3-month follow-up data collection. The sample for PAFAS scales included 842 individuals who all completed baseline and 6-month follow-up data collection, 787 completed the 1-month data collection, and 703 completed the 3-month follow-up data collection.

¹ Maximum likelihood estimation uses all available data. Records for each individual are included for all available time points, regardless of whether the individual has missing data at one or more time point. Despite these missing records, maximum likelihood projects what would happen in later time points by “borrowing” information from values of the dependent variable at earlier time points while also accounting for the uncertainty of the projection in calculating standard errors and test statistics.

1. Key findings

1. We found no statistically significant treatment effects (using an alpha of .05 to indicate statistical significance) at the six-month study period for relationship satisfaction, emotion regulation, individual distress, or parenting and family relationships for couples who completed the six-month follow-up. We found small, statistically significant treatment effects for dyadic coping indicating that treatment group couples experienced small, positive changes to their dyadic coping communication at the six-month follow-up when compared to the wait-list control group couples.

Couples indicated change in the desired direction over the six-month study period; however, we found no statistically significant differences in the rate of change over the six-month study period between the treatment and wait-list control groups who completed the six-month follow-up. This means that treatment couples who completed the six-month study follow-up survey did not improve at a rate that was different than the wait-list control couples over the six-month study period. Below, we present specific findings by research question for both the primary and secondary questions.

Primary research questions:

1. What is the effect of relationship education on couples' perceptions of overall relationship satisfaction?
 - The mean difference in RDAS scores between intervention and comparison groups six months following the program was estimated to be 0.5. This difference was not statistically significant ($p = .5058$).
2. What is the effect of relationship education on couples' perceptions of their dyadic coping?
 - Controlling for baseline scores, the mean difference in DCI scores between intervention and comparison groups six months following the program was statistically significant ($p = .0032$). Individuals with average baseline score who receive the intervention were expected to score 3.3 points, on average, greater than those who did not receive the intervention. The standardized effect size for the mean difference was 0.20, representing a small effect.
3. What is the effect of relationship education on couples' perceptions of their ability to regulate emotions?
 - Controlling for baseline scores, the mean difference in DERS scores between intervention and comparison groups six months following the program was estimated to be -1.7. This difference was not statistically significant ($p = .2294$).
4. What is the effect of relationship education on couples' perceptions of their overall individual distress levels?
 - The mean difference in OQ scores between intervention and comparison groups 6 months following the program was estimated to be -1.8. This difference was not statistically significant ($p = .2238$).

Secondary research questions:

The secondary research questions address parenting and family adjustment outcomes for the subset of couples who participated as parents (research question 1). The remaining questions examine the rate of change between the treatment group and wait-list control group couples. We include specific results by research question below. Figures 1-4 in the appendix provide a visual depiction of the growth rate for research questions 2-5.

1. What is the effect of relationship education on couples' perceptions of parenting and family adjustment six months after study completion for participating couples who have children living at home under the age of 18?
 - The mean difference in PAFAS Parental Adjustment scores between intervention and comparison groups 6 months following the program was estimated to be -0.2. This difference was not statistically significant ($p = .4296$).
 - Controlling for baseline scores, the mean difference in PAFAS Family Relationship scores between intervention and comparison groups 6 months following the program was estimated to be -0.3. This difference was not statistically significant ($p = .3290$).
 - Controlling for baseline scores, the mean difference in PAFAS Parental Teamwork scores between intervention and comparison groups 6 months following the program was estimated to be -0.3. This difference was not statistically significant ($p = .1185$).
2. How does relationship education affect the rate of change over the six-month study period in couples' perceptions of relationship satisfaction?
 - The difference in mean baseline RDAS score between intervention and comparison groups of 0.3 was not statistically significant ($p = .6856$). Overall, the mean score on RDAS accelerated at the one month data collection (positive coefficient of the linear term) then slowed over the three and six month follow up assessments (negative coefficient of the quadratic term). Further, the difference in RDAS growth rates over 6 months between intervention and comparison groups were 0.14 for the coefficients of the linear term and -0.02 for the coefficients of the quadratic term. These differences were not statistically significant ($p = .6537$ for the coefficients of the linear term, $p = .7009$ for the coefficients of the quadratic term).
3. How does relationship education affect the rate of change over the six-month study period in couples' perceptions of dyadic coping?
 - The difference in mean baseline DCI scores between intervention and comparison groups of 1.8 was not statistically significant ($p = .2171$). Further, the difference in DCI growth rates over 6 months between intervention and comparison groups were 0.82 for the coefficients of the linear term and -0.08 for the quadratic term. These differences were not statistically significant ($p = .1959$ for the coefficients on the linear term, $p = .4142$ for the coefficients on the quadratic term).
4. How does relationship education affect the rate of change over the six-month study period in couples' perceptions of emotion regulation?
 - The difference in mean baseline DERS score between intervention and comparison groups of -1.6 was not statistically significant ($p = .2133$). Further, the difference in DERS growth rates over 6 months between intervention and comparison groups were -0.08 for the coefficients of the linear

term and 0.01 for the quadratic term. These differences were not statistically significant ($p = .8716$ for the coefficients on the linear term, $p = .8775$ for the coefficients on the quadratic term).

5. How does relationship education affect the rate of change over the six-month study period in couples' perceptions of individual distress?

- The difference in mean baseline OQ score between intervention and comparison groups was of -2.0 was not statistically significant ($p = .1601$). Further, the difference in OQ growth rates over 6 months between intervention and comparison groups were -0.02 for the coefficients of the linear term and 0.01 for the coefficients on the quadratic term. These differences were not statistically significant ($p = .9725$ for the coefficients on the linear term, $p = .8820$ for the coefficients on the quadratic term).

Table V.1. Covariates included in impact analyses

Covariate	Description of the covariate
DCI score at baseline	Score on the DCI measure at baseline was included in the model for the DCI outcome at the 180-day follow up
DERS score at baseline	Score on the DERS measure at baseline was included in the model for the DERS outcome at the 180-day follow up
PAFAS FR score at baseline	Score on the PAFAS FR measure at baseline was included in the model for the PAFAS FR outcome at the 180-day follow up
PAFAS PT score at baseline	Score on the PAFAS PT measure at baseline was included in the model for the PAFAS PT outcome at the 180-day follow up

Table V.2. Post-intervention estimated effects using data from the RDAS, DCI, DERS, and OQ at 180-day follow up data collection to address the primary research questions

Outcome measure	Intervention mean estimate	Comparison mean estimate	Mean difference between groups (standard error)	p-value
Primary RQ1: RDAS	51.4066	50.8715	0.5351 (0.8039)	.5058
Primary RQ2: DCI ^a	142.5680	139.2300	3.3380 (1.1272)	.0032
Primary RQ3: DERS ^a	64.2461	65.3182	-1.0721 (0.8912)	.2294
Primary RQ4: OQ	36.2608	38.0483	-1.7875 (1.4679)	.2238

Source: Follow-up surveys including the Revised Dyadic Adjustment Scale (RDAS), Dyadic Coping Inventory (DCI), Difficulty in Emotion Regulation Scale (DERS), and Outcome Questionnaire (OQ) measures were administered 6 months after the program.

Notes: p-values are included in the last column. See Table IV.2 for a more detailed description of each measure and Appendix D for a description of the impact estimation methods.

^aThe outcome measure score at baseline was included as a covariate to control for a small baseline difference between intervention and comparison groups.

Table V.4. Post-intervention estimated effects using data from the parenting measure at the 180-day follow up and from the RDAS, DCI, DERS, and OQ over the four data collection time points to address the secondary research questions

Outcome measure	Intervention estimate	Comparison estimate	Estimated intervention effect (standard error)	p-value
Secondary RQ1: PAFAS parenting scales				
Parental Adjustment mean	4.1228	4.3647	-0.2419 (0.3060)	.4296
Family Relationship mean ^a	3.1681	3.4202	-0.2521 (0.2580)	.3290
Parental Teamwork mean ^a	2.2825	2.6049	-0.3224 (0.2062)	.1185
Secondary RQ2: RDAS				
Intercept (baseline mean)	50.6905	50.3703	0.3202 (0.7906)	.6856
Linear term	0.5683	0.4285	0.1398 (0.3115)	.6537
Quadratic term	-0.0783	-0.0597	-0.0186 (0.0483)	.7009
Secondary RQ3: DCI				
Intercept (baseline mean)	139.6139	137.7700	1.8439 (1.4928)	.2171
Linear term	1.7573	0.9371	0.8202 (0.6338)	.1959
Quadratic term	-0.2098	-0.1295	-0.0803 (0.0983)	.4142
Secondary RQ4: DERS				
Intercept (baseline mean)	68.8942	70.4545	-1.5603 (1.2527)	.2133
Linear term	-1.8704	-1.7864	-0.0843 (0.5211)	.8716
Quadratic term	0.1778	0.1653	0.0125 (0.0811)	.8775
Secondary RQ5: OQ				
Intercept (baseline mean)	38.6725	40.6883	-2.0158 (1.4337)	.1601
Linear term	-1.7582	-1.7377	-0.0205 (0.5947)	.9725
Quadratic term	0.2315	0.2178	0.0137 (0.0825)	.8820

Source: For question 1, follow-up surveys including the PAFAS scales were administered 6 months after random assignment. For questions 2-5, surveys including the RDAS, DCI, DERS, and OQ measures were administered at baseline and again 1, 3, and 6 months after the random assignment.

Notes: p-values are included in the last column. The sample size for estimating the parenting scales models was 904-909 individuals. The sample size for estimating the growth models was 1,345-1,353 individuals. See Table IV.3 for a more detailed description of each measure and Appendix D for a description of the impact estimation methods.

^aThe outcome measure score at baseline was included as a covariate to control for a small baseline difference between intervention and comparison groups.

C. Sensitivity analyses

Sensitivity analysis for the four outcome measures associated with secondary research questions 2-5 were conducted using samples of participants who completed all four rounds (pre, 1-month post, 3-month follow-up, and 6-month follow-up). These samples include all participants with complete data and no missing data points during the 6-month data collection period. The same multilevel growth models specified for the analytic samples for secondary research questions 2-5 (participants who completed the six-month follow-up but may have had missing timepoints) were estimated using data from this smaller sample of complete data. Comparisons between the

models estimated using the two samples were made based on differences in magnitude of the intercept (baseline scores) and the slope (growth rate) parameter estimates for intervention and comparison groups. The reduced sample included 1,030 with data from all time points. Participants with complete data throughout the study period may be different in some way than those who had missing data. For example, participants with complete data may be more motivated, have more resources/support, or experience relationship stress differently than those who missed sessions during the study. Therefore, we conducted this sensitivity analysis to determine if any potential differences would result in changes to outcomes examining intervention effects.

1. Key findings

The results of the statistical significance tests using the reduced sample of individuals with complete data did not change for any of the analyses associated with research questions 2-5. The magnitude of the changes in parameter estimates were small. The largest differences were observed for the OQ outcome where the mean baseline estimates for both the intervention and comparison groups differed by about 0.4 points. This suggests the results of the original analyses are robust to changes in how the sample is defined.

Table V.3. Differences in parameter estimates between intervention and comparison groups estimated using alternative methods for secondary research questions 2-5

Outcome	Reduced sample: Sample that completed all four measures (pre, post [1-month], 3-month, and 6-month)	Difference in intervention estimate (primary - reduced sample)	Difference in comparison estimate (primary - reduced)
Secondary RQ2: RDAS	No change in significance test decisions		
Intercept (baseline mean)		0.0776	-0.1680
Linear growth		0.0860	0.0783
Quadratic growth		-0.0387	-0.0314
Secondary RQ3: DCI	No change in significance test decisions		
Intercept (baseline mean)		0.1878	-0.1100
Linear growth		0.1488	0.0642
Quadratic growth		-0.0402	-0.0223
Secondary RQ4: DERS	No change in significance test decisions		
Intercept (baseline mean)		-0.0215	0.3378
Linear growth		-0.2137	-0.2157
Quadratic growth		0.0439	0.0540
Secondary RQ5: OQ	No change in significance test decisions		
Intercept (baseline mean)		-0.3715	0.5035
Linear growth		-0.1531	-0.1785
Quadratic growth		0.0339	0.0560

Source: For secondary research questions 2-5, surveys including the RDAS, DCI, DERS, and OQ measures were administered at baseline and again 1, 3, and 6 months after the program.

Notes: The reduced sample size of participants with complete data was 1,030 individuals.

VI. DISCUSSION

The primary objective of the current evaluation was to examine the impact of PREP's *Within Our Reach* curriculum on couple functioning. Primary outcomes included relationship satisfaction, dyadic coping, emotion regulation, and individual distress. We also examined intervention effects on parenting and family adjustment for a sub-sample of couples who reported having children under the age of 18. Results of the six-month impact analyses indicated no statistically significant effects, except for dyadic coping which showed a positive and statistically significant impact ($d = .20, p = .0032$). Couples across both groups demonstrated positive change trajectories over the six-month study period. However, growth models indicated no statistically significant differences in the rate of change over the study period between the treatment and wait-list control couples.

With a randomized sample of 1,418 couples, this evaluation represents one of the largest tests of community-based relationship education to date. Previous large-scale randomized trials of relationship education resulted in mixed-findings (Wood et al, 2010; Hseuh et al, 2012). When effects were detected, they tended to be small. Similar to previous randomized trials of RE, results of the current study are indeterminate. While most contrasts between the treatment and control group couples were not statistically significant, only one contrast--dyadic coping--showed a statistically significant positive impact of the WOR curriculum. Unlike previous studies, the current impact evaluation included a wait-list control group with couples receiving an abbreviated version of the WOR curriculum after six-month data collection. Additionally, the wait-list was an active control group, with both treatment and wait-list couples receiving booster workshops on topics related to career and finances. It was anticipated that this design would encourage wait-list couples to remain in the study as they would not be stagnant during the six-month waiting period. However, it is possible that receiving these booster services diminished testable intervention effects because wait-list couples likely felt more hopeful about their relationship because they were receiving some services and had the knowledge that they would be receiving the intervention (in a retreat format) after the six-month study period.

Moreover, wait-list couples demonstrated lower attrition than treatment couples at the 30, 90, and 180-day follow-ups. The study team monitored attrition regularly, and met several times with the implementation team to discuss strategies to improve retention, especially for treatment couples. The project staff implemented a number of efforts to improve both overall and differential attrition, including (a) offering Walmart gift cards at specific points throughout the study, as well as other incentives to treatment couples; (b) providing childcare and meals; (c) following up regularly and confirming appointments; and (d) assessing *intent to attend* the next scheduled visit. Participants' *intent to attend* was assessed by asking all participants one question at the conclusion of each of their first three assessment meetings (baseline, one and three month follow-ups): *How likely is that you will attend the next workshop?* Participants responded on a scale of 1 (not likely) to 5 (very likely). Program staff reviewed responses and made attempts to follow-up with participants who indicated a response of '4' or lower. The majority of participant responses were '5' as people did not typically intend to miss scheduled meetings.

The evaluation team also conducted focus groups with couples who told us the primary reasons they did not attend at various times were due to factors such as changes to work schedules, unexpected issues with children (e.g., illness), or not being able to navigate rush-hour traffic in order to reach the facility in time for the start of the workshop. After speaking with couples and program staff during focus groups, we speculated that the differential study attrition was due in large part to treatment couples being asked to attend the center two additional times after completing the intervention, which is two more times than the wait-list control group. Couples told us that despite being enthusiastic about the intervention, they often experienced structural and contextual factors that prevented their attendance. Some examples that were mentioned included opportunities to work overtime, caring for sick children, and leaving too late to attend the workshop on time (Orlando is a sprawling metropolitan area and many couples drove across town to attend the sessions). Thus, the more visits requested of couples presented more opportunities for last-minute structural barriers.

It is possible that providing the relationship education workshops to treatment couples in a weekend-retreat format would have mitigated some attrition by reducing the number of visits and thereby lowering some of the burdens participants experienced. However, delivering the curriculum in a weekend retreat format would also entail a reduction in content hours because there is not enough time to cover twelve hours in a weekend. Moreover, treatment group couples maintained relatively strong retention to the four-week intervention. Given that couples reported that the primary reason for participating in the program was to improve their relationship, it is possible that they no longer felt as committed to the larger study after completing the relationship education intervention. The follow-up booster workshops focused on economic well-being, which may not have been of major interest to some couples. Programs that implement briefer, but intense delivery of RE early on, and then follow-up with booster sessions that include content from the RE workshops may generate better retention from couples. Additionally, such a format might contribute to longer-term effects by continuing to reinforce, or teach, relationship skills over a longer duration.

The biggest challenge to the current study was overall and differential attrition. Enrollment benchmarks were met or exceeded, and most treatment couples completed the 30-day *Within Our Reach* intervention. The implementation analysis provides some insight into the conclusion about differential attrition. When asked about intentions to return, over 80% of respondents indicated a high likelihood. It is possible that respondents reported this intention before experiencing the structural barriers that prevented their future attendance. Additionally, the implementation analysis revealed that most couples (99%) said they had a positive experience participating in the workshops; would recommend (or did recommend) the program to family or friends (92%); and thought the program was important or worthwhile (98%). However, the perceived value may not have been enough to overcome the challenges that low-income couples in Orlando experience when attending a program that requires consecutive sessions and time commitments.

We implemented a novel tool, *intent-to-attend*, to assess participants' overall intentions to return to their next scheduled visit. Participants responded on a five-point scale (1 – not at all likely, and 5 – very likely) indicating *How likely are you to attend your next schedule visit?* This

question was administered at baseline, 30- and 90-day follow-up. We hoped that by asking this question three times over the six-month study period, we could proactively address participants who indicated a low likelihood of returning. We found participant responses to be somewhat predictive of attendance at the next session, but also found that most participants indicated a strong likelihood to return. This could be associated with their overall positive experiences with the program services. We attempted to follow-up with couples who indicated less than the highest likelihood of returning (a score of 4 or less). We hoped these outreach attempts and conversations would help us address participant concerns and prevent them from dropping out of the study, or missing a session. In some cases, we found these conversations helpful, while in others, participants told us they did not indicate the highest response because they could not predict the future and unforeseen events. Family case managers carried a high caseload, which also prevented them from following-up with all couples who indicated less than the highest likelihood of returning. Thus, despite being satisfied with the program services, couples experienced, and some anticipated, unexpected challenges to their schedule. They told us these challenges included things like kids being sick, or opportunities to work overtime. Additionally, because family case managers had high caseloads, they may not have been able to spend as much individualized time with couples as necessary. The information provided by participants on the *intent-to-intend* question was helpful, and has additional potential, pending staff availability to address each person's response. Future studies should include this question, or similar questions, to proactively address participant concerns in order to minimize attrition. Additionally, case managers should have smaller caseloads, allowing them more time to spend time with couples and to engage in a coaching-style relationship. This may require investing in more managers, enrolling fewer participants, or developing a strategy that includes a tiered case management approach such that participants might receive different levels of case management contact depending on factors such as baseline distress level, needs assessment, or program participation.

Despite higher attrition at six-months, the current study revealed positive intervention effects for treatment couples on dyadic coping. Dyadic coping, as measured in this evaluation, assessed couples' stress coping and communication. Thus, *Within Our Reach* helped couples improve on their overall stress communication six-months after random assignment, and five-months after completing the intervention. *Within Our Reach* emphasizes healthy communication, creating relationship safety, listening with empathy, and intentionality in communication. It is possible that this emphasis, along with the structure of the communication tools, contributed to couples feeling more supported and encouraged in their relationships. The current evaluation did not measure specific mechanisms of change, thus we can only speculate what aspects of the intervention contributed to improved stress communication for couples.

There are limitations to note in the current evaluation. First, as previously discussed, there was large overall and differential attrition during the six-month study period. The final analysis sample did not demonstrate baseline effects greater than .08, thus indicating relative equivalence, but the loss of power due to attrition may have influenced the ability to detect intervention effects. Additionally, participants completed large surveys with several questions, that may have contributed to test fatigue. Participants spent between 30 minutes to 1 hour and 30 minutes

completing the assessments. In a focus group conducted during the evaluation, one noted area of feedback was to reduce the number of questions that participants were asked to complete.

We did not examine treatment effects in this report at the post-assessment (30 days post random assignment) and first follow-up (90-days post random assignment). It is possible that intervention effects existed, but faded by the six-month follow-up. We also included only total scores on all assessments, and did not examine sub-scales on outcome measures. It is also possible that intervention effects could be detected with a more nuanced examination on the specific scale scores.

Finally, family case managers on the implementation/program team carried high caseloads with numerous couples at any given time. This reduced their ability to contact couples more frequently and did not allow for the possibility to implement coaching, or more tailored instruction and skill reinforcement to couples. Coaching has been shown to help increase intervention effects for couples in relationship education (Roddy, Rothman, & Doss, 2018). Future RE programming should include smaller case manager caseloads and include coaching for couples.

VII. REFERENCES

- Amato, P., & Booth, A. (1996). A prospective study of divorce and parent-child relationships. *Journal of Marriage and Family*, 58(2), 356-365. doi:10.2307/353501.
- Barden, S. M., Carlson, R. G., Daire, A. P., Finell, L. R., Christopher, K., & Young, E. (2015). Investigating the influence of relationship education on parental attitudes. *Marriage & Family Review*, 51(3), 246-263.
- Bodenmann, G., & Cina, A. (2005). Stress and coping among stable-satisfied, stable-distressed, and separated/divorced swiss couples: A 5-Year prospective longitudinal study. *Journal of Divorce & Remarriage*, 44(1-2), 71-89. https://doi.org/10.1300/J087v44n01_04.
- Busby, D. M., Christensen, C., Crane, D. R., & Larson, J. H. (1995). A revision of the dyadic adjustment scale for use with distressed and nondistressed couples: Construct hierarchy and multidimensional scales. *Journal of Marital and Family Therapy*, 21(3), 289-308. <https://doi.org/10.1111/j.1752-0606.1995.tb00163.x>.
- Carlson, R. G., Barden, M. S., Daire, A. P., & Swartz, M. (2014). Examining parental alliance for low-income couples who attend relationship education with or without a partner. *Journal of Couple and Relationship Therapy*, 13(2), 153-170.
- Clavarino, A., Hayatbakhsh, M. R., Williams, G. M., Bor, W., O'Callaghan, M., & Najman, J. M. (2011). Depression following marital problems: Different impacts on mothers and their children? A 21-year prospective study. *Social psychiatry and psychiatric epidemiology*, 46(9), 833-841.
- Charles, P., Orthner, D. K., Jones, A., & Mancini, D. (2006). Poverty and couple relationships: Implications for welfare policy. *Marriage and Family Review*, 39, 27-52.
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences (2nd ed.)*. Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.
- Falconier, M. K., Nussbeck, F., & Bodenmann, G. (2013). Dyadic coping in Latino couples: Validity of the Spanish version of the Dyadic Coping Inventory. *Anxiety, Stress & Coping: An International Journal*, 26(4), 447-466. <https://doi.org/10.1080/10615806.2012.699045>.
- Gratz, K. L., & Roemer, L. (2004). Multidimensional assessment of emotion regulation and dysregulation: Development, factor structure, and initial validation of the difficulties in emotion regulation scale. *Journal of Psychopathology and Behavioral Assessment*, 26(1), 41-54.
- Hair, E. C., Sidorowicz, K., Martin, L., & Milot, A. (2011). The mental health of vulnerable youth and their transition to adulthood: Examining the role of the child welfare, juvenile justice, and runaway/homeless systems. *Free Inquiry in Creative Sociology*, 39(2), 35-48.
- Hardie, J. H., & Lucas, A. (2010). Economic factors and relationship quality among young couples: Comparing cohabitation and marriage. *Journal of marriage and the family*, 72(5), 1141-1154. <https://doi.org/10.1111/j.1741-3737.2010.00755.x>.
- Hsueh, J., Alderson, D. P., Lundquist, E., Michalopoulos, C., Gubits, D., Fein, D., & Knox, V. (2012). *The Supporting Healthy Marriage evaluation: Early impacts on low-income families*. Washington, DC: Administration for Children and Families, Office of Planning, Research, and Evaluation.
- Johnson, C. A., Stanley, S. M., Glenn, N. D., Amato, P. R., Nock, S. L., Markman, H. J., & Dion, M. R. (2002). *Marriage in Oklahoma: 2001 baseline statewide survey on marriage and divorce*. Oklahoma City, OK: Oklahoma Department of Human Services.

- Karney, B. R., & Bradbury, T. N. (2005). Contextual influences on marriage: Implications for policy and intervention. *Current directions in psychological science, 14*(4), 171-174.
- Kashy, D.A., Donnellan, M.B., Burt, S.A., & McGue, M. (2008). Growth curve models for indistinguishable dyads using multilevel modeling and structural equation modeling: The case of adolescent twins' conflict with their mothers. *Developmental Psychology, 44*(2), 316–329. Online appendix: <http://dx.doi.org/10.1037/0012-1649.44.2.316.supp>
- Kenny, D.A., Kashy, D.A., & Cook, W.L. (2006). *Dyadic Data Analysis*. The Guilford Press: New York, NY.
- Lambert, M. J., Morton, J. J., Hatfield, D. R., Harmon, C., Hamilton, S., Shimokawa, K., et al. (2004). *Administration and scoring manual for the Outcome Questionnaire (OQ 45.2)*. (3 ed.). Wilmington, DE: American Professional Credentialing Services LLC.
- Ledermann, T., Bodenmann, G., Gagliardi, S., Charvoz, L., Verardi, S., Rossier, J., Bertoni, A., & Iafrate, R. (2010). Psychometrics of the dyadic coping inventory in three language groups. *Swiss Journal of Psychology, 69*(4), 201-212.
- Moore, K. A., Kinghorn, A., & Bandy, T. (2011). *Parental relationship quality and child outcomes across subgroups* (Publication No. 2011-13). Retrieved from Child Trends website: <http://www.childtrends.org/?publications=parental-relationship-quality-andchild-outcomes-across-subgroups>.
- Neff, L. A., & Karney, B. R. (2004). How does context affect intimate relationships? Linking external stress and cognitive processes within marriage. *Personality and Social Psychology Bulletin, 30*(2), 134-148.
- Owen, J., Antle, B., & Barbee, A. (2013). Alliance and group cohesion in relationship education. *Family Process, 52*, 465–476. doi: 10.1111/famp.12039.
- Randall, A., Hilpert, P., Jimenez-Arista, L. E., Walsh, K. J., & Bodenmann, G. (2016). Dyadic coping in the U.S. psychometric properties and validity for use of the english version of the dyadic coping inventory. *Current Psychology, 35*(4), 570-582. <https://doi.org/10.1007/s12144-015-9323-0>.
- Roddy, R. K., Rothman, K., & Doss, B. D. (2018). A randomized controlled trial of different levels of coach support in an online intervention for relationship distress. *Behaviour Research and Therapy, 110*, 47-54. <https://doi.org/10.1016/j.brat.2018.09.002>.
- Sanders, M. R., Morawska, A., Haslam, D. M., Filus, A., & Fletcher, R. (2014). Parenting and family adjustment scales (PAFAS): Validation of a brief parent-report measure for use in assessment of parenting skills and family relationships. *Child Psychiatry & Human Development, 45*, 255-272. <https://doi.org/10.1007/s10578-013-0397-3>.
- Schramm, D. G., & Harris, V. W. (2011). Marital quality and income: An examination of the influence of government assistance. *Journal of family and economic issues, 32*(3), 437-448.
- Singer, J.D. (1998). Using SAS PROC MIXED to fit multilevel models, hierarchical models, and individual growth models. *Journal of Educational and Behavioral Statistics, 23*(4), 323-355.
- Stanley, S. M., Carlson, R. G., Rhoades, G. K., Markman, H. J., Ritchie, L. L., & Hawkins, A. J. (2020). Best practices in relationship education focused on intimate relationships. *Family Relations, 69*, 497-519.
- Waite, L. J., & Gallagher, M. (2000). *The case for marriage: Why married people are happier, healthier and better off financially*. New York: Doubleday.

Wood, R. G., Moore, Q., Clarkwest, A., Killewald, A., & Monahan, S. (2012). *The long-term effects of Building Strong Families: A relationship skills education program of unmarried parents*. Washing, DC: Administration for Children and Families, Office of Planning, Research, and Evaluation.

VIII. APPENDICES

A. Logic Model (or Theory of CHANGE) for the program

ASSUMPTIONS: Research: RCT design- treatment/wait-list-control group Intervention (PREP) – evidence-based approach Best practices: Incentives/sensitivity to common barriers for low income families; inclusion of women and minorities Experience: Existing history, recruitment, structure of project, facilities, lead staff, community partnerships					
Inputs	Target Population	Activities/Approach	Goals/Objectives	Outputs (ST)	Outcome (LT)
Marriage and Family Research Institute, University of Central Florida Partnerships/ Community Organizations Key Staff	300 low-income couples per year (150 enrolled in treatment group; 150 enrolled in wait-list-control) <i>Note: This was adjusted to 450 couples per year for years 3-4</i> Inclusion criteria: a) low-income b) married or unmarried, and c) parent or non-parent	RCT – randomly assign participants into the treatment or wait-list control <ul style="list-style-type: none"> Treatment- <ul style="list-style-type: none"> – WOR curricula – Case management – Career Pathways Assessments Intake, pre, post, 90 days and 6-month follow-ups <ul style="list-style-type: none"> Waitlist- Control- <ul style="list-style-type: none"> – Career Pathways – 90 day delayed start for WOR/WMR curricula – Case Management Assessments Intake, 30 day, 60 day, pre, post, 90 day follow-up	<ol style="list-style-type: none"> 1. Enroll low-income couples 2. Increase relationship satisfaction 3. Increase co-parenting alliance 4. Increase couples in healthy relationships, 5. Increase conflict resolution 6. Increase communication skills 7. Improve access to community social services, (financial literacy and job/career preparedness); (8) Increase dyadic coping, (9) Decrease parental stress 	70% of participants will report improvement in: <ul style="list-style-type: none"> • Healthy relationships Communication Conflict resolution, Relationship satisfaction, Dyadic Coping • Improved co-parenting alliance and decreased parental stress • Economic stability- Job training, job skills, financial literacy 	Improved family functioning <i>Proximal:</i> Increase in communication skills, dyadic coping, conflict resolution <i>Distal:</i> Increase in relationship satisfaction and parental alliance/stress Increased economic stability and mobility <i>Proximal:</i> Increase in career-related skills <i>Distal:</i> Reduced poverty

B. Data and study sample

Implementation data

This section provides data sources for the implementation study, and outlines the measures that were used to address the implementation study research questions. We also present the implementation data analysis strategy.

Recruitment. Items used to assess participant recruitment came from the nFORM applicant characteristics survey, attendance records, and internal program tracking documents. In order to understand the recruitment sources from which enrolled participants first heard about the study, we examined a frequency distribution for the coded responses from enrolled participants. The initial sample for this frequency distribution was 1,390 couples, or 2,780 individuals. Next, we coded attendance at each round of data collection (1-4) as either attended (1) or not attended (0). We also categorized each of the recruitment sources into a binary variable to represent the two primary recruitment strategies of active (1) or passive (0). Recruitment strategies were considered active if a program staff member actively recruits a participant by describing the study, collecting their contact information, and following-up with him or her at a later time. Examples of active recruitment strategies included talking to people at the library, community events, or in waiting rooms at partner agencies. A strategy was considered passive if the participant learned about the study on his or her own, and called the program office unsolicited to find out about the study (e.g., website, word or mouth, flyers posted in the community). Once we coded attendance and recruitment strategy, we conducted a binomial logistic regression to examine the relationship between recruitment strategy and attendance. We coded recruitment strategy as either active (1) or passive (0), and attendance as either completed the six-month follow-up (1) or not completed (0).

Quality. We administered a program satisfaction survey to all study participants. Participants assigned to the treatment group completed the satisfaction survey twice (at one month follow-up immediately following the intervention, and again at six-month follow-up), while control participants only completed the survey at the six-month follow-up. Responses to each question were indicated on a Likert scale from (1) *Strongly Agree* to (5) *Strongly Disagree*. We examined the distribution of responses to these items. We presented the distribution by random assignment group (i.e., treatment/control). Additionally, we examined responses to the PREP Alliance Measure that was administered to all treatment group participants at the one-month follow-up immediately following the intervention. Participants indicated responses on a 7-point Likert scale from (1) *Strongly Disagree* to (7) *Strongly Agree*. We presented the distribution of responses to items 3, 7, and 9 because those items deal specifically with participants' beliefs about the workshop leaders.

Engagement/Retention. To understand participant motivation for attending and completing the study, we examined participant responses on the nFORM applicant characteristics questionnaire (E2) as well as the participant satisfaction survey for all program participants who enrolled and provided a response for these. nFORM item E2 provides a list of options that participants may

select from, and the participant satisfaction survey asked participants to rank order pre-identified reasons that motivated them to start the study. The nFORM data was collected at baseline, while the participant survey data was collected at six-month follow-up, thus allowing us to visually compare the distribution of responses. We provided a distribution of responses for each item. To understand why participants may have missed a scheduled meeting, we provided a distribution of responses to the participant satisfaction survey where participants were asked to rank order pre-selected items that best described why they missed a meeting. Finally, at baseline, one month and three month follow-up, we asked participants to rate the likelihood they would attend their next scheduled meeting, with 1 indicating *not at all likely* and 5 indicating *very likely*. We presented the distribution of responses for this question at each of the first three data collection time points (baseline, one month, and three months). We also assessed how closely responses corresponded with actual attendance.

Table B.1. Data used to address implementation research questions

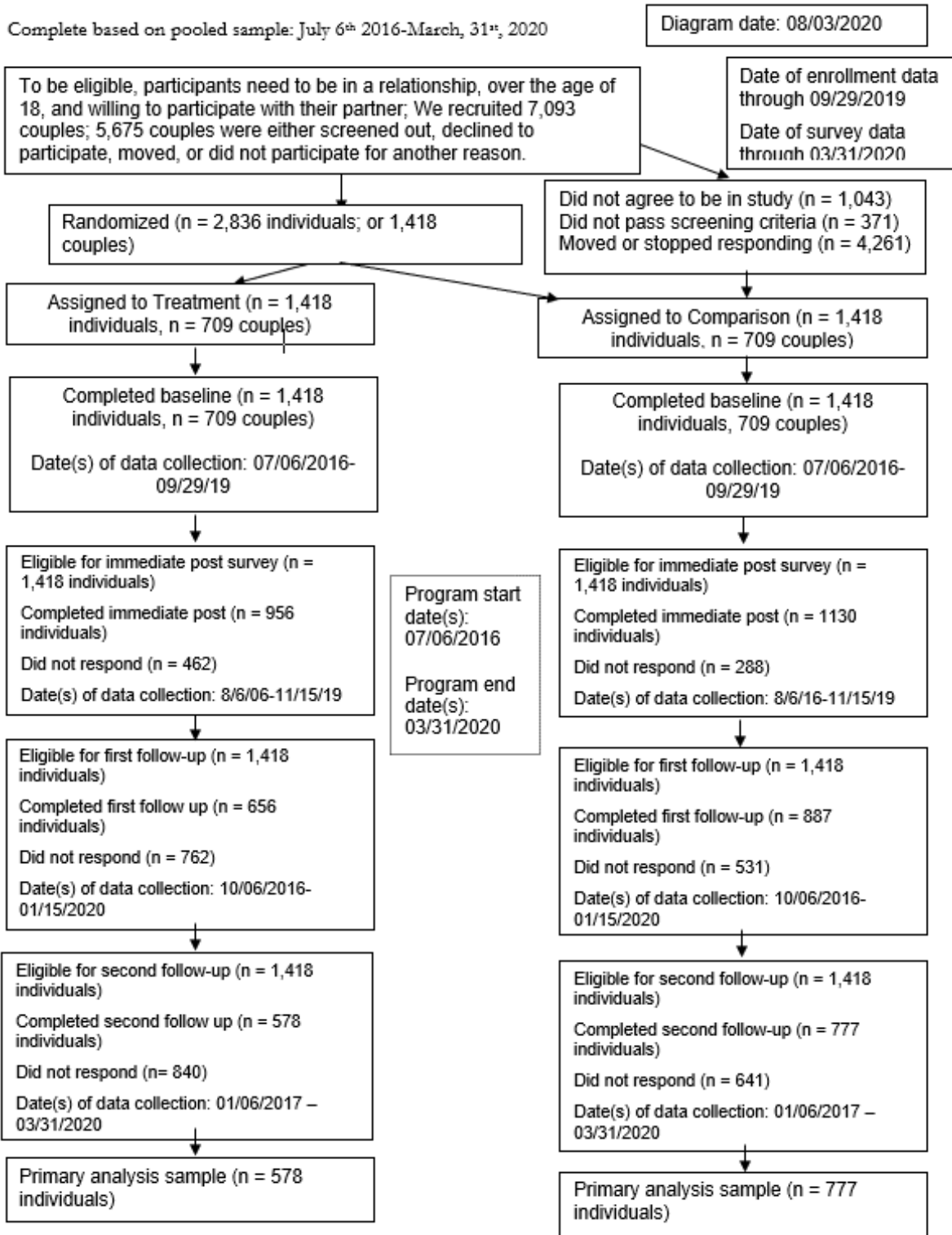
Implementation element	Research question	Data source	Timing/ frequency of data collection	Party responsible for data collection
Recruitment	What were the primary recruitment sources for enrolled participants?	Responses on nFORM applicant characteristics question asking participants to list how they heard about the program (E1); Recruitment staff tracking spreadsheet	Enrollment	Intervention staff
Recruitment	Additionally, were recruitment strategies (i.e., passive versus active) related to study completion?	nFORM attendance records: We will group recruitment sources into either active or passive categories	Every workshop session attended	Intervention staff
Quality	How satisfied were study participants with the overall program model?	Select responses from the PH Participant satisfaction survey items: I am satisfied with the content of the workshops I am satisfied with the person(s) who taught my workshops I am satisfied with the services provided by my family case manager Would you recommend (or have you recommended) this study to your family or friends?	180-Day follow-up	Intervention staff
Quality	What relationship did participants have with program staff, including relationship educators?	Select responses from the PREP Alliance Measure: I believe the educators liked me I felt that the educators appreciated me The educators and I trusted one other	30-day follow-up (treatment only)	Intervention staff
Engagement /Retention	What were primary motivators for enrolling in the program?	Question E2 from nFORM applicant characteristics survey: Why did you choose to enroll in this program?; Question 4 from PH Satisfaction Survey: Please rank order the primary reasons that motivated you to start the program	180-day follow-up	Intervention staff
Engagement/ Retention	What were the main reasons for missing scheduled workshops?	Question 5 from the PH Satisfaction Survey: Please rank order the primary reasons that motivated you to complete the program?	180-day follow-up	Intervention staff
Engagement/ Retention	Did participants plan to continue participating in the study when asked about their future intentions to return to their next scheduled workshop? Did responses predict attendance?	From the intent to attend questionnaire: How likely is it that you will attend your next scheduled visit?	Enrollment; 30- and 90-day follow-up	Intervention staff

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

	Data source	Timing of data collection	Mode of data collection	Party responsible for data collection	Start and end date of data collection
Intervention	Intervention group study participants	<ul style="list-style-type: none"> • Enrollment (baseline) • End of intervention (30-days after enrollment) • Booster session 1 (90-days after enrollment) • Booster session 2 (180-days after enrollment) 	In-person online survey (nFORM; SMIS)	Program staff	July 2016 through March 2020
Counterfactual	Comparison group study participants	<ul style="list-style-type: none"> • Enrollment (baseline) • Booster session 1 (30-days after enrollment) • Booster session 2 (90-days after enrollment) • 180-day Follow-up 	In-person online survey (nFORM)	Program staff	July 2016 through March 2020

C. CONSORT diagram for individual clients, for studies in which consent occurred before assignment

Figure C.1. CONSORT diagram



D. Data preparation

The impact evaluation data merging process involved combining data from two primary data sources: 1) nFORM – pre and post-surveys as well as baseline characteristics; and 2) supplemental management information system (SMIS) – a researcher-developed app and web-based method of collecting all outcome measure data. The merging of these two data sources involved a multi-step process that the evaluation team completed every month.

The data cleaning and merging process began with the evaluation team accessing nFORM data for the desired data date range. We then recoded the nFORM export labels into the Statistical Packaging for the Social Sciences (SPSS - the platform used to maintain the large, master data file) labels. We relabeled both pre- and post-survey files and deleted variables not collected in the current study. The cleaned pre- and post-survey data was then transferred to the corresponding participant into the SPSS master file. Because the unique participant id existed in both files, this process was a simple merge on participant id.

The second data cleaning and merging process for Project Harmony included exporting the SMIS raw data directly from the web-based side of the software. The evaluation team had a unique username and password for accessing the outcome data in real-time. We structured the data to coincide with participants' nFORM data, and exported that data into the appropriate round's (i.e., data collection wave) section in the SPSS master file. We then transferred all relevant data into the main data file. Throughout the process of data merging, the evaluation team conducted quality control checks to ensure that the data corresponded with the appropriate participant by cross-referencing their exported data to the master (excel and SPSS) files. We also performed monthly random data checks to ensure quality control.

E. Impact estimation

Baseline Equivalence

Chi-square tests of independence were used to test for associations between treatment group and categorical demographic variables for the analytic sample mapped to each research question and outcome. The hypotheses of the chi-square test are as follows:

H₀: There is no association between the two variables (treatment group membership and the demographic variable of interest).

H_a: There is an association between the two variables (treatment group membership and the demographic variable of interest).

The formula for the chi-square statistic is as follows:

$$\chi^2_{df} = \sum \frac{(O_i - E_i)^2}{E_i}$$

where O_i represents the observed cell count, E_i represents the expected cell count assuming independence, and the sum is taken over all cells of the contingency table. The critical value for the chi-square statistic was determined by the level of alpha (statistical significance, which was .05 in this study) and the degrees of freedom. The degrees of freedom for the chi-square are calculated using the following formula: $df = (r-1)(c-1)$ where r is the number of rows and c is the number of columns. If the observed chi-square test statistic is greater than the critical value, the null hypothesis can be rejected. In addition, the phi coefficient (see formula below) was computed as an effect size for each treatment group and demographic variable association at baseline.

$$\phi = \sqrt{\frac{\chi^2}{n}}$$

Outcome Measures at Baseline

Independent two sample t-tests were conducted between intervention and comparison groups for the outcome measures of interest to determine whether groups differed significantly at baseline. The critical value for the t-test statistic was determined by the level of significance (.05 was used) and the degrees of freedom. If the observed t-test statistic is greater than the critical value, the null hypothesis can be rejected. In addition, Cohen's d for standardized mean difference was computed as an effect size measure for each measure at baseline. Effect size of 0.2 is considered small, 0.5 is considered medium, and 0.8 is considered large (Cohen, 1988). Based on What Works Clearinghouse guidelines, statistical adjustments were made in the analysis of outcomes for effect sizes greater than 0.05 and less than 0.25.

The hypotheses for the independent samples t-test are as follows:

H₀: The true mean difference between treatment and comparison groups = 0.

H_a: The true mean difference between treatment and comparison groups \neq 0.

The formula for the t-test statistic with pooled variance is as follows:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \text{ where } s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

Cohen's d was calculated as $d = \frac{|\bar{x}_1 - \bar{x}_2|}{s_p}$.

Analysis for Primary Research Questions

Primary outcome measures were analyzed for couples at the six-month follow-up period.

1. What is the effect of relationship education on couples' perceptions of overall relationship satisfaction?
2. What is the effect of relationship education on couples' perceptions of their dyadic coping?
3. What is the effect of relationship education on couples' perceptions of their ability to regulate emotions?
4. What is the effect of relationship education on couples' perceptions of their overall individual distress levels?

To address the primary research questions, multilevel models that accounted for the dyadic structure of the data were estimated using the outcomes measure of interest at 180-day follow up as the dependent variable. A treatment group indicator, X, was included as an independent variable to determine if differences exist between the treatment and comparison groups. Kenny, Kashy and Cook (2006) described the estimation process for multilevel models involving dyadic data and provided sample SAS code to estimate the models using PROC MIXED. In the multilevel framework, dyads can be viewed as groups composed of two persons. With dyadic data, the outcome variable, Y, is measured once for each person within the dyad. Thus, each dyad has two scores on Y. The application of multilevel models to dyadic data requires that slopes (i.e., the effect of X on Y for each dyad) be constrained to be equal across all dyads. That is, the model must be constrained to include only a fixed effect with respect to the effect of X on Y. The intercepts for the dyads can vary, and the nonindependence in the persons' scores within dyads is modeled through the variation of intercepts. In the formulas below, person is at level 1 and is nested in couple at level 2 where *i* represents the person (*i*=1, 2) and *j* represents the couple. The models are as follows:

$$\text{Level 1 model: } Y_{ij} = b_{0j} + b_{1j}X_j + e_{ij}$$

$$\text{Level 2 models: } b_{0j} = a_0 + u_{0j}$$

$$b_{2ij} = a_{20}$$

For measures that require statistical adjustment based on baseline equivalence assessment, the score on the measure at baseline (centered on the mean) was included as a person-level covariate in the model to

control for baseline differences. These include DCI for primary research question 2 and DERS for primary research question 3. Individual scores are treated as repeated measures in the dyad and the residual error structure is specified as compound symmetry. This means that the variances of the residuals at each time point for both members of the couple are set to be equal and that the covariances between the couples' residuals at the four time points are also equal. Nonindependence is estimated as a covariance. .

After estimating model parameters, the statistical significance of the slope coefficient indicates whether a difference exists on the mean score of the outcome at the 180-day follow up between treatment and comparison groups. For measures that require statistical adjustment based on baseline equivalence assessment, the score on the measure at baseline (centered on the mean) was included as a covariate in the model to control for baseline differences. SAS PROC MIXED with maximum likelihood estimation was used for the multilevel analysis (Kenny, Kashy, & Cook, 2006).

For statistically significant findings, a standardized effect size for the mean difference between the intervention and comparison groups was calculated based on Spybrook (2008). The formula is provided below.

$$\hat{\delta} = \left| \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\sigma^2 + \tau}} \right|$$

where τ represents the variance between dyads and σ^2 represents the variance within dyads.

Analysis for Secondary Research Questions

Secondary research question 1 addressed outcomes for sub-populations (e.g., parents) within the primary sample. Additionally, secondary questions 2-5 included examination of the rate of change, or the growth curve, for major intervention target areas over the six-month study period. Thus, the growth curve models estimate the rate of change over three measurement points (one month, three months, and six months post-intervention).

1. What is the effect of relationship education on couples' perceptions of parenting and family adjustment six months after study completion for participating couples who have children living at home under the age of 18?
2. How does relationship education affect the rate of change over the six-month study period in couples' perceptions of relationship satisfaction?
3. How does relationship education affect the rate of change over the six-month study period in couples' perceptions of dyadic coping?
4. How does relationship education affect the rate of change over the six-month study period in couples' perceptions of emotion regulation?
5. How does relationship education affect the rate of change over the six-month study period in couples' perceptions of individual distress?

Multilevel models for parenting measures at 180-day follow up were used to address secondary research question 1. The two-level model accounting for dyadic data structure presented in the previous section was used for this analysis as well. Three models were estimated using each of the three family adjustment PAFAS scales as the dependent variable: Parental Adjustment,

Family Relationship, and Parental Teamwork. For measures that required statistical adjustment based on baseline equivalence assessments, the score on the measure at baseline (centered on the mean) was included as a covariate in the model to control for baseline differences. These included the Family Relationship and the Parental Teamwork scales.

Multilevel growth models using the outcome measure of interest as the dependent variable were used to address secondary research questions 2-5. The means for each measure at each time point were calculated and plotted for treatment and control groups. Observed patterns revealed initial changes from baseline to the 1-month post-assessment, with a leveling off for the 3-month and 6-month follow up assessments. Thus, a curvilinear growth model that included both a linear and a quadratic term to account for change in growth rate over time was selected.

The analysis was based on methods described in Kashy et al. (2008). Growth modeling of dyadic data starts with growth functions for each individual. Two unique aspects of dyadic growth models compared to individual growth models are: (a) Certain parameter estimates must be pooled across dyad members, and (b) additional parameters are included that capture the degree of correspondence between dyad members' outcomes. The data were organized into a person-period data structure. Time was coded 0, 1, 3, and 6 representing data collection at baseline, 30 days, 90 days, and 180 days, respectively. The intercept is thus interpreted as the baseline and the follow-up data collection points account for the number of months from baseline data collection. SAS PROC MIXED with maximum likelihood estimation was used for the multilevel growth analysis (Singer, 1998; Kashy et al., 2008). We estimated separate models using the relevant outcome measure as the dependent variable. The single individual-level or lower level equation for outcome measure Y_{ijk} for person i in dyad (couple) j at time k where $i=1, 2$, j represents the dyad number, and $k=1, 2, 3, 4$ time periods is represented as follows.

$$Y_{ijk} = b_{0ij} + b_{1ij}T_k + b_{2ij}T_k^2 + b_{3ij}X_j + b_{4ij}T_k * X_j + b_{5ij}T_k^2 * X_j + e_{ijk}$$

The lower level intercepts and growth parameters are aggregated across the sample, as is true for growth models for individuals. However, the aggregation occurs both within dyads and across dyads, as presented in the following equations:

$$b_{0ij} = a_0 + u_{0ij}$$

$$b_{1ij} = a_1 + u_{1ij}$$

$$b_{2ij} = a_2$$

The individual intercepts represent the mean of the outcome measure for each member of the couple at the beginning of the study and are aggregated over individuals and dyads. This results in a single estimate of the average intercept, a_0 , and is modeled as a random effect. Similarly, the individual linear growth parameters estimating the initial growth rate of the outcome measure over time are aggregated over individuals and dyads. Thus, the average initial growth rate of the outcome measure across the sample is a single estimate of the linear growth parameter, a_1 , and is modeled as a random effect. Further, the individual quadratic growth parameters estimating the change in growth rate of the outcome measure over time are aggregated over individuals and dyads. Thus, the average change in growth rate of the outcome measure across the sample is a single estimate of the quadratic growth parameter, a_2 . When specifying the

quadratic growth term as random, the models failed to converge and thus this term was modeled as a fixed effect.

The b_{3ij} parameter represents the differential effect of the treatment group on the intercepts at baseline and is modeled as a fixed effect. The direction and significance of the b_{3ij} parameter estimate indicates if one group or the other has a greater mean at baseline. The b_{4ij} parameter represents the differential effect in linear growth rate for the treatment group (i.e., time by groups interaction) and is modeled as a fixed effect. The direction and significance of the b_{4ij} parameter estimate indicates if one group or the other has a faster initial rate of change over the six months. The b_{5ij} parameter represents the differential effect in quadratic growth rate for the treatment group (i.e., time by groups interaction) and is modeled as a fixed effect due to convergence issues when specifying as random. The direction and significance of the b_{5ij} parameter estimate indicates if one group or the other has a faster change in growth rate over the six months.

In this model, there are three random effects whose variances can be estimated as well as five covariances. The between-persons covariances represent the degree to which members within dyads are similar on the associated fixed effects parameter. That is, the $Cov(u_{01j}, u_{02j})$ measures the degree to which the couples are similar in outcome measure score at the beginning of the study. Similarly, the $Cov(u_{11j}, u_{12j})$ measures the degree to which the couples are similar in their linear growth in the outcome measure score over time.

The two random intercepts are constrained to the same value, and the two random slopes are similarly constrained to be equal. Equality constraints for the random effects were specified in SAS as described in the online appendix to Kashy et al. 2008 (<http://dx.doi.org/10.1037/0012-1649.44.2.316.supp>). I1 and I2 are dummy variables (coded 0 or 1) where I1 represents individual 1 and I2 represents individual 2 in a given couple. Note that the assignment of 1 or 2 is arbitrary as the data are indistinguishable. I1 and I2 represent the two members of the couples' intercept. I1*TIME and I2*TIME represent the two members of the couples' slope. The random effects equality constraints are accomplished by setting up a covariance matrix data set in SAS that specifies the values associated with each random effect parameter to ensure equality of associated pairs. The model further specifies that the intercepts and slopes can covary both within and between couple dyads, and the residual error structure is specified as compound symmetry. Treatment group and the interaction of treatment group with time were treated as fixed.

F. Sensitivity analyses and alternative model specifications

Sensitivity Analysis for Secondary Research Questions 2-5

Sensitivity analysis for the four outcome measures associated with secondary research questions 2-5 were conducted using samples of participants who completed all four rounds (pre, 1-month post, 3-month follow-up, and 6-month follow-up). These samples included all participants with complete data and no missing data points during the 6-month data collection period. The same multilevel growth models specified for the analytic samples for secondary research questions 2-5 (participants who completed the 6-month follow-up but may have had missing timepoints) were estimated using data from this smaller sample of complete data. Comparisons between the models estimated using the two samples were made based on differences in magnitude of the intercept (baseline scores) and the slope (growth rate) parameter estimates for intervention and comparison groups.

The following figures show growth patterns (associated with secondary research questions 2-5 and selection of curvilinear growth model).

Figure F.1. Growth patterns on relationship satisfaction over the six-month study period. This was modeled for secondary research question two.

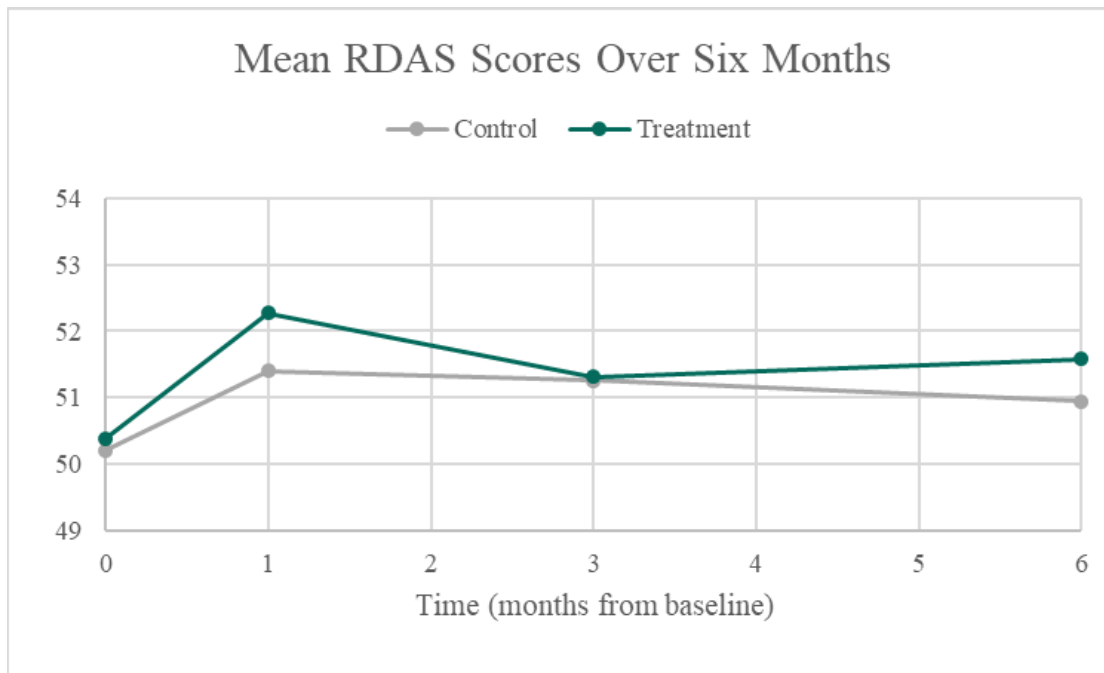


Figure F.2. Growth patterns on dyadic coping over the six-month study period. This was modeled for secondary research question three.

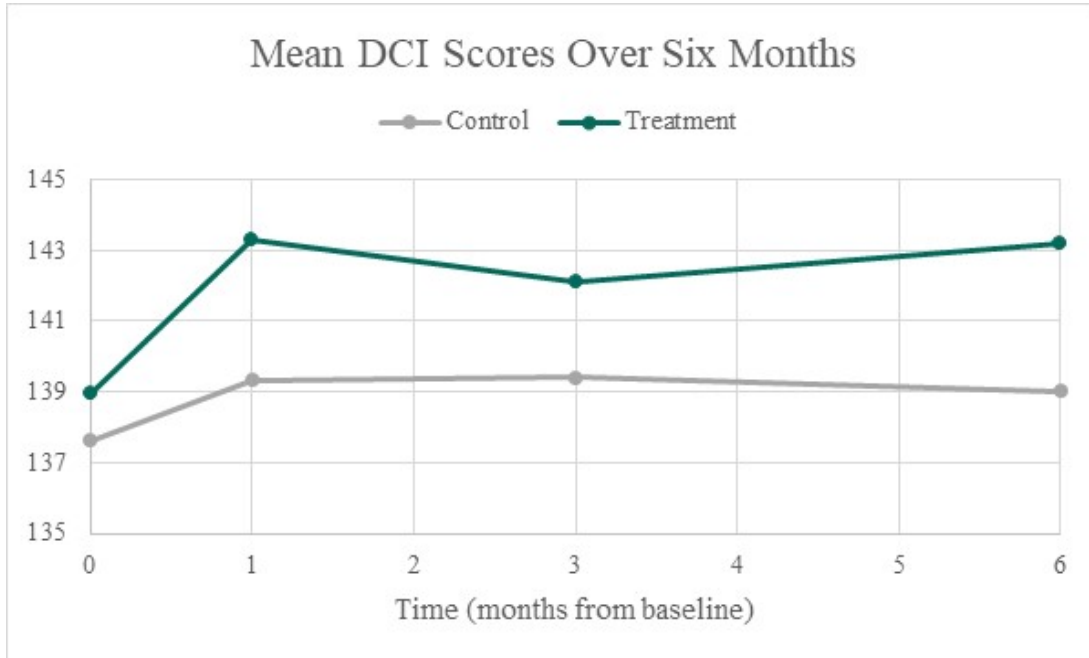


Figure F.3. Growth patterns on emotion regulation over the six-month study period. This was modeled for secondary research question four.

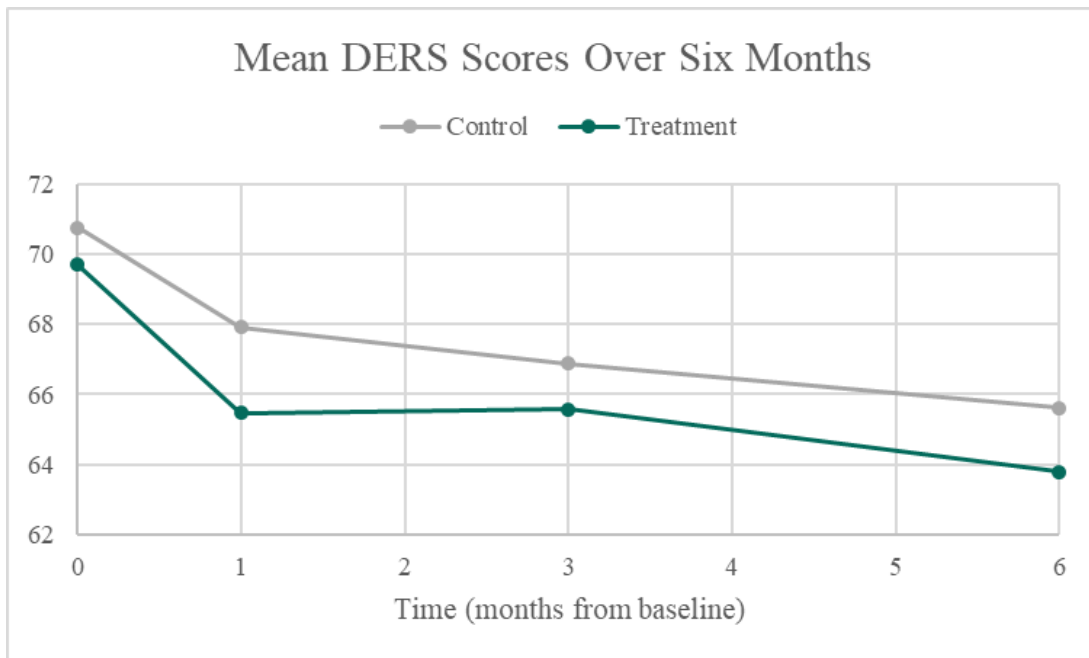


Figure F.4. Growth patterns on individual distress over the six-month study period. This was modeled for secondary research question five.

