



UNIVERSITY of  
DENVER



## Evaluation of Family Expectations in Oklahoma City

### Final Impact Evaluation Report for the Administration for Children and Families

November 17, 2020

Prepared by

Galena K. Rhoades, PhD, Scott M. Stanley, PhD,

Maggie O'Reilly Treter, MA,

Lane L. Ritchie, PhD, University of Denver

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:

Rhoades, G. K., Stanley, S. M., O'Reilly Treter, M., & Ritchie, L. L. (2020). *Evaluation of Family Expectations in Oklahoma City*. Denver, CO: University of Denver.

Acknowledgements:

We would like to thank the Administration for Children and Family's Office of Planning, Research, and Evaluation (OPRE) for their support of this project, as well as Mathematica for their guidance and consultation. We would also like to thank the staff at Public Strategies and the University of Denver for their collaboration in managing and executing the project.

Disclosure:

The authors of this report have no conflicts of interest.

This publication was prepared under Grant Number #90FM0064 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:** “Evaluation of Family Expectations in Oklahoma City”

**Objective.** Having a baby provides a window of opportunity for preventative interventions to help strengthen couples and families. This evaluation examines the impacts of Family Expectations—a 12-week relationship education program for couples that uses a 36-hour evidence-based curriculum (Becoming Parents Program)—on communication skills, destructive conflict, relationship stability, and depressive symptoms 12 months after enrollment. This evaluation represents an extension of previous federally funded randomized controlled trials of Family Expectations.

**Study design.** The current impact evaluation includes 1,320 committed couples (married and unmarried) recruited in Oklahoma City and surrounding areas who were pregnant at enrollment or gave birth to a child in the three months prior to enrollment. Couples were randomly assigned to either the Family Expectations program or to a no-treatment control group. Couples responded to surveys regarding communication skills, destructive conflict, depressive symptoms, and relationship stability at enrollment and 12 months later. The final sample analyzed in the current report included couples wherein at least one partner provided data on the outcomes of interest at baseline and the 12-month follow-up.

**Results.** Results indicate that there was a statistically significant difference between groups on destructive conflict, such that those who were assigned to Family Expectations reported lower levels of destructive conflict at the 12-month follow-up compared to those who were assigned to the control group. However, there were no statistically significant differences between Family Expectations and the control group in communication skills or depressive symptoms, or on the likelihood of being in a relationship together at the 12-month follow-up. Family Expectations was implemented with fidelity and the majority of participants received a high dosage of the program. Further, participants reported satisfaction with the program and their interactions with staff.

**Conclusion.** Couples were satisfied with Family Expectations and, in line with previous randomized controlled trials of this program, this impact evaluation provides further evidence that the program reduced their destructive conflict. Significant differences between Family Expectations and the control group did not emerge for the other outcomes tested, which showed impacts in prior studies. Overall, these findings demonstrate the robustness of this program for strengthening couple relationships by decreasing destructive conflict.

## Contents

I.	INTRODUCTION .....	1
A.	Introduction and study overview .....	1
B.	Primary research questions.....	2
C.	Additional research questions .....	2
II.	INTERVENTION AND COUNTERFACTUAL CONDITIONS .....	3
A.	Description of program as intended.....	3
B.	Description of counterfactual condition as intended.....	6
C.	Research questions about the intervention and counterfactual conditions as implemented .....	6
III.	STUDY DESIGN .....	7
A.	Sample formation and research design .....	7
B.	Data collection .....	8
IV.	ANALYSIS METHODS.....	10
A.	Analytic sample .....	10
B.	Outcome measures .....	11
C.	Sample characteristics and baseline equivalence .....	11
V.	FINDINGS AND ESTIMATION APPROACH .....	16
A.	Implementation evaluation .....	16
B.	Primary impact evaluation .....	18
C.	Sensitivity analyses.....	20
D.	Additional analyses .....	22
VI.	DISCUSSION.....	24
VII.	REFERENCES .....	26
VIII.	APPENDICES .....	28
A.	Data and study sample.....	28
B.	Data preparation .....	31
D.	Impact estimation .....	32
E.	Sensitivity analyses and alternative model specifications .....	35
F.	Additional analyses .....	37
G.	Additional tables of implementation findings.....	39

## 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.	Research questions about implementation of the intervention.....	6
IV.1.	Individual sample sizes by intervention group .....	10
IV.2.	Outcome measures used for primary impact analyses research questions .....	11
IV.3.a.	Summary statistics of <i>relationship stability</i> and baseline equivalence across study groups, for individuals completing 12-month follow-up.....	12
IV.3.b.	Summary statistics of <i>communication skills</i> and baseline equivalence across study groups, for individuals completing 12-month follow-up.....	13
IV.3.c.	Summary statistics of <i>destructive conflict</i> and baseline equivalence across study groups, for individuals completing 12-month follow-up .....	14
IV.3.d.	Summary statistics of <i>depressive symptoms</i> and baseline equivalence across study groups, for individuals completing 12-month follow-up.....	14
V.1.	Covariates included in impact analyses .....	18
V.2.	Post-intervention estimated effects on relationship stability using data from 12-month follow-up.....	19
V.3.	Post-intervention estimated mean differences on destructive conflict, communication skills, and depressive symptoms using data from 12-month follow-up .....	20
V.4.	Covariates included in sensitivity analyses.....	20
V.5.	Differences in means between Family Expectations and control groups estimated using alternative methods .....	21
V.6.	Post-intervention estimated effects using data from 12-month follow-up – sensitivity analyses using only participants who remained together .....	23
V.7.	Post-intervention estimated effects on the conflict-stability index using data from 12-month follow-up .....	23
A.1.	Key features of the implementation analysis data collection.....	28
A.2.	Key features of the impact analysis data collection .....	29
D.1.	Post-intervention estimated effects on destructive conflict, communication skills, and depressive symptoms using data from 12-month follow-up .....	35

E.1.	Post-intervention estimated effects using data from 12-month follow-up – sensitivity analyses with additional covariates – multilevel models .....	36
G.1.	Individual item responses regarding the <i>quality of</i> staff-participant interactions during Family Expectations .....	39
G.2.	Individual item responses regarding program feedback .....	40

## Figures

A.3. Family Expectations CONSORT diagram .....	30
--	----

# Impact Evaluation of Family Expectations In Oklahoma City

## I. INTRODUCTION

### A. Introduction and study overview

A substantial body of research confirms that children typically live healthier, longer lives under better economic conditions when raised in stable home environments with two actively engaged parents. The mere physical presence of two parents, however, is not sufficient to confer the fullest benefits to children; it is also vital that the interactions between parents are healthy, stable, and cooperative. While research has shown that becoming a parent has a negative effect on marital satisfaction (Twenge et al., 2003), this life transition is also considered a “teachable moment” in the couple relationship (Markman et al., 1986), providing a window of opportunity for preventative interventions to help strengthen couples and families.

Beginning around 2003, the federal government, through the Administration for Children and Families (ACF) began to fund community-based, relationship education services to be provided to couples, individuals, and families generally, and specifically, to those with socio-demographic disadvantages. As part of this funding, two large, federal, multi-site trials evaluated services provided to economically disadvantaged unmarried expectant couples (The Building Strong Families Study: BSF), or to economically disadvantaged married couples (The Supporting Healthy Marriage Study: SHM). These two studies generally concluded that there were non-significant impacts (BSF; Wood et al., 2012) or modest but significant and sustained effects (SHM; Lundquist et al., 2014) on relationship quality, psychological distress, and coparenting and parenting. However, the Family Expectations program in Oklahoma City was the only site that showed positive effects on partners’ relationship quality at 15 months and family stability at 15 and 36 months (Devaney & Dion, 2010).

Family Expectations is a 12-week program that uses a 36-hour evidence-based curriculum (Becoming Parents Program) developed by Pam Jordan that includes relationship education (adapted from the Prevention and Relationship Education Program [PREP]; Markman, Stanley, Blumberg, Jenkins, & Whiteley, 2001) and content on caring for a newborn (Frei & Jordan, 2016).

This impact evaluation represents an extension of the aforementioned federally funded randomized controlled trials (RCTs) of Family Expectations. Since the BSF and SHM studies were conducted, a few modifications were made to Family Expectations, including a reduction in case management from 12-15 meetings to meetings as needed in the current version, and an increase in curriculum content to 36 hours (from 30 hours). The current report assesses implementation of the modified Family Expectations intervention and presents program impacts



on the following outcomes: communication skills, destructive conflict, relationship stability, and depressive symptoms compared to a control group 12 months after enrollment.

Below, we describe our primary and additional research questions. In the following five sections we describe (II.) the intervention and comparison conditions, (III.) our study design, (IV.) analysis methods, (V.) implementation and impact findings and estimation approach, and (VI.) a discussion of our program implementation and impact findings.

## B. Primary research questions

This section presents research questions that assessed the impact of Family Expectations on key outcomes 12 months after enrollment compared to a no-treatment control group.

1. Compared to the control group, does being randomly assigned to participate in Family Expectations improve couples' **relationship stability**?
2. Compared to a control group, does being randomly assigned to participate in Family Expectations lead to better **communication skills**?
3. Compared to a control group, does being randomly assigned to participate in Family Expectations lead to lower levels of **destructive conflict**?
4. Compared to a control group, does being randomly assigned to participate in Family Expectations lead to lower levels of **depressive symptoms**?

This study is registered with [clinicaltrials.gov](https://clinicaltrials.gov/ct2/show/study/NCT03058549) (NCT03058549).

## C. Additional research questions

1. Compared to a control group, what is the impact of Family Expectations on **destructive conflict, communication skills, and depressive symptoms** 12 months after enrollment among those couples who remained together by the end of the study period?
2. Compared to a control group, are couples randomly assigned to participate in Family Expectations more likely to still be in a relationship together characterized as not having high levels of conflict?

## II. INTERVENTION AND COUNTERFACTUAL CONDITIONS

This section outlines the intended program and counterfactual conditions for the Family Expectations impact evaluation, including the program content, planned dosage and implementation schedule, and the education and training of program staff. The program condition was made up of couples assigned to Family Expectations who received the Becoming Parents Program curriculum and case management as needed. The counterfactual condition was a control group wherein participants did not receive the Becoming Parents Program curriculum or case management services, though they were able to seek and obtain any services they wished in the community.

### A. Description of program as intended

The Family Expectations program model includes workshops on relationship skills and caring for a newborn as well as case management (see Table II.1). The target population included committed couples, married or unmarried, who are pregnant or had a baby in the past 3 months.

#### 1. Key components of the program model

Family Expectations utilizes the evidence-based Becoming Parents Program curriculum. Becoming Parents Program uses a mix of interactive and didactic learning models, including couple activities, group activities, videos, and workbook assignments. Core lessons included learning communication skills; resolving and managing conflict and problem solving; learning the benefits of having a healthy marriage; and handling stress and managing anger. Additionally, program incentives were offered as couples reached participation benchmarks, thereby encouraging couples to complete the sessions.

The multi-faceted Family Expectations program model addresses variables that affect a couple's ability to establish and maintain healthy relationships and families: 1) marriage and relationship education to help married, engaged, and unmarried couples build healthy and enduring partnerships; 2) communication, interpersonal, and conflict resolution skill development that is useful in romantic relationships, with family members, and in the workforce; 3) parenting skills to ensure safe, nurturing environments for children, particularly newborns; 4) economic stability and mobility activities to enhance participants' employability skills, help them secure job opportunities, and improve their financial literacy; 5) connection to local resources and services including mental health and substance abuse assistance; and 6) formation of social support networks and positive relationships that improve the multiple relational and emotional dimensions of socioeconomic disadvantage.

In addition, couples enrolled in Family Expectations received case management from Family Support Navigators (FSNs). FSNs supported participant attendance and engagement through reminder phone calls and text messages; coordinated supportive assistance such as transportation and childcare to help couples overcome barriers to attendance; provided access and referrals to community resources, such as employment development services and mental health/substance

abuse agencies; and reinforced curriculum content and skills through interactions that support couple's personal and family goals. See Table II.2 for more information on education and training of staff.

## **2. Co-located employment stability services**

Family Expectations participants who were under- or unemployed also had access to co-located job training and career advancement services based on their needs. Although not a core component of Family Expectations, It's My Community Initiative (IMCI), a grant partner that operates a comprehensive employment services program, WorkReady Oklahoma, is co-located with Public Strategies, offering a one-stop integration and easy access to optional services for participants. These services included: 1) employment assessments, which identify interests and abilities; 2) individualized career planning, including content focused on establishing goals and identifying appropriate education/training opportunities; 3) intensive coaching and short-term specialized services to remove external barriers not related to education/training, including resume review and refinement, interview coaching, and social media/LinkedIn profile creation and refinement; 4) education and career training; and 5) job placement, including access to a variety of employer partners representing in-demand occupations. Participants in this study did not participate in IMCI's fatherhood program, TRUE Dads, or the related evaluation of the TRUE Dads program.

## **3. Intake and schedule of program participation**

During Family Expectations' initial contact with a couple, information was gathered to determine both program eligibility and their availability to participate in the program. Eligible couples were then scheduled for an intake appointment for the study at the Family Expectations site. After random assignment, those assigned to the intervention track were invited to an office visit with an FSN that occurred, typically, within two weeks of the initial intake appointment. During the visit, the FSN provided an overview of what to expect throughout the program, completed a needs assessment, built rapport, and facilitated connections to resources both at Public Strategies and in the community that help stabilize home and family dynamics. After completing the intake process, the couples' assigned FSN remained the couples' primary point of contact for the duration of their participation in Family Expectations. All Family Expectations services were delivered on-site at Public Strategies' facilities in Oklahoma County.

Family Expectations workshops were facilitated by two-to-three Educators in weekly 180-minute classes/sessions totaling 36 hours of curriculum delivered over the course of 12 weeks.

Weeknight and weekend workshop options were available to accommodate various schedules, with new workshop series beginning nearly every two weeks. Due to the various life circumstances that couples in the target population faced (for example, family emergencies, loss of job, birth of a baby, or bedrest), participants were given the option to make up missed content by joining another workshop series; additionally, couples were eligible to receive program services, including make-up sessions, for up to one year from their date of intake.

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

Component	Curriculum and content	Dosage and schedule	Delivery	Target Population
<b>Intervention</b>				
Parenting and Relationship Skills Workshops	Becoming Parents Program curriculum: Relationship and co-parenting education, and caring for a newborn	36 hours, with 3-hour sessions occurring once per week for 12 weeks	Group lessons provided at the intervention's facility by 2-3 trained facilitators in every session	Committed couples (married or unmarried) who are pregnant or had a baby in the past 3 months
Case Management/Family Support	Support around program attendance/completion, reinforce curriculum content, facilitate connections to community resources (including WorkReady services)	Initial office visit; additional support as needed	In-person meetings	All couples assigned to Family Expectations
<b>Counterfactual</b>				
No-Treatment control Group	Couples did not receive the intervention. They did not receive referrals but were able to seek and obtain services in the community.	n.a.	n.a.	Couples assigned to the control group

n.a. = not applicable

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

Component	Education and initial training of staff	Ongoing training of staff
<b>Intervention</b>		
Parenting and Relationship Skills Workshops	Educators were male and female and held at least a bachelor's degree and received 30 hours of initial training with the curriculum developers.	Educators received on-going booster sessions, continuing education, and ad-hoc feedback based on observations from staff and master trainers.
Case Management/Family Support	Family Support Navigators were male and female and held at least a bachelor's degree and completed initial training in case management.	FSNs participated in curriculum-based information and role-playing sessions, attended weekly planning meetings, completed annual domestic violence awareness and prevention training, and received training from Oklahoma Department of Human Services on preventing and detecting child maltreatment.
<b>Counterfactual</b>		
None	n.a.	n.a.

n.a. = not applicable

## B. Description of counterfactual condition as intended

Couples assigned to the control group received no intervention services from Family Expectations or Public Strategies, nor referral assistance to other community resources that may have been available (see Table II.1). However, they were not excluded from pursuing such services, including employment services that were available from WorkReady located in the same building as the Family Expectations program. Although control group couples were not referred to WorkReady, and these services were not advertised in the building, they may have learned about these services on their own or from other community agencies.

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

Table II.3 outlines research questions regarding the implementation of Family Expectations as experienced by participants in this study, including fidelity to the program, program dosage received by participants, quality of staff-participant interactions, participants' feedback about the program, and services utilized by the Family Expectations program group in addition to the intervention. These findings provide context and aid interpretation of the impact analysis findings for couples assigned to Family Expectations.

**Table II.3. Research questions about implementation of the intervention**

Implementation element	Research question
Fidelity	<ul style="list-style-type: none"> <li>Did couples in the Family Expectations group have access to the full set of workshop sessions offered?</li> <li>Were there adaptations to the workshops over the course of the study?</li> </ul>
Dosage	<ul style="list-style-type: none"> <li>What percentage of couples in the Family Expectations group completed the workshops?</li> <li>What was the average attendance of couples (number of workshops) in the Family Expectations group?</li> <li>What percentage of couples in the Family Expectations group received case management, and what was the average dosage?</li> </ul>
Quality	<ul style="list-style-type: none"> <li>What was the quality of staff—participant interactions during the intervention (i.e., workshops and individual service contacts)?</li> </ul>
Engagement	<ul style="list-style-type: none"> <li>What was intervention couples' overall feedback on the program?</li> </ul>
Context	<ul style="list-style-type: none"> <li>What outside services were utilized by couples in the Family Expectations group during the evaluation period?</li> </ul>

### III. STUDY DESIGN

This section provides an overview of the study design, sample, and data collection. The Family Expectations impact evaluation included 1,320 committed couples (married or unmarried) who were either pregnant or had a baby in the past three months. Couples were randomly assigned to the Family Expectations program or the control group and responded to a survey regarding communication skills, destructive conflict, depressive symptoms, and relationship stability 12 months after enrollment.

#### A. Sample formation and research design

The University of Denver Institutional Review Board approved the study design and data collection plans on May 4<sup>th</sup>, 2016 (DU IRB Protocol #: 860135-1). A continuing review occurred on May 24, 2017. Minor amendments to the IRB protocol were routinely submitted and approved related to various changes and additions to the data collection and for changes in research assistants. Study enrollment began in June 2017. Baseline data collection occurred from June 2017 to May 2019 and 12-month follow-up data collection occurred from June 2018 to June 2020.

##### 1. Recruitment and eligibility

Family Expectations' primary source of recruitment for the study was word-of-mouth recommendations from friends, families, and acquaintances, including from current or past program participants. In addition, program staff cultivated referral relationships with local OB/GYN offices and hospitals, childcare centers, social services agencies, probation and parole offices, and community agencies (like Infant Crises Services) who also serve Family Expectations' target demographic of low-income, pregnant couples.

Couples were recruited from Oklahoma County and surrounding areas based on the following inclusion criteria for the study:

- Both parents were willing to participate and were 18 years old or older.
- Parents have a commitment to be together as a couple.
- Both parents are the biological parents of a child in a current pregnancy or of a child born within the past 3 months.
- Both parents are fluent in English.
- Neither parent has taken part in the Family Expectations program before or is related to a Family Expectations employee.

##### 2. Consent process

At the couple's intake appointment, an Intake Specialist explained the study, the voluntary nature of participation, and how and when their data would be collected. The Intake Specialist was

aided by a video developed by Public Strategies to help explain the study and the random assignment process. The specialist clearly explained that there were two study tracks, one that involved participating in Family Expectations and one that involved participating in a control group, and that group assignment would be made by random assignment. Participants were asked to give consent to participation without knowing to which track they would be assigned. Prior to consenting, couples were asked if they wanted to proceed with program enrollment and participate in the study together. If both partners wished to proceed, each partner in the couple completed their own consent form on paper, and either partner could decide not to consent. Either partner choosing not to consent resulted in exclusion from the study.

### **3. Random assignment process**

Couples were randomly assigned to participate in the Family Expectations program or a control group after they consented to participate in the evaluation and completed baseline data collection. Random assignment was completed using a computerized procedure that was built into the Family Expectations management information system (MIS) for this study. Random assignment was made in blocks of 50 couples, with 60% probability of assignment to the intervention group and 40% to the control. This procedure produced allocations in line with the target probability. After baseline data were entered in the MIS system, and participants completed baseline surveys, a Family Expectations staff member completed random assignment in the MIS and the results were sent to a supervisor. Those assigned to the intervention received a phone call within 24 hours to schedule their participation in Family Expectations. Those assigned to the control condition were contacted within a few days by a research specialist by phone, letter, or email.

After random assignment, couples in the Family Expectations program group chose which upcoming workshop series they wanted to attend based on their scheduling preferences. Each series met weekly for 12 weeks to receive the Becoming Parents Program curriculum as a group. The discontinuation of study participation by one partner (either by not completing the 12-month follow-up survey or by withdrawing from the study altogether) had no bearing on the ability of the other partner to participate in the 12-month follow-up.

## **B. Data collection**

This section outlines the Family Expectations impact evaluation data collection procedures, first discussing the implementation evaluation and then the impact evaluation.

### **1. Implementation analysis**

All data were collected via internal staff notes; case notes entered into the MIS; participant survey items collected using Qualtrics survey software; and the federally required Information, Family Outcomes, Reporting, and Management survey (nFORM). See Table B.1 in the Appendix for more information regarding data collection for the implementation analysis.

## 2. Impact analysis

There were two survey platforms used to collect participant data in the study: one required by all federal grantees called Information, Family Outcomes, Reporting, and Management (nFORM) and one developed for the evaluation using Qualtrics survey software. Using these two platforms, program staff collected surveys at baseline, in a post-intervention assessment of the intervention group, and at 12-months after enrollment. See Table B.2 in the Appendix for more information regarding data collection for the impact analysis.

### *Baseline assessment.*

Participants completed surveys at baseline administered by program staff using Qualtrics and nFORM in a private setting on tablets or computers. All participants, regardless of study condition, completed these baseline surveys. First, after the initial consent process, participants completed the nFORM Applicant Characteristics survey. Next participants completed a Qualtrics survey. Lastly, participants completed the pre-test survey in nFORM. Participants were paid \$50 for completing these surveys.

### *Post-intervention assessment (intervention group only)*

The nFORM post-test survey was delivered by program staff at the final intervention session for the intervention group. This was only completed by those that attended this session. Thus, this post-test was not collected from program dropouts or non-completers and these data were not included in the impact evaluation. Participants were not paid for completing this assessment.

### *12-month follow-up assessment*

Participants were contacted by program staff 12 months after enrollment (when they completed their baseline assessment) and invited to return to Public Strategies to complete the follow-up survey. Program staff used calls, emails, texts, and, on occasion, field locators to reach participants. If participants indicated that they could not come in, they were sent a link to complete the Qualtrics survey electronically. Participants were paid \$50 for completing this assessment.



## IV. ANALYSIS METHODS

The following section describes the analytic sample, study attrition, and measures for each outcome of interest (relationship stability, communication skills, destructive conflict, and depressive symptoms). In addition, we present an assessment of baseline equivalence of the Family Expectations program and control groups across several demographic and relationship characteristics.

### A. Analytic sample

The analytic sample used to estimate impacts of the intervention is the sample of individuals for whom one or both partners in a couple provided data on the outcome measures at both baseline and the 12-month follow-up (see Table IV.1). Based on the What Works Clearinghouse cautious attrition standards, Family Expectations is a low-attrition RCT based on the rate of overall and differential attrition at the cluster (couple) and individual level. The overall rate of couple attrition was 10.2% with differential attrition of 0.2 percentage points. The rate of individual attrition for each outcome measure was between 6.0 and 20.9 percent, with rates of differential attrition between 0.2 and 1.2 percentage points.

**Table IV.1. Individual sample sizes by intervention group**

Number of individuals	FE sample size	control sample size	Total sample size	Total response rate	FE response rate	control response rate
<b>Clusters</b>						
Clusters: Assigned to condition	786	534	1,320	n.a.	n.a.	n.a.
Clusters: Contributed at least one individual at baseline	786	534	1,320	100%	100%	100%
Clusters: Contributed at least one individual at follow-up survey (12 months after baseline)	707	479	1,186	89.8%	89.9%	89.7%
<b>Individuals in non-attributing clusters</b>						
Assigned to condition	1,414	958	2,372	n.a.	n.a.	n.a.
Contributed a baseline survey	1,414	958	2,372	100%	100%	100%
Contributed to follow-up survey (12 months after baseline)	1,326	904	2,230	94.0%	93.8%	94.4%
Relationship Stability (12 months after baseline)	1,326	904	2,230	94.0%	93.8%	94.4%
Communication Skills (12 months after baseline)	1,126	751	1,877	79.1%	79.6%	78.4%
Destructive Conflict (12 months after baseline)	1,279	864	2,143	90.3%	90.4%	90.2%
Depressive Symptoms (12 months after baseline)	1,311	883	2,194	92.5%	92.7%	92.2%

n.a. = not applicable

Note: Sample sizes for each outcome account for item non-response.

## B. Outcome measures

Table IV.2 describes the outcome measures used for the impact analyses. Relationship stability is measured at the couple-level wherein if either partner reports not being in the same relationship at the 12-month follow-up with the partner they enrolled in the study with, it is considered a breakup. Communication skills, destructive conflict, and depressive symptoms are measured at the individual level using measures that have acceptable reliability based on published sources.

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

Outcome measure	Description of the outcome measure	Source	Timing of measure
Relationship Stability	This outcome is a single dichotomous item that measures whether partners report still being in the same relationship on a steady basis or not at follow-up.	Building Strong Families (Wood, Moore, Clarkwest, & Killewald, 2014)	12-Month Follow-up
Communication Skills	This outcome measures communication skills in a relationship using the average of 13 survey items (value range 1 to 7); $\alpha = .85$ (Stanley et al., 2014). Includes items such as, "When our discussions begin to get out of hand, we agree to stop them and talk later."	Communication Skills Test (Saiz & Jenkins, 1995)	Baseline and 12-Month Follow-up
Destructive Conflict	This outcome measures destructive conflict and communication in a relationship using the average of 9 survey items (value range 0 to 3); $\alpha = .87$ (Wood et al., 2014). Includes items such as, "When we argue, one of us is going to say something we will regret."	Building Strong Families (Wood, Moore, Clarkwest, & Killewald, 2014)	Baseline and 12-Month Follow-up
Depressive Symptoms	This outcome is a brief depression inventory using the adjusted sum (the sum multiplied by 1.67) in order to compare to clinical cutoffs for the 20-item CES-D (value range 0 to 3). $\alpha = .92$ for men and .89 for women (Wood, Moore, Clarkwest, Killewald, & Monahan, 2012). Includes items such as, "During the past week...I felt that everything I did was an effort."	Center for Epidemiological Studies Depression Scale (CES-D; Radloff, 1977)	Baseline and 12-Month Follow-up

## C. Sample characteristics and baseline equivalence

### 1. Sample characteristics.

Based on the entire sample of participants who completed baseline, in terms of race, 51% of participants identified as White, 34% as Black/African American, 12% as Native American, 2% as Asian, 1% as Native Hawaiian or other Pacific Islander, and 13% as another race. Participants selected all racial categories they identified with, thus racial categories endorsed sum to greater than 100%. Nineteen percent identified as Hispanic/Latino/a.

All couples reported being romantically involved at the time of enrollment, 86% reported living together, and 29% reported being married. Thirty-one percent had another child together beside the pregnancy/baby. The mean relationship length was 3.25 years ( $SD = 2.92$ ). The median combined income was \$25,000-29,999.

## 2. Baseline equivalence.

The Family Expectations program group and the control group were compared at baseline on each outcome variable, as well as key demographic and relationship characteristics, using a cutoff of  $p = .05$  to determine group equivalence. This was repeated for the analytic sample used for each outcome measure (see Tables IV.3.a through IV.3.d). Tests of continuous variables at baseline were conducted by comparing mean scores between groups within multilevel models that account for the non-independence of partner scores (in which partners in a couple can have different values, for example age) or by  $t$ -tests (in which each couple has one value, for example, relationship duration). Similarly, tests of dichotomous variables at baseline were conducted using logistic regression in multilevel models (where each partner can have different values, for example high school graduate or not) or by regular logistic regression (where each couple has one value, for example married or not). Effect sizes were calculated using Hedges  $g$  for continuous variables and the Cox index for dichotomous variables. Details of the statistical tests are provided in Appendix E.

At baseline, there was a statistically significant difference in participant age between the Family Expectations group and the control group ( $p < .01$ ). The effect size difference was between .15 and .17 in each analytic sample. Therefore, age was controlled for in the primary impact analyses. No other covariates, and no outcome measures, demonstrated statistically significant baseline differences between the groups.

**Table IV.3.a. Summary statistics of *relationship stability* and baseline equivalence across study groups, for individuals completing 12-month follow-up**

Baseline measure	Family Expectations M (SD)	Control M (SD)	Family Expectations vs. control p-value
<b>Men Only</b>			
Age	27.8 (6.8)	26.7 (6.1)	.008
Hispanic	20%	19%	.589
Native American	12%	10%	.332
Black/African-American	36%	38%	.917
White	33%	35%	.542
Other race	5%	3%	.308
Employed full time	57%	55%	.522
High school diploma	68%	70%	.386
<b>Women Only</b>			
Age	25.5 (5.2)	24.7 (4.9)	.008
Hispanic	18%	18%	.822
Native American	12%	12%	.908
Black/African-American	30%	32%	.924
White	41%	41%	.987
Other race	4%	3%	.405
Employed full time	28%	30%	.524

Baseline measure	Family Expectations M (SD)	Control M (SD)	Family Expectations vs. control p-value
High school diploma	77%	80%	.173
<b>Men and Women (Couple-Level)</b>			
Married	32%	29%	.321
Living together	86%	86%	.746
Prior child together	33%	31%	.477
Combined Income (Mdn)	\$25,000-29,999	\$20,000-24,999	.272
Relationship length (months)	40.1 (35.8)	38.9 (34.1)	.554
Sample size	707 couples (1,326 individuals)	479 couples (904 individuals)	n.a

n.a. = not applicable.

Notes: For testing baseline imbalances on stability, individual covariates were analyzed separately for men and women given that analyses of stability included (where indicated) scores for both partners. Income was measured by categories and analyzed as a continuous variable (see Table V.5). Race/ethnicity variables sum to more than 100% because participants could endorse more than one category.

**Table IV.3.b. Summary statistics of *communication skills* and baseline equivalence across study groups, for individuals completing 12-month follow-up**

Baseline measure	Family Expectations M (SD)	Control M (SD)	Family Expectations vs. control p-value
Age	27.0 (6.3)	26.0 (5.7)	.004
Hispanic	20%	19%	.513
Native American	12%	11%	.656
Black/African-American	30%	33%	.303
White	39%	39%	.767
Other race	4%	3%	.249
Employed full time	44%	45%	.618
High school diploma	74%	77%	.197
Married	35%	31%	.241
Living together	88%	87%	.885
Prior child together	31%	30%	.433
Combined Income (Mdn)	\$25,000-29,999	\$25,000-29,999	.551
Relationship length (months)	42.2 (36.8)	40.9 (34.5)	.547
Communication skills (Range: 1 to 7)	4.6 (1.4)	4.6 (1.3)	.874
Sample size	617 couples (1126 individuals)	420 couples (751 individuals)	n.a.

n.a. = not applicable.

Notes: Income was measured by categories and analyzed as a continuous variable (see Table V.5). Race/ethnicity variables sum to more than 100% because participants could endorse more than one category.

**Table IV.3.c. Summary statistics of *destructive conflict* and baseline equivalence across study groups, for individuals completing 12-month follow-up**

Baseline measure	Family Expectations M (SD)	Control M (SD)	Family Expectations vs. control p-value
Age	26.8 (6.1)	25.9 (5.7)	.006
Hispanic	19%	19%	.724
Native American	12%	11%	.609
Black/African-American	32%	35%	.343
White	37%	38%	.846
Other race	4%	3%	.338
Employed full time	42%	43%	.764
High school diploma	73%	76%	.222
Married	33%	29%	.273
Living together	86%	86%	.955
Prior child together	31%	30%	.743
Combined Income ( <i>Mdn</i> )	\$25,000-29,999	\$25,000-29,999	.402
Relationship length (months)	40.6 (36.1)	39.7 (34.4)	.684
Destructive conflict (range: 0 to 3)	1.3 (0.8)	1.3 (0.8)	.475
Sample size	680 couples (1,279 individuals)	462 couples (864 individuals)	n.a.

n.a. = not applicable.

Notes: Income was measured by categories and analyzed as a continuous variable (see Table V.5). Race/ethnicity variables sum to more than 100% because participants could endorse more than one category.

**Table IV.3.d. Summary statistics of *depressive symptoms* and baseline equivalence across study groups, for individuals completing 12-month follow-up**

Baseline measure	Family Expectations M (SD)	Control M (SD)	Family Expectations vs. control p-value
Age	26.8 (7.0)	25.7 (5.6)	.002
Hispanic	19%	18%	.718
Native American	13%	11%	.257
Black/African-American	33%	34%	.449
White	37%	38%	.590
Other race	4%	3%	.273
Employed full time	42%	42%	.921
High school diploma	73%	76%	.112
Married	32%	29%	.319
Living together	85%	86%	.793
Prior child together	31%	29%	.514

Baseline measure	Family Expectations M (SD)	Control M (SD)	Family Expectations vs. control p-value
Combined Income ( <i>Mdn</i> )	\$25,000-29,999	\$25,000-29,999	.392
Relationship length (months)	40.1 (35.8)	38.5 (33.0)	.438
Depressive symptoms (range: 0 to 60)	15.5 (11.7)	15.4 (11.7)	.837
Sample size	705 couples (1,311 individuals)	474 couples (883 individuals)	n.a.

n.a. = not applicable.

Notes: Income was measured by categories and analyzed as a continuous variable (see Table V.5). Race/ethnicity variables sum to more than 100% because participants could endorse more than one category.

## V. FINDINGS AND ESTIMATION APPROACH

### A. Implementation evaluation

#### Key Findings:

Results demonstrate that the program was implemented with fidelity and participants received an adequate dosage of program content and case management services. The complete program (12 sessions) was offered to all cohorts. Eighty-nine percent (89%) of participants attended at least one program session. Including those not attending any sessions, those assigned to Family Expectations completed an average of 7.5 program sessions, receiving an average of 22.4 hours of curriculum content (out of 36 hours of content). Forty-six (46%) percent of those assigned to Family Expectations completed all 12 sessions. Participants reported high-quality staff-participant interactions during the intervention and overall positive feedback about the program. Finally, 75% of participants in the Family Expectations group received services from other programs, in addition to Family Expectations, during their time in the program. Two recruitment strategies that resulted in the largest share of participants starting Phase II were Residential Center A, making up 30.6% of 434 participants who started Phase II workshops, followed by Residential Center B, contributing 26.7% of 434 participants who started Phase II. The third largest contributor was family/friends/walk-in/word-of-mouth, contributing 11.8% of 434 participants. The remaining 30.9% of participants were recruited from five additional recruitment strategies, for a total of eight different recruitment strategies. This highlights the importance of multiple recruitment strategies in order to meet target enrollments.

#### 1. Fidelity

The program was implemented with fidelity, meaning that the full set of 12 parenting and relationship skills workshops were offered to each cohort. Adaptations to the program occurred in late April 2019 with other changes in November 2019. These changes would not likely have impacted the findings in this report, as the changes in late April occurred with barely more than one month left in recruitment for the study, thus only affecting only couples entering the study in the last month of random assignment. Further, the changes in November 2019 came after most couples would have completed the program. Nevertheless, the changes made were as follows: To provide more time for participant engagement and discussion, Family Expectations worked in partnership with the Becoming Parents Program curriculum developer to revise content that was repetitive in nature or that the participants found confusing. Several of the activities were reconfigured to get the participants out of their seats and moving around, and videos were added to make this a more interactive experience and less didactic in nature. Additionally, better descriptions of the concepts were added to anchor discussions after key points in the curriculum. The program facilitators also placed additional focus on creating a conversational style of atmosphere in the classroom by utilizing co-facilitation skills.

## 2. Dosage

Regarding dosage, eighty-nine percent (89%) of the Family Expectations sample attended at least one program session. Including those not attending any sessions, those assigned to Family Expectations completed an average of 7.5 program sessions, receiving an average of 22.4 hours of content. Forty-six (46%) percent of those assigned to Family Expectations completed the entire program consisting of 12 sessions, receiving all 36 hours of content. Of those ever attending any workshop, the average number of workshop sessions attended was 8.4 out of 12 (receiving 25 hours of content, on average). The majority of these participants (52%) attended all 12 workshop sessions (36 hours).

By program design, couples received their first case management meeting prior to attending workshops. Based on need and interest, couples could attend additional meetings over the course of their involvement in Family Expectations. Of those assigned to Family Expectations, 91.3% received at least one case management visit (8.7% received no office visits), 23.5% received two visits, and 17.7% received three or more visits. The average participant received 1.7 visits (including those who received no visits).

## 3. Quality

Participants reported high-quality staff-participant interactions during the intervention, as captured by participant responses to a seven-item post-program survey (rated from 1 = Strongly disagree to 7 = Strongly agree;  $M$  working alliance = 6.37,  $SD$  = 1.07). At least 89% of participants *somewhat agreed* to *strongly agreed* that staff-participant interaction was positive on each item. Specific item responses are outlined in the Appendix in Table H.1.

## 4. Engagement

Participants reported overall positive feedback regarding Family Expectations on measures of program satisfaction, engagement, and knowledge gained. The majority of participants agreed that the program helped them learn to communicate and manage conflict in their relationship, work as a team, and invest in their relationship. In addition, participants endorsed feeling more supported in their relationship and feeling more equipped to care for their baby. See Table H.2 in the Appendix for information regarding specific item responses.

## 5. Context

The majority of participants in the Family Expectations group (75%) received outside services in addition to services provided as part of Family Expectations. Fourteen percent (14%) reported attending classes, workshops, or group sessions for relationship help; and 13% reported meeting with a social worker, counselor, or clergy member for relationship help. In addition, 71% of participants reported receiving outside services unrelated to relationship help, including programs for learning parenting skills (25%), acquiring a job (15%), help with mental health, alcohol or substance use (10%), infant crisis services (23%), and SNAP employment and training (31%).



### Key limitations of the implementation data:

Despite having quantitative reports of program fidelity (for example, number of sessions delivered), a key limitation of this impact evaluation was the lack of qualitative fidelity data. Without direct observation of program delivery and fidelity, it is difficult to fully characterize the fidelity in *how* the program was delivered to participants. However, facilitators received 30 hours of in-person core curriculum training and then ongoing boosters/continuing education and observation feedback from staff and master trainers on an ad-hoc basis. Thus, while we do not have a full picture of program delivery, the ongoing education and support offered to facilitators supported their ability to offer the program with fidelity.

## B. Primary impact evaluation

### Key Findings:

Participants assigned to Family Expectations had significantly lower levels of destructive conflict 12 months after enrollment compared to the control group. However, the Family Expectations and control groups did not significantly differ in relationship stability, communication skills, or depressive symptoms 12 months after enrollment.

An intent-to-treat (ITT) framework was used in all impact analyses, such that each couple was analyzed based on the group to which they were randomly assigned, regardless of uptake or actual services received. All analyses used the standard threshold of  $p < .05$ , two-tailed test, to evaluate statistical significance. For each outcome variable, the analytic sample consisted of individuals who did not have any missing data for any of the variables included in the estimation model.

The primary analyses evaluated the impact of the intervention on the four primary outcome measures: relationship stability, communication skills, destructive conflict, and depressive symptoms. Age and the baseline measure of the outcome was controlled for in the analysis of primary outcomes (Table V.1).

**Table V.1. Covariates included in impact analyses**

Covariate	Description of the covariate
Age	Age (in years) as of the baseline data collection
Outcome value at baseline	Mean levels of <i>Destructive Conflict</i> , <i>Communication Skills</i> , or <i>Depressive Symptoms</i> as of baseline data collection ( <i>Relationship Stability</i> does not include a baseline outcome covariate)

Notes: The outcome value at baseline was only included as a covariate in the impact analysis of that outcome.

### 1. Relationship stability

At the 12-month follow-up period, couples were coded as stable in their relationship if both partners reported that they were still together, in a steady, romantically involved relationship. A couple was coded as not stable if either or both partners reported having broken up or that the relationship was an “on-again-off-again” relationship. For couples in which only one partner

responded at follow-up, the answer of the responding partner was coded as the stability outcome for the dyad. Thus, every couple for which at least one partner responded to the follow-up and indicated relationship status was entered in these analyses and stability was thus evaluated as a couple-level outcome. In order to evaluate the impact on relationship stability, Ordinary Least Squares (OLS) regression analyses were conducted with intervention (Family Expectations vs. control) as the independent variable and relationship stability as the outcome variable, controlling for age. See Appendix E for more information on the analytic approach.

Results indicated that those assigned to Family Expectations were not statistically significantly more likely to still be in a relationship together at the 12-month follow-up than those who were assigned to the control group ( $p = 0.38$ ; see Table V.2).

**Table V.2. Post-intervention estimated effects on relationship stability using data from 12-month follow-up**

Outcome measure	Family Expectations %	No-Treatment control %	FE versus control % difference (p-value of difference)	Effect size
Relationship Stability	70.9%	68.5%	2.4 (.378)	.07
Sample Size	707	479	n.a.	n.a.

n.a. = not applicable

Source: Baseline surveys and 12-month follow-up surveys.

Notes: The estimate for the group difference is the difference in percentages between groups based on OLS regression, adjusted for participant age. The percentage for Family Expectations is the adjusted value based on the analysis. The effect size is estimated by the Cox Index (see Appendix E).

## 2. Destructive conflict, communication skills, and depressive symptoms

The continuous outcome variables (destructive conflict, communication skills, depressive symptoms) were evaluated at the level of the individual, rather than the couple. For these outcomes, multilevel modeling was used to account for the interdependence of nested data; that is, of individuals within couples. Hierarchical Linear Modeling Version 7.0 (HLM 7.0) was used to analyze data and estimate impacts (Raudenbush, Bryk, Fai, Congdon, & du Toit, 2011). The HLM model included two levels: individual and couple. Age and the baseline value of each outcome variable were entered at the first level and group assignment (Family expectations versus control) was entered at the second.

Results indicated that there was a statistically significant difference between groups, such that those who were assigned to Family Expectations reported lower levels of destructive conflict ( $M = 1.05$ ) at follow-up compared to those who were assigned to the control group ( $M = 1.13$ ,  $p = 0.04$ ; see Table V.3). There were no statistically significant differences between groups for communication skills ( $p = 0.51$ ) or depressive symptoms ( $p = 0.18$ ).

**Table V.3. Post-intervention estimated mean differences on destructive conflict, communication skills, and depressive symptoms using data from 12-month follow-up**

Outcome measure	Family Expectations M (SD)	Control M (SD)	FE versus Control mean difference (p-value of difference)	Effect size
Destructive Conflict	1.05 (0.84)	1.13 (0.90)	-0.08 (.039)	0.10
Communication Skills	4.90 (1.40)	4.86 (1.40)	0.04 (.507)	0.03
Depressive Symptoms	13.63 (14.00)	14.45 (13.98)	-0.82 (.177)	0.07

n.a. = not applicable

Source: Baseline surveys and 12-month follow-up surveys. See Figure B.3. in the Appendix for the Family Expectations and control group sample sizes for each outcome measure.

Notes: Means are adjusted to account for covariates. Effect size is Hedges' g, calculated by dividing the differences in adjusted means by the weighted pooled standard deviation for the analytic sample at baseline.

Table E.1. in the Appendix provides the regression output from each impact analysis, which corresponds to the means, mean differences, and statistical tests of significance reported in Table V.3.

## C. Sensitivity analyses

### Key Findings:

Results of sensitivity analyses demonstrate that the primary impact evaluation results are relatively unchanged after including additional baseline covariates.

This section presents findings from sensitivity analyses conducted to check the robustness of the primary impact findings to the inclusion of an expanded set of covariates listed in Table V.4. The expanded set of covariates was chosen based on theory in order to include a variety of commonly used control variables at both the individual or couple level. All covariates were entered into the model at once. Including such covariates could control for any baseline differences remaining after randomization and potentially improve the precision of estimated impacts. See Appendix F for more information on the analytic approach.

**Table V.4. Covariates included in sensitivity analyses**

Covariate	Description of the covariate
Age	Age (in years)
Sex	Male (0) or female (1)
Employment status	Full-time employment (1) or not (0)
Education	No high school diploma (0) or High school diploma (1)
Race/Ethnicity	Three separate binary variables: Hispanic (1) or not (0); Black/African American (1) or not (0); Native American (1) or not (0)
Marital status	Marital status (1 = married; 0 = not married)
Cohabitation status	Cohabitation status (1 = cohabiting, 0 = not cohabiting)

Covariate	Description of the covariate
Have prior child(ren) together	Prior child(ren) together prior to the pregnancy with the target child (1 = Yes, 0 = No)
Combined income	Combined total couple-level income in the past 12 months before taxes and deductions in categorical increments of \$5,000. Scores ranging from 0 (no income) to 31 (\$150,000 - \$155,000 in combined income) and were analyzed as a continuous variable.
Relationship length	Length of the current romantic relationship in months
Outcome value at baseline	Mean levels of <i>Destructive Conflict</i> , <i>Communication Skills</i> , or <i>Depressive Symptoms</i> as of baseline data collection

Notes: All covariates included in the sensitivity analyses refer to participants' responses at baseline. Sex was not included in sensitivity analyses for *relationship stability* because it was a couple-level analysis. Outcome value at baseline was only included in analyses of *destructive conflict*, *communication skills*, and *depressive symptoms*.

The impact results using the additional covariates are presented in Table V.5 and are relatively unchanged from the primary impact findings. As in the primary findings, those who were assigned to Family Expectations reported lower levels of destructive conflict ( $M = 1.05$ ) at follow-up than those who were assigned to the control group ( $M = 1.13$ ,  $p = 0.03$ ,  $ES = 0.10$ ). There were no statistically significant differences between groups for relationship stability ( $p = 0.31$ ), communication skills ( $p = 0.51$ ), or depressive symptoms ( $p = 0.13$ ). Table F.1. in the Appendix provides the regression output from the sensitivity analyses accounting for additional covariates.

**Table V.5. Differences in means between Family Expectations and control groups estimated using alternative methods**

Outcome	Primary impact analyses	With addition of covariates from Table V.4
Relationship Stability	2.4 (.378)	2.6 (.305)
Destructive Conflict	-0.08 (.039)*	-0.08 (.030)*
Communication Skills	0.04 (.507)	0.04 (.507)
Depressive Symptoms	-0.82(.177)	-0.92(.130)

Source: Baseline surveys and 12-month follow-up surveys.

\*\*/\*/+ Differences are statistically significant at the .01/.05/.10 levels, respectively.

## D. Additional analyses

### Key Findings:

Two sets of exploratory analyses were conducted. In the first, participants' outcomes differed from the primary impact findings when based on a subgroup of couples who remained together in a steady, stable relationship at the 12-month follow-up. Among these couples, the Family Expectations group reported significantly lower levels of depressive symptoms than the control group; though, other outcome measures did not differ significantly. In the second exploratory analysis, the two groups were compared on the likelihood that couples were steadily involved in relationships that did not have intimate partner violence or high levels of conflict. All three characteristics had to be simultaneously true to be coded as having this outcome. The Family Expectations group was significantly more likely at 12 months to be in these stable, safe, low-conflict relationships than the control group.

#### 1. Impacts for couples in a relationship at follow-up

Additional analyses were conducted in order to evaluate the impact of the intervention on destructive conflict, communication skills, and depressive symptoms among those couples who remained together by the end of the study period (976 individuals and 509 couples in the Family Expectations group; 634 individuals and 328 couples in the control group). The rationale for this set of analyses was that the impacts of Family Expectations may be different based on whether a couple has remained together in a steady, stable relationship. We believed this might be especially warranted for analyses of depressive symptoms because the program is primarily a relationship-focused intervention, not a mental health intervention per se, with presumed effects on wellbeing likely linked to whether the relationship remained intact or not.

These analyses were conducted using the same multilevel modeling structure described above, with individuals nested within couples. The expanded set of baseline covariates from Table V.4 were included in these models.

Results revealed that, among those who remained together, those who were assigned to the Family Expectations group reported significantly lower levels of depressive symptoms ( $M = 11.08$ ) than those who were assigned to the control group ( $M = 12.96$ ;  $p = 0.003$ ; Table V.6). Among those who remained together, there were no statistically significant differences between the Family Expectations group and the control group on communication skills ( $p = 0.21$ ) or destructive conflict ( $p = 0.14$ ). See Table G.1. in the Appendix for the full analysis results.

**Table V.6. Post-intervention estimated effects using data from 12-month follow-up – sensitivity analyses using only participants who remained together**

Outcome measure	Family Expectations M(SD)	control M(SD)	FE versus control mean difference (p-value of difference)	Effect size
Destructive Conflict	0.89 (0.75)	0.94 (0.80)	-0.06 (.140)	0.08
Communication Skills	5.03 (1.34)	4.97 (1.40)	0.09 (.206)	0.07
Depressive Symptoms	11.08 (12.07)	12.96 (13.11)	-1.88 (.003)	0.17
Sample Size	508	327	n.a.	n.a.

Source: 12-month follow-up surveys.

Notes: Means are adjusted to account for covariates. Effect size is Hedges' g, calculated by dividing the differences in adjusted means by the weighted pooled standard deviation for the analytic sample at baseline.

## 2. Impacts on whether couples were in a steady, safe, low conflict relationship at follow-up

An additional analysis was conducted to investigate theory-based predictions about the impact of the Family Expectations program. Based on a study indicating negative outcomes later in life for children whose parents divorced or had highly distressed marriages (Amato, 2001), we constructed a single dichotomous outcome (called the conflict-stability index) that captures whether couples were in a steady relationship characterized by low levels of destructive conflict and no physical aggression at follow-up. Appendix G provides more information on the construction of the outcome measure and the methods used in the analysis.

Results indicated that those who were assigned to the Family Expectations group were significantly more likely to still be in a relationship together characterized as not having high levels of conflict at follow-up (38.4%), than those assigned to the control group (32.1%;  $p = 0.02$ ,  $ES = 0.22$ ; see Table V.7).

**Table V.7. Post-intervention estimated effects on the conflict-stability index using data from 12-month follow-up**

Outcome measure	Family Expectations %	No-Treatment control %	FE versus control mean difference (p-value of difference)	Effect size
Conflict-stability Index	38.4%	32.1%	6.3 (.015)	0.22
Sample Size	691	473	n.a.	n.a.

n.a. = not applicable

Source: Baseline surveys and 12-month follow-up surveys.

Notes: The estimate for the group difference is the difference in percentages between groups based on OLS regression, adjusted for the covariate age. The percentage for Family Expectations is the adjusted value based on the analysis. The effect size is estimated by the Cox Index (see Appendix E).

## VI. DISCUSSION

This study evaluated the effect of the Family Expectations program for couples expecting a baby or who had recently had a baby. In the primary analyses, those assigned to the Family Expectations group reported significantly lower levels of destructive conflict at 12-month follow-up, compared to those assigned to the control group. This finding remained statistically significant when controlling for the characteristics of participants at baseline. Conversely, in both primary analyses and sensitivity analyses, there were no statistically significant impacts on measures of relationship stability, communication skills, or depressive symptoms.

Family Expectations was implemented with fidelity. The 12 sessions (36 hours) of the Becoming Parents Program curriculum were offered to each cohort with nearly half of participants attending all 12 sessions. The implementation findings for the current evaluation of Family Expectations also demonstrated similar, albeit slightly higher, fidelity and dosage compared to the BSF evaluation of Family Expectations. For example, Family Expectations participants in the current report received on average 22.4 hours of content (compared to 20 hours in BSF) and 89% attended at least one session (compared to 76% in BSF). These implementation findings demonstrate that the program has been consistently and successfully implemented with high attendance and retention (Devaney & Dion, 2010). Participants also reported high satisfaction and engagement with the program, including feeling supported by Family Expectations staff. Further, consistent with the analyses of program impact, the majority of participants reported that Family Expectations helped them learn to communicate and manage conflict in their relationships.

When taken together, these impact and implementation findings suggest that Family Expectations was delivered appropriately, was well-received by couples, and made a meaningful impact in reducing couples' destructive conflict. Perhaps the combination of the healthy communication skills and overall support couples received from the program made Family Expectations particularly effective for reducing or halting negative communication patterns among low-income couples. Destructive conflict is an especially important target for couples' interventions because higher levels of conflict are associated with lower relationship satisfaction and greater risk for breakup (e.g., Markman et al., 2010; Gottman, 1994). Further, parental conflict is associated with negative, long-term outcomes for children (Cummings & Davies, 1994; Grych & Fincham, 1990).

This effect on destructive conflict is similar to prior findings in the BSF and SHM evaluations for follow-ups of a similar length (cf. Devaney & Dion, 2010; Hsueh et al, 2012; Wood et al., 2010). At the same time, prior significant findings were not replicated on relationship stability or symptoms of depression (except for among those who remained together). It is possible that changes in the program since the prior studies or subtle changes in recruitment may have resulted in fewer effects, suggesting consideration of future program modifications or enhancements.

Exploratory analyses demonstrated that couples in the Family Expectations group were more likely than couples in the control group to be in a stable, safe, low conflict relationships at the



12-month follow up. Testing this outcome was based on the research by Amato (2001) that lower conflict parenting relationships benefitted children. This outcome is of potential importance for both the wellbeing of adults and their children. Other exploratory findings suggested that Family Expectations may lead to reduced depressive symptoms among the partners who remain together. This finding could be of importance because it provides some evidence that, although Family Expectations focuses on the relationship of the parents, it may also influence individual wellbeing.

The study design provided a stringent test of program impacts. Specifically, an RCT design with an ITT framework was used, such that couples were randomly assigned to the intervention or control group and data were analyzed based on the group to which each couple was assigned, regardless of whether they actually received services. Importantly, couples were not restricted in their ability to access other services. For example, a couple assigned to the control group was free to receive any services that they otherwise would have sought. Indeed, once couples attended the intake appointment and learned more about Family Expectations, they may have been motivated to seek similar services elsewhere if assigned to the control group, or, even if not, they could have found the idea of improving their relationship something to work toward and made changes in their relationships.

As described in the implementation section, some key changes occurred with regard to program content during the course of the study. Specifically, in April 2019, content, but not workshop length, was reduced in almost every unit of the curriculum. Although these changes were made intentionally with the aim of improving the program by allowing more time for core content, it is possible that these changes affected the program's impact, albeit only for a very small portion of those assigned to Family Expectations in this study. In November 2019, another set of minor changes were made to program content. One of these changes was increasing the amount and quality of training in a key communication skill, the Speaker-Listener Technique. At that time, the educators who delivered the program engaged in additional training about how to better coach couples in practicing this skill. These changes were minor, in our opinion, and did not affect most couples in the study due to their timing. It is possible that such changes could enhance the effectiveness of the program going forward.

Family Expectations has 15-year history of being guided by research and participating in rigorous random assignment studies. It is not atypical in this field to see impacts on some dimensions while not finding them on others, in such studies as this one. There was evidence for program impacts on destructive conflict, which is critically important in the functioning of couples and families.



## VII. REFERENCES

- Angrist, J. D., & Pischke, J. S. (2009). *Mostly harmless econometrics: An empiricist's companion*. Princeton, New Jersey: Princeton University Press.
- Amato, P. R. (2001). Good enough marriages: Parental discord, divorce, and children's well-being. *The Virginia Journal of Social Policy & the Law*, 9, 71-94.
- Cummings, E. M., & Davies, P. (1994). *Children and marital conflict*. New York: Guilford.
- Devaney, B., & Dion, R. (2010). *15-Month Impacts of Oklahoma's Family Expectations Program* (p. 76). Mathematica Policy Research. <https://www.mathematica.org/our-publications-and-findings/publications/15month-impacts-of-oklahomas-family-expectations-program>
- Frei, A. E., & Jordan, P. L. (2016). *Becoming Parents Program: Leader's Manual*. Seattle, WA: author's self-publication.
- Hellevik, O. (2009). Linear versus logistic regression when the dependent variable is a dichotomy. *Quality & Quantity*, 43 (1), 59–74. doi: 10.1007/s11135-007-9077-3
- 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. Administration for Children and Families, Office of Planning, Research, and Evaluation. Washington DC.
- Lundquist, E., Hsueh, J., Lowenstein, A. E., Faucetta, K., Gubits, D., Michalopoulos, C., & Knox, V. (2014). A family-strengthening program for low-income families: Final impacts from the Supporting Healthy Marriage evaluation. *New York, NY: MDRC*.
- Gomila, R. (in press). Logistic or linear? Estimating causal effects of treatments on binary outcomes using regression analysis. *Journal of Experimental Psychology: General*. Retrieved from: <https://psyarxiv.com/4gmbv/>
- Grych, J., & Fincham, F. (1990). Marital conflict and children's adjustment. *Psychological Bulletin*, 108, 267-290.
- Gottman, J. (1994). *What predicts divorce? The relationship between marital process and marital outcomes*. Lawrence Erlbaum, New Jersey.
- Markman, H. J., Stanley, S. M., Blumberg, S. L., Jenkins, N. H., & Whiteley, C. (2004). *12 Hours to a Great Marriage*. New York: Wiley and Sons.
- Markman, H., Floyd, F., Stanley, S. & Lewis, H. (1986). Prevention. In N. Jacobson & A. Gurman (eds), *Clinical Handbook of Marital Therapy*. NY, NY: The Guilford Press.
- Markman, H. J., Rhoades, G. K., Stanley, S. M., Ragan, E., & Whitton, S. (2010). The premarital communication roots of marital distress: The first five years of marriage. *Journal of Family Psychology*, 24, 289-298.
- Morris, S. B. (2008). Estimating effect sizes from pretest-posttest-control group designs. *Organizational Research Methods*, 11(2), 346 – 386.
- Raudenbush, S. W., Bryk, A. S., Fai, C. Y., Congdon, R., & du Toit, M. (2011). *HLM 7: Hierarchical linear and nonlinear modeling*. Lincolnwood, IL: Scientific Software International, Inc.
- Sánchez-Meca, J., Chacón-Moscoso, S., & Marín-Martínez, F. (2003). Effect-size indices for dichotomized outcomes in meta-analyses. *Psychological Methods*, 8, 448 – 467.

- Twenge, J. M., Campbell, W., & Foster, C. A. (2003). Parenthood and marital satisfaction: a meta-analytic review. *Journal of Marriage and Family*, 65(3), 574-583. Retrieved from EBSCOhost.
- What Works Clearinghouse. (2020). What Works Clearinghouse procedures handbook, version 4.1. Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance. This report is available on the What Works Clearinghouse website at <https://ies.ed.gov/ncee/wwc/handbooks>.
- Wood, R. G., McConnell, S., Moore, Q., Clarkwest, A., & Hsueh, J. (2010). The Building Strong Families Project. Strengthening unmarried parents' relationships: The early impacts of Building Strong Families. Administration for Children and Families, Office of Planning, Research, and Evaluation. Washington DC.

## VIII. APPENDICES

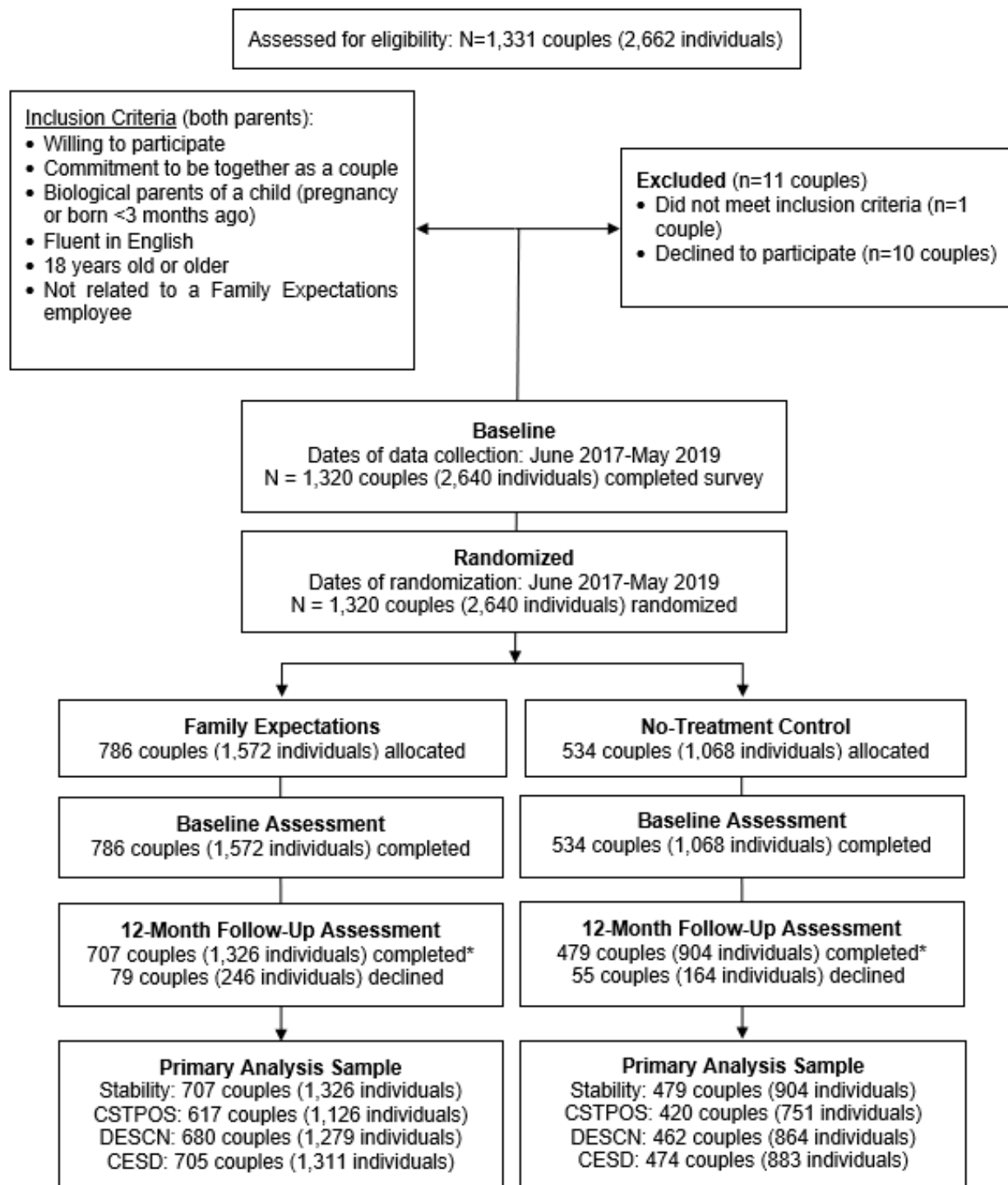
### A. Data and study sample

**Table A.1. Key features of the implementation analysis data collection**

Implementation element	Research question	Data source	Timing/frequency of data collection	Party responsible for data collection
Fidelity	Did each cohort in the Family Expectations group have access to the full set of workshop sessions offered?	Workshop sessions in nFORM	All sessions delivered	Intervention staff
Fidelity	Were there adaptations to the workshops over the course of the study?	Internal notes	Reviewed quarterly	Local evaluator
Dosage	What percentage of couples in the Family Expectations group completed the workshops?	Family Expectations MIS; nFORM	All sessions delivered	Intervention staff
Dosage	What was the average attendance of couples (number of workshops)?	Family Expectations MIS; nFORM	All sessions delivered	Intervention staff
Dosage	What percentage of couples in the Family Expectations group received case management, and what was the dosage?	Family Expectations MIS; nFORM	Daily	Intervention staff
Quality	What was the quality of staff—participant interactions during the intervention?	Survey items on working alliance	At Post-Program Survey and 12-Month Follow-Up Survey	Study staff
Engagement	What was intervention couples' overall feedback on the program?	Survey items on program feedback	At Post-Program Survey and 12-Month Follow-Up Survey	Study staff
Context	What outside services were utilized by couples in the in the Family Expectations group during the evaluation period?	Survey items on other services used during evaluation period	At 12-Month Follow-Up Survey	Study staff

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

	<b>Data source</b>	<b>Timing of data collection</b>	<b>Mode of data collection</b>	<b>Party responsible for data collection</b>	<b>Start and end date of data collection</b>
Intervention	Intervention group study participants	Enrollment (baseline); End of intervention (post-test); 12 months after baseline (follow-up)	In-person online survey (nFORM and/or Qualtrics)	Program staff	June 2017 through June 2020
Counterfactual	Comparison group study participants	Enrollment (baseline); 12-months after baseline (follow-up)	In-person online survey (nFORM and/or Qualtrics)	Program staff	June 2017 through June 2020

**Figure A.3. Family Expectations CONSORT diagram**

\*At least one member of the couple completed the follow-up survey.

## B. Data preparation

Data was merged from three sources by using participants' unique ID number: baseline and post-test data stored in nFORM and Qualtrics, follow-up data stored in Qualtrics only, and information regarding random assignment and program attendance stored in the MIS. The evaluation team downloaded data from each time point from Qualtrics. The program team provided data extracts from nFORM and the MIS.

Multiple data quality checks were embedded into the measures in order to prevent errors. For example, each participant entered the first name of their partner and child. This step provided an opportunity to catch any errors in the pairing of data across partners. For obvious data entry errors, (e.g., incorrect birthdates), we consulted other records to replace the data with the correct information.

For measures of continuous outcomes (communication skills, destructive conflict, and depressive symptoms), the planned analyses allowed each partner to have their own score (and own covariate) per description of the models below. However, relationship stability was scored as a single value for each dyad. Participants were asked, "Which of the following statements best describes your current relationship with [partner]?" Possible responses were "We are not in a romantic relationship," "We are involved in an on-again and off-again relationship," or "We are romantically involved on a steady basis." There is ample evidence suggesting that on-again-off-again relationships are of substantially lower quality than steady relationships (e.g., Dailey et al., 2009). Individuals report less conflict and interpersonal violence in relationships that are either steady or broken up for good compared to those in on-again-off-again relationships (Halpern-Meekin et al., 2013). Thus, we scored relationships as stable if the couple reported being romantically involved on a steady basis and as unstable if the couple reported anything else. It is possible for two partners to not agree on the nature of their relationship. For a couple to be scored as stable, both partners were required to report that they were romantically involved on a steady basis. Either or both reporting that they were not was otherwise considered an unstable relationship.

Missing data were addressed using a variety of theoretically- and empirically-guided strategies. At the level of each questionnaire measure, small amounts of missing data were addressed by using mean scores. For example, if a participant accidentally or intentionally skipped an item on a longer scale, the mean score would reflect the average value of their responses to other items aimed at measuring the same construct. Means were computed as long as the participant had completed at least 80% of the items on the measure. For each outcome variable, the analytic sample consists of individuals who did not have any missing data for any of the variables included in the model.

## D. Impact estimation

### 1. Baseline Equivalence

All tests of baseline equivalence include only those research participants who are in the analytic sample for the outcome in question.

**Relationship Stability.** Analyses for stability were not multilevel analyses, as the couple-level outcome was tested with a couple-level predictor (GROUP). For baseline equivalence tests related to stability, dichotomous, couple-level baseline covariates (e.g., married or not) were tested by comparing the groups at baseline using logistic regression.

$$\log(p[\text{MARRIED}]/1 - [\text{MARRIED}]) = b_0 + b_1 * (\text{GROUP})$$

In the same manner, dichotomous, individual-level covariates (e.g., high school graduate or not) were also tested using logistic regression. For all individual-level covariates used in analyses of stability, male and female scores were tested separately (and they are entered separately in the impact analyses), for example, female education.

$$\log(p[\text{F EDU}]/1 - [\text{F EDU}]) = b_0 + b_1 * (\text{GROUP})$$

Continuous couple-level (e.g., duration of relationship in months) or individual-level covariates (e.g., age) were tested using t-tests comparing the two groups.

$$t = \frac{m_A - m_B}{\sqrt{\frac{S^2}{n_A} + \frac{S^2}{n_B}}}$$

**Continuous Measures of Relationship Outcomes.** All equations for estimating the baseline equivalence related to continuous, individual-level covariates (e.g., age, destructive conflict) were of the same form as those for testing the primary analyses of continuous measures of relationship outcomes (see below). Therefore, below we only present the equation for tests of dichotomous, individual-level covariates. Dichotomous, individual-level covariates used in tests of the continuous outcomes (e.g., Hispanic or not) were tested using multilevel modeling with HLM's Bernoulli routine for analyzing a binary outcome with individuals nested within couples, comparing the two groups. This is essentially a logistic regression testing if the probability of an individual being in one category or the other differs by groups at baseline, accounting for the dependency between partners.

The equations below are for the two levels of this model using the example of testing for a group difference in the probability of identifying as being Hispanic. In the equations below, subscript  $i$  represents individuals and  $j$  represents couples.  $\varphi_{ij}$  is the probability of an individual being Hispanic given their couple membership (individuals are nested within couples).  $B_j$  denotes the couple ( $j$ ) in which an individual is nested. The probability of an individual being Hispanic is

transformed into logged odds, where  $\eta_{ij}$  is the predicted log odds of individual  $i$  in couple  $j$  being Hispanic, which is estimated by the intercept  $\beta_{0j}$ . The log odds for being Hispanic ( $\beta_{0j}$ ) is then predicted using the level-2 model which includes a variable denoting intervention assignment (group) for couples and a random error term ( $u_{0j}$ ) allowing variation between couples on the odds of being Hispanic. The analysis yields the predicted log odds of the control group ( $\gamma_{00}$ ), the difference between that value and the log odds for the intervention group (Group;  $\gamma_{01}$ ), and the random error term for level-2 ( $u_{0j}$ ).  $\gamma_{01}$  is then exponentiated to yield an odds ratio reflecting the difference in the odds for individuals in the intervention and control groups being Hispanic, accounting for the nesting of individuals within couples.

### **Level-1 Model**

$$Prob(HISP_{ij} = 1 | \beta_j) = \phi_{ij}$$

$$\log[\phi_{ij} / (1 - \phi_{ij})] = \eta_{ij}$$

$$\eta_{ij} = \beta_{0j} \quad \eta_{ij} = \beta_{0j}$$

### **Level-2 Model**

$$\beta_{0j} = \gamma_{00} + \gamma_{01} * (GROUP_j) + u_{0j}$$

Dichotomous, couple-level covariates (e.g., married or not) used in analyses of continuous outcomes were tested using logistic regression.

$$\log(p[MARRIED] / 1 - [MARRIED]) = b_0 + b_1 * (GROUP)$$

Continuous, couple-level covariates (e.g., duration of relationship in months) used in analyses of continuous outcomes were tested using independent t-tests.

## **2. Impact Analyses**

**Relationship Stability.** Program impact on relationship stability (and the supplemental analysis on the conflict-stability index) was evaluated with OLS Linear Probability Modeling, with the binary outcome of stability regressed on group assignment and other covariates. While such dichotomous outcomes are often evaluated with logistic regression, OLS has an advantage of providing easily interpretable differences in the percentage of couples in the groups who are together at the follow-up, net of baseline covariates included in the models. It has been argued that such methods are particularly appropriate in randomized controlled trials (Gomila, in press). Further, the percentages of couples still together at follow-up will be in a range where OLS and logistic regression should yield nearly identical p values (Angrist & Pischke, 2009; Hellevik



2009; 2009). Indeed, there would be no difference in interpretation in the results from these analyses conducted by either OLS or logistic regression, hence, we favor the more interpretable results from OLS.

The functional form of the equation is:

$$STABILITY = b_0 + b_1(GROUP) + b_2(MALE\_AGE) + b_3(FEMALE\_AGE)$$

Where coefficient  $b_0$  reflects the intercept for the control group, assuming average age and average level of the dependent variable at baseline. The coefficient  $b_1$  refers to the effect of the dependent variable at baseline. The coefficients  $b_2$  and  $b_3$  refers to the effects of male and female participant's age at baseline.

**Continuous Measures of Relationship Outcomes.** The functional form of all equations for 0dccc estimating the impact on continuous measures of relationship outcomes (destructive conflict, communication skills, depressive symptoms) were the same for the primary analyses. The equations represent a two-level model with individuals ( $i$ ) nested within couples ( $j$ ).

#### **Level-1 Model**

$$\text{DEPENDENT VARIABLE AT FOLLOWUP}_{ij} = \beta_{0j} + \beta_{1j} * (\text{AGE}_{ij}) + \beta_{2j} * (\text{DEPENDENT VARIABLE AT BASELINE}_{ij}) + r_{ij}$$

#### **Level-2 Model**

$$\beta_{0j} = \gamma_{00} + \gamma_{01} * (\text{GROUP}_j) + u_{0j}$$

$$\beta_{1j} = \gamma_{10}$$

$$\beta_{2j} = \gamma_{20}$$

\*Bolted terms are grand mean centered.

At level 1, individuals' scores (subscripted  $ij$ ) are estimated from their couple means ( $B_{0j}$ ), controlling for their own ages and their values for the dependent variable at baseline, with residual variation ( $r_{ij}$ ) around their couple means. At level 2, the intercept for each couple ( $B_{0j}$ ) is estimated based on the sample grand mean ( $\gamma_{00}$ ), a parameter ( $\gamma_{01}$ ) representing the intervention assignment of the couple (GROUP), and a random error term allowing each couple's intercept to differ from the grand mean by  $u_{0j}$ , which makes this a random intercept model. In such two-level models, the error terms  $r$  and  $u$  are assumed to be random, normally distributed, with a mean value of 0, and to have constant variance. In the model shown, GROUP = 0 for the control group and GROUP = 1 for those assigned to Family Expectations. The parameter  $\gamma_{00}$  is an intercept estimating the mean of the control group and  $\gamma_{01}$  estimates the difference between the intercepts for the intervention and control groups.

**Table D.1. Post-intervention estimated effects on destructive conflict, communication skills, and depressive symptoms using data from 12-month follow-up**

Outcome measure	Destructive Conflict		Communication Skills		Depressive Symptoms	
	B	SE	B	SE	B	SE
Intercept	1.13***	0.03	4.86***	0.05	14.45***	0.47
Intervention	-0.08*	0.04	0.04	0.07	-0.82	0.61
Outcome value at baseline	0.57***	0.02	0.47***	0.02	0.52***	0.02
Age	-0.01	0.00	0.01	0.00	-0.19***	0.05

Source: Baseline surveys and 12-month follow-up surveys.

Notes: Significant results are noted with asterisks: \*  $p < .05$ , \*\*  $p < .01$ , \*\*\*  $p < .001$ . In these multilevel models, “intercept” coefficients are the estimated means for the control group, and the coefficients for Family Expectations reflect the estimated difference from those means for the intervention group.

### 3. Estimation of Effect Sizes.

**Continuous measures.** We report Hedges’  $g$  for estimation of effect sizes for continuous outcomes. That is, we divide the mean difference of the groups at the follow-up by the pooled standard deviation for the measure at baseline.

**Dichotomous measures.** As noted earlier, for estimating the percentage point differences between groups for outcomes such as stability, we used Linear Probability models (OLS applied to binary outcomes). We report effect sizes for these outcomes using the Cox index (which estimates effect sizes for dichotomous outcomes in a scaling roughly comparable to Hedges’  $g$ . Cox index values are calculated based on odds ratios (Sánchez-Meca, Chacón-Moscoso, & Marín-Martínez, 2003), using the guidelines presented in the What Works Clearinghouse Procedures Handbook (2020). To obtain the odds ratios for conversion to the Cox index values, we ran the identical models as reported from OLS in logistic regressions. These two different approaches in the form of regression returned virtually identical  $p$  values.

### E. Sensitivity analyses and alternative model specifications

**Relationship Stability.** Below is the regression equation used to test stability as an outcome (using OLS).

$$\text{STABILITY} = b_0 + b_1(\text{GROUP}) + b_2(\text{M\_AGE}) + b_3(\text{F\_AGE}) + b_4(\text{M EMPLOY}) + b_5(\text{F EMPLOY}) + b_6(\text{M EDU}) + b_7(\text{F EDU}) + b_8(\text{M HISPANIC}) + b_9(\text{F HISPANIC}) + b_{10}(\text{M NAT AM}) + b_{11}(\text{F NAT AM}) + b_{12}(\text{M BLACK/AA}) + b_{13}(\text{F BLACK/AA}) + b_{14}(\text{MARRIED}) + b_{15}(\text{COHAB}) + b_{16}(\text{CHTOG}) + b_{17}(\text{INCOME}) + b_{18}(\text{LENGTH}) + e_{ij}$$

\*M=Male, F=Female

**Continuous Measures of Relationship Outcomes.** All equations for estimating the impact of continuous measures of relationship outcomes were the same for the sensitivity analyses; therefore, we only present the equation for destructive conflict as an example.

**Level-1 Model**

$$\text{DESCNM}_{ij} = \beta_{0j} + \beta_{1j} * (\text{DESCNT}_{0ij}) + \beta_{2j} * (\text{AGET}_{0ij}) + \beta_{3j} * (\text{SEXT}_{0ij}) + \beta_{4j} * (\text{ED\_HST}_{0ij}) + \beta_{5j} * (\text{EMPLFTT}_{0ij}) + \beta_{6j} * (\text{HISPT}_{0ij}) + \beta_{7j} * (\text{BLACKT}_{0ij}) + \beta_{8j} * (\text{NATAMT}_{0ij}) + r_{ij}$$

**Level-2 Model**

$$\beta_{0j} = \gamma_{00} + \gamma_{01} * (\text{GROUP}_j) + \gamma_{02} * (\text{MART}_{0j}) + \gamma_{03} * (\text{LVTGT}_{0j}) + \gamma_{04} * (\text{CHTGT}_{0j}) + \gamma_{05} * (\text{LENGHT}_{0j}) + \gamma_{06} * (\text{INCT}_{0j}) + u_{0j}$$

$$\beta_{1j} = \gamma_{10}$$

$$\beta_{2j} = \gamma_{20}$$

$$\beta_{3j} = \gamma_{30}$$

$$\beta_{4j} = \gamma_{40}$$

$$\beta_{5j} = \gamma_{50}$$

$$\beta_{6j} = \gamma_{60}$$

$$\beta_{7j} = \gamma_{70}$$

$$\beta_{8j} = \gamma_{80}$$

\*Bolded terms are grand mean centered.

**Table E.1. Post-intervention estimated effects using data from 12-month follow-up – sensitivity analyses with additional covariates – multilevel models**

Outcome measure	Destructive Conflict		Communication Skills		Depressive Symptoms	
	B	SE	B	SE	B	SE
Intercept	1.13***	0.03	4.85***	0.05	14.54***	0.47
<b>Intervention</b>	<b>-0.08*</b>	<b>0.04</b>	<b>0.04</b>	<b>0.07</b>	<b>-0.92</b>	<b>0.60</b>
Baseline value of the outcome variable	0.56***	0.02	0.47***	0.02	0.50***	0.02
Age	0.00	0.00	0.00	0.01	-0.14*	0.05
Sex	0.08**	0.03	0.02	0.05	0.29	0.52
Baseline employment status	-0.05	0.03	0.12	0.06	-1.73**	0.60
Education	-0.02	0.04	0.10	0.07	-0.36	0.64
Race: Black/African American	-0.05	0.04	-0.04	0.07	0.02	0.65
Race: Native American	-0.01	0.05	0.05	0.09	0.50	0.82
Ethnicity	-0.10*	0.04	0.10	0.08	-0.59	0.74
Baseline marital status	-0.14*	0.04	0.15	0.08	-0.05	0.74
Baseline cohabitation status	-0.09	0.05	0.11	0.10	-2.10*	0.88
Number of previous children together	0.02	0.04	-0.08	0.08	1.19	0.70

Outcome measure	Destructive Conflict		Communication Skills		Depressive Symptoms	
	B	SE	B	SE	B	SE
Combined income	0.01	0.00	0.00	0.01	-0.09	0.07
Relationship length	-0.00*	0.00	0.00	0.00	-0.01	0.01

Source: Baseline surveys and 12-month follow-up surveys.

Notes: Significant results are noted with asterisks: \*  $p < .05$ , \*\*  $p < .01$ , \*\*\*  $p < .001$ . In these multilevel models, “intercept” coefficients are the estimated means for the control group, and the coefficients for Family Expectations reflect the estimated difference from those means for the intervention group.

## F. Additional analyses

Table G.1. Post-intervention estimated effects using data from 12-month follow-up – additional analyses using only participants who remained in a relationship – multilevel models

Outcome measure	Destructive Conflict		Communication Skills		Depressive Symptoms	
	B	SE	B	SE	B	SE
Intercept	0.94***	0.03	4.97***	0.05	12.96***	0.50
<b>Intervention</b>	<b>-0.05</b>	<b>0.04</b>	<b>0.06</b>	<b>0.07</b>	<b>-1.88**</b>	<b>0.64</b>
Baseline value of the outcome variable	0.51	0.02	0.46***	0.02	0.48***	0.03
Age	0.00	0.00	0.00	0.01	-0.10	0.05
Sex	0.03	0.03	0.03	0.06	1.08	0.56
Baseline employment status	-0.06***	0.03	0.09	0.06	-1.34*	0.63
Education	0.02	0.04	0.13	0.07	-0.21	0.69
Race: Black/African American	-0.09*	0.04	-0.00	0.07	-0.59	0.71
Race: Native American	-0.04	0.05	0.00	0.09	-0.34	0.86
Ethnicity	-0.12**	0.04	0.10	0.08	-0.88	0.76
Baseline marital status	-0.07	0.04	0.11	0.08	0.79	0.74
Baseline cohabitation status	-0.12	0.07	0.06	0.12	-0.18	1.01
Number of previous children together	0.02	0.04	-0.08	0.08	1.54*	0.73
Combined income	0.00	0.00	-0.00	0.01	-0.08	0.07
Relationship length	-0.00**	0.00	0.00	0.00	-0.01	0.01

Source: Baseline surveys and 12-month follow-up surveys.

Notes: Significant results are noted with asterisks: \*  $p < .05$ , \*\*  $p < .01$ , \*\*\*  $p < .001$ . In these multilevel models, “intercept” coefficients are the estimated means for the control group, and the coefficients for Family Expectations reflect the estimated difference from those means for the intervention group.

The **conflict/stability index**, a dichotomous variable used in the additional analyses, is based on the following criteria (all must be met for a score of 1).

1. The partners report that they are steadily involved together in a romantic relationship (not just co-parenting, and not in an on-again-off-again relationship).
2. The individual partners’ scores on the measure of destructive conflict fall below the top third of the distribution of the sample at follow up.
3. Neither partner reports physical aggression in their relationship within the three months prior to the follow-up. This means neither partner reported throwing something that could hurt the

other, pushing, shoving, slapping, kicking, or hitting the other. Further, neither partner reported there being a physical injury because of a fight between them.

The intervention impact on this variable was then tested in an OLS regression equation of the same form used in the analyses of relationship stability noted above. All of the baseline covariates listed in Table V.4 were included: age, employment status, education, race, and ethnicity of each partner, as well as each couple's marital status, cohabitation status, prior children together, combined income, and relationship length at baseline. We also included the baseline scores for the two measures that combine with stability to make the conflict/stability index (low destructive conflict and no recent history of physical aggression).

## G. Additional tables of implementation findings

**Table G.1. Individual item responses regarding the *quality* of staff-participant interactions during Family Expectations**

Working Alliance survey Item	% of Participants who endorsed <i>Somewhat Agree, Agree, or Strongly Agree</i>
I believe the Family Expectations staff liked me	90.8
The workshop leaders understood what I wanted to accomplish in the program	90.3
I was confident that the workshop leaders knew how to help us	90.3
The Family Expectations team worked with us toward goals that made sense to me	91.1
I believe the way we worked on strengthening our relationship in Family Expectations was a good fit for us	89.0
The Family Expectations team truly cared about us	91.4
I felt respected by the Family Expectations team	93.0

**Table G.2. Individual item responses regarding program feedback**

As a result of attending Family Expectations...	% of Participants who endorsed <i>Somewhat Agree, Agree, or Strongly Agree</i>
[My partner] and I are better able to talk about things in a positive way	79.4
I will invest more time in my relationship with [my partner]	81.7
I think [my partner] and I will work more as a team	83.8
I have learned skills that will help me in all of my relationships	89.4
We will be better able to keep arguments from getting overheated	82.9
We know more about how to get help and support for our relationship	85.4
I am more willing to get help when we need it	85.5
[My partner] and I work better as a team to manage our family's needs	83.5
I can keep myself calmer when I don't like how we're handling something	83.7
I can keep my thoughts about [my partner] more positive when we're having a conflict	82.1
I better understand how a baby communicates what she/he needs	90.9
I feel more confident that I can comfort a baby when she/he is upset	91.1
I feel that I have more support as a parent from other people	82.8
I better understand a baby's basic emotional and physical needs	90.6
I know more about when a baby wants to play and when a baby needs a break	90.0
I better understand how to create a safe sleep environment for [my baby]	92.6
I feel more confident getting the support I need to meet [my baby]'s needs	91.7