Impact Evaluation of the New Pathways to Responsible Fatherhood

Family Formation Program in Saint Louis, Missouri

Final Impact Evaluation Report for
Fathers & Families Support Center, Inc.

11.02.2020

Prepared by

Courtney King, MPH, MSW, Melissa Krauss, MPH, Shih-Ying Cheng, MSW
Nancy Mueller, MPH, Patricia Kohl, PhD, Patrick Fowler, PhD,

Brown School Evaluation Center at Washington University in Saint Louis

Recommended Citation:

Brown School Evaluation Center at Washington University in St. Louis. (2020). *Final Impact Evaluation Report: The Evaluation of the New Pathways to Responsible Fatherhood Family Formation Program.* Administration for Children and Families. https://www.acf.hhs.gov/ofa/programs/healthy-marriage

Acknowledgements:

We would like to thank Halbert Sullivan, Cheri Tillis and Destini Goodwin and the program staff of the Fathers & Families Support Center (FFSC) and AMTC & Associates for their partnership in this evaluation and their guidance and contributions to the evaluation and final report. We would also like to acknowledge the contributions of Amy Sorg, Sarah Bobmeyer, Courtney Cantwell, Marlene Cerritos-Rivas, Meihsi Chiang, and the student interviewers at the Brown School Evaluation Center.

We are grateful to the FFSC team for their continued dedication and service to fathers, their children and families in the St. Louis community.

Disclosure:

The authors declare that there is no conflict of interest.

This publication was prepared under Grant Number 90FK0069 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:

Objective. This evaluation assessed the effectiveness of the New Pathways to Responsible Fatherhood Family Formation (NPFF) project, an evidence-based program, provided by Fathers & Families Support Center (FFSC) for economically disadvantaged fathers. FFSC offers two programs: (a) Family Formation (FF), a 6-week/240-hour program that implements curricula on economic stability and mobility, responsible fatherhood, and healthy relationships, as well as case management and legal services; (b) Economic Stability (ES), a 4-week/80-hour program that implements an economic stability curriculum (the same as FF) and limited employment-related case management and legal services. The evaluation objective was to examine differences in parenting, relationship, financial stability, and well-being outcomes between fathers receiving the FF versus ES program.

Study design. A randomized controlled trial (RCT) was used to compare fathers in FF (n=350) versus ES program (total n=342). Surveys were administered at enrollment and 3- and 12-months post-intervention via telephone. Primary and secondary outcomes, including father involvement, co-parenting/parenting skills, father/child well-being, and financial responsibility and stability, were compared between the study groups at 3-months (secondary outcomes, n=415) and 12-months post-intervention (primary outcomes, n=378). Three staff focus groups and twelve participant focus groups were conducted to assess program experience and implementation.

Results. Approximately 38% of fathers attended ≥75% of program sessions (FF 40% vs ES 35%) and 37% graduated from the program (FF 39% vs ES 35%). Nearly all fathers were non-white (94% African American, 4% other/mixed race). While both FF and ES groups experienced improvements at 3-months and 12-months post-intervention, the FF group did not fare better on outcomes compared to the ES group. Fathers felt that both programs provided a supporting community of fathers and valuable knowledge to help gain employment and improve relationship and parenting skills. Program challenges included low program completion for both groups and staff turnover.

Conclusion. Although both groups experienced improvements in outcomes over time, FF fathers did not experience better outcomes than ES fathers. The lack of difference in outcomes could be due to a similar core focus on employment-related curriculum for both groups. It is possible that gaining financial stability contributed to positive improvements in other fatherhood domains.

Contents

I. Introduction 1

A. Introduction and study overview 1
B. Primary research question(s) 3
C. Secondary research question(s) 3

II. Intervention and counterfactual conditions 4

A. Description of program as intended 4
B. Description of counterfactual condition as intended 6
C. Research Questions about the Intervention and Counterfactual Conditions as Implemented 6

III. Study design 8

A. Sample formation and research design 8
B. Data collection 9

IV. Analysis methods 10

A. Analytic sample 10
B. Outcome measures 13
C. Baseline equivalence and sample characteristics 16

V. Findings and estimation approach 22

A. Implementation evaluation 22
B. Primary impact evaluation 28
C. Sensitivity analyses 30
D. Additional analyses 31

VI. Discussion 36

VII. References 40

VIII. APPENDIX TABLES AND FIGURES TO SUPPLEMENT FINAL IMPACT REPORT 44

Appendix A: Logic Model & Implementation Schedules 45

Appendix B: Data, Sample, and Measures 47

Appendix C: Baseline Equivalence of the RCT Design for Secondary Outcomes 53

Appendix D: Data Preparation 63

Appendix E: Impact Estimation 64

Appendix F: Sensitivity Analyses and Alternative Model Specifications 66

Appendix G: Additional Analyses 67

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 6

IV.1. Individual sample sizes by intervention status for primary outcomes at 12-month follow-up survey 12

IV.2. Outcome measures used for primary impact analyses research questions 15

IV.4a. Summary statistics of key baseline measures and baseline equivalence across study groups, for fathers completing the Father Engagement Caregiving/Play subscale in the follow-up survey at 12 months post-intervention (n=329) 17

IV.4b. Summary statistics of key baseline measures and baseline equivalence across study groups, for fathers completing the Co-Parenting Relationship Skills Alliance subscale in the follow-up survey at 12 months post-intervention (n=300) 18

IV.4c. Summary statistics of key baseline measures and baseline equivalence across study groups, for fathers completing the Parenting Skills Psychological Aggression subscale in the follow-up survey at 12 months post-intervention (n=280) 19

IV.4d. Summary statistics of key baseline measures and baseline equivalence across study groups, for fathers completing the Father Well-Being Mental Health Composite Scale in the follow-up survey at 12 months post-intervention (n=326) 20

IV.4e. Summary statistics of key baseline measures and baseline equivalence across study groups, for fathers completing the Child Well-Being Aggressive Behavior subscale in the follow-up survey at 12 months post-intervention (n=190) 21

V.1. Covariates included in impact analyses[a] 29

V.2. Post-intervention estimated effects using data from survey 12 months after the program to address the primary research questions 30

V.3. Post-intervention estimated effects using data from survey 3 months after the program to address the secondary research questions 32

V.4. Mixed model results examining change in scores over time and between groups using data from baseline and survey 12 months after the program 35

B.1. Data used to address implementation research questions 47

B.2. Key features of the impact analysis data collection 48

B.3. Individual sample sizes by intervention status for secondary outcomes 51

B.4. Outcome measures used for secondary impact analyses research questions 52

C.1. Summary statistics of key baseline measures and baseline equivalence across study groups, for fathers completing the Father Engagement Caregiving/Play subscale in the follow-up survey at 3 months post-intervention (n=362) 54

C.2. Summary statistics of key baseline measures and baseline equivalence across study groups, for fathers completing the Co-Parenting Relationship Skills Alliance subscale in the follow-up survey at 3 months post-intervention (n=336) 55

C.3. Summary statistics of key baseline measures and baseline equivalence across study groups, for fathers completing the Parenting Skills Psychological Aggression subscale in the follow-up survey at 3 months post-intervention (n=304) 56

C.4. Summary statistics of key baseline measures and baseline equivalence across study groups, for fathers completing the Financial Responsibility item in the follow-up survey at 3 months post-intervention (n=156) 57

C.5 Summary statistics of key baseline measures and baseline equivalence across study groups, for fathers completing the Employment item in the follow-up survey at 3 months post-intervention (n=359) 58

C.6. Summary statistics of key baseline measures and baseline equivalence across propensity score matched study groups, for fathers completing the Parenting Skills Psychological Aggression subscale in the follow-up survey at 3 months post-intervention (n=262) 60

C.7. Summary statistics of key baseline measures and baseline equivalence across propensity score matched study groups, for fathers completing the Financial Responsibility item in the follow-up survey at 3 months post-intervention (n=130) 61

C.8. Summary statistics of key baseline measures and baseline equivalence across propensity matched study groups, for fathers completing the Employment item in the follow-up survey at 3 months post-intervention (n=316) 62

F.1. Post-intervention estimated effects using data from survey 12 months after the program to address the primary research questions, not adjusting for baseline covariates 66

G.1. Lesson topics attended by program assignment 68

G.2. Post-intervention estimated effects using data from survey 3 months after the program to address the secondary research questions, not adjusting for baseline covariates 69

G.3. Differences in means between intervention and comparison groups for primary outcomes estimated using treatment-on-treated approach 71

G.4. Differences in means between intervention and comparison groups for secondary outcomes using treatment-on-treated approach 72

G.5. Mixed model results examining change in scores over time and between groups using data from baseline and survey 3 months after the program to address the secondary research questions 74

Figures

B1. CONSORT diagram 49

G1. Unadjusted primary outcomes by time and intervention group 75

G2. Unadjusted secondary outcomes by time and intervention group 76

Impact Evaluation of the New Pathways to Responsible Fatherhood Family Formation Program in Saint Louis, Missouri

I. Introduction

A. Introduction and study overview

Fathers & Families Support Center (FFSC) is a St. Louis, MO nonprofit organization that was founded in December of 1997. FFSC provides a comprehensive array of services to parents with a particular focus on fathers. Services include parenting and healthy relationship classes, economic and mobility services, legal services, case management and much more. For over 20 years, FFSC has been the one-stop-shop for low-income fathers in the St. Louis community, primarily African-American fathers. There are four major problems that face St. Louis low-income African-American fathers and their families: (1) low levels of healthy family functioning (couple relationships, parenting, co-parenting); (2) low levels of adult and child well-being; (3) low levels of economic stability; and (4) high rates of poverty (U.S. Census Bureau, 2018a; 2018b; 2018c; MO DHSS, 2018; U.S. DOJ, 2019; & CDC, 2018a; 2018b).

In 2011, FFSC along with three agencies in Missouri and Minnesota participated in the Parents and Children Together (PACT) evaluation to determine the effectiveness of their responsible fatherhood (RF) programs using a randomized controlled trial (RCT) (Avellar, Covington, Moore, Patnaik & Wu, 2018). The PACT study randomly assigned 5,522 fathers who enrolled in one of the four RF programs from 2012 to 2015 to (a) a treatment group with RF programs provided, or (b) a control group with no services offered. The study found that compared to fathers in the control group, fathers in the RF group performed better in parenting and employment outcomes at follow up. Specifically, fathers in the RF group engaged in more nurturing behaviors (e.g., encouraging the child to talk about their feelings) and age-appropriate activities with children (e.g., reading books, feeding the child), and were employed for a longer length of time.

In October 2015, FFSC began a five-year project funded by the U.S. Department of Health & Human Services, Administration for Children & Families, Office of Family Assistance that builds upon the prior study to compare two levels of programming, the Family Formation program (FF) versus the Economic Stability program (ES). The overarching goal of the New Pathways to Responsible Fatherhood Family Formation (NPFF) project is to provide evidence-based responsible fatherhood education, economic stability and mobility services, and healthy marriage education to economically disadvantaged fathers in St. Louis, Missouri. The major types of responsible fatherhood interventions include: (1) parent training (social-emotional

wellbeing, counseling, relationship support, and relational skill building); (2) education and/or vocational training; and (3) harm reduction approaches to high-risk behaviors (e.g., substance abuse, unprotected sex). FF is a comprehensive intervention that addresses all three domains, and ES focuses on vocational training.

There is a growing body of evidence demonstrating the effectiveness of RF programs. Much of the prior research has focused exclusively on parent training and a number of randomized, and quasi-experimental studies have been done to compare the outcomes of parent training programs to control groups (Fagan & Iglesias, 1999; Lundahl, Risser, & Lovejoy, 2006; Rienks, Wadsworth, Markman, Einhorn, & Etter, 2011). Cowan and colleagues (2009) conducted the Promoting Father Involvement study, an RCT, to evaluate the impact of a 16-week group for fathers, a 16-week group for couples and a low-dose comparison condition (parents attended one 3-hour group session). Findings showed that the interventions positively impacted parent well-being (particularly father engagement and father sense of self) and child well-being.

Less focus has been placed on vocational training combined with parent training; however, the limited results suggest positive outcomes. Robbers et al. (2009) found that young fathers who participated in a parenting program while they attended job training workshops demonstrated improved father involvement with their child. No job placement or employment outcomes were reported. Another study found that fathers participating in the combined intervention secured employment and had increased contact with their children, though it is uncertain which component of the intervention was most essential to achieving these positive outcomes (Barthelemey & Coakley, 2017).

According to Bronte-Tinkew, Burkhauser, & Metz (2012), RF programs appear to be effective; however, it is imperative to further the current knowledge by investigating the key promising elements of RF programs to answer questions about “what works”. This project sought to disentangle these components.

The NPFF project includes an implementation evaluation and an impact evaluation. The Brown School Evaluation Center at Washington University in St. Louis led both evaluations and worked in collaboration with AMTC & Associates. FFSC, AMTC, and the Evaluation Center communicated on a regular basis about program implementation and the evaluations.

The FFSC NPFF Impact Evaluation seeks to quantify the added benefit of combining an economic stability curriculum with curricula on parenting, father-child engagement, and father well-being (FF program) compared to course content containing economic stability material only (e.g., job search, resume development, interviewing skills; ES program). The impact evaluation utilized an RCT design to compare fathers in FF (intervention group) to those in ES (comparison group) on short- and long-term outcomes. The impact evaluation answers 5 primary research questions and 5 secondary research questions. This report provides a summary of the FFSC NPFF Impact Evaluation findings and is organized into the following six sections: I.

Introduction, II. Intervention and Counterfactual Conditions, III. Study Design, IV. Analysis methods, V. Findings and Estimation Approach, and VI. Discussion.

B. Primary research question(s)

Primary research questions examined the effect of FF on long-term outcomes (assessed 12-months after the intervention) of changes in father and child well-being, co-parenting relationship quality, and the father-child relationship. Specific primary research questions include:

- What is the impact of FF relative to ES services only on *Father Involvement* twelve months after the end of the intervention?
- What is the impact of FF relative to ES services only on the *Co-Parenting Relationship* twelve months after the end of the intervention?
- What is the impact of FF relative to ES services only on *Parenting Skills* twelve months after end of the intervention?
- What is the impact of FF relative to ES services only on *Father Well-being* twelve months after the end of the intervention?
- What is the impact of FF relative to ES services only on *Child Well-being* twelve months after the end of the intervention?

C. Secondary research question(s)

Secondary research questions examined the effect of FF on short-term outcomes (assessed 3-months after the intervention) of improved parenting and co-parenting skills, increased father-child involvement, increased financial responsibility of fathers, and progress towards greater economic stability. Specific secondary research questions include:

- What is the impact of FF relative to ES services only on *Father Involvement* three months after the end of the intervention?
- What is the impact of FF relative to ES services only on the *Co-Parenting Relationship* three months after the end of the intervention?
- What is the impact of FF relative to ES services only on *Parenting Skills* three months after the end of the intervention?
- What is the impact of FF relative to ES services only on *Financial Responsibility* three months after the end of the intervention?
- What is the impact of FF relative to ES services only on *Economic Stability* three months after the end of the intervention?

This evaluation was registered by FFSC at clinicaltrails.gov (Identifier: NCT03413709).

II. Intervention and counterfactual conditions

FFSC offers two programs, the Family Formation program (FF; intervention group) and the Economic Stability program (ES; the comparison program). This section describes for each program the intended program components, content, dosage, method of delivery, target population and education and training of staff. Program components, the target population, and staff training are summarized in Tables II.1 and II.2.

A. Description of program as intended

Intended components: FF implements a set of curricula focusing on responsible fatherhood, healthy relationships, and economic stability and mobility. In addition, FF participants receive case management and a variety of employment, legal and support services for up to one year following the completion of the curriculum. Specific legal and support services for FF fathers include access to legal services related to children (i.e., legal visitation, child custody, and child support payment modifications) as well as other services (i.e., warrant recalls, background check and driver's license status).

Intended content: FF content integrates four curricula *Family Formation Curriculum, Within My Reach,* Federal Deposit Insurance Corporation (FDIC) *Money Smarts* and the FFSC *Employment Readiness and Retention* with comprehensive case management services. *Family Formation Curriculum* topics include but are not limited to fatherhood, father/child bonding, effective parenting skills, communication and relationship skills, and conflict resolution. *Within my Reach* focuses on teaching healthy marriage and relationship skills. *Money Smarts* is a curriculum developed by the FDIC designed to help "low- and moderate-income individuals outside the financial mainstream enhance their financial skills and create positive banking relationships." *Employment Readiness and Retention* is an FFSC-developed curriculum designed to focus on helping fathers who have not worked or have had poor work experiences to identify areas of growth to improve their ability to function and succeed in a work environment.

Intended dosage: FF is a six-week, 240-hour program, with sessions occurring 5 days a week for 8 hours per session.

Intended delivery: Group sessions are provided at FFSC facilities in St. Louis, Missouri by trained male facilitators. FFSC facility locations include: (1) Prince Hall Family Support Center (Department of Social Services); (2) North County (Saint Peter's United Church-Christ); (3) MET Center (Metropolitan Education and Training Center); and (4) SLATE (St. Louis Agency on Training and Employment).

Target population: The program is intended to be delivered to low-income (earning below the federal poverty threshold), custodial and non-custodial fathers living in St. Louis.

Education and training of staff: The minimum educational requirement for facilitators is a high school diploma or GED. New facilitators receive a full day of training by experienced FFSC

facilitators in the curriculum and observe class sessions prior to curriculum implementation. During curriculum implementation, experienced facilitators observe the new facilitators daily and correct any facilitation problems that arise. Ongoing training of staff includes annual trainings in the interventions curricula by experienced FFSC facilitators.

Of note, FFSC implemented an attendance and graduation policy for FF: three unexcused missed days resulted in being dropped from the program. And to graduate from the program, you must have attended at least 75% of the program sessions. A logic model describing the activities of FF, implementation objectives and outcomes, as well as FF implementation schedule are included in Appendix A.

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

Component	Family Formation	Economic Stability
	Intervention	***Counterfactual***
Curriculum and content	Economic Stability Legal Services[1] Case Management[2] Responsible Parenting Healthy Relationships	Economic Stability Limited Legal Services[3] Limited Case Management[4]
Dosage and schedule	6 week program, 240 hours; sessions occur five days a week for eight hours per session	4 week program, 80 hours; sessions occur five days a week for four hours per session
Delivery	Group lessons provided at FFSC's facilities by trained male facilitators	Group lessons provided at FFSC's facilities by trained male facilitators
Target Population	Low-income custodial and non-custodial fathers living in St. Louis	Low-income custodial and non-custodial fathers living in St. Louis

[1] Legal Services: Legal visitation, child custody, child support payment modifications, warrant recalls, background check, and driver's license status.

[2] Case Management: Fathers meet with the Social Service and Employment Case Managers a minimum of once a week during the 6 weeks. They receive follow-up after completing the curriculum for up to 1 year, meeting once every two weeks.

[3] Limited Legal Services: Warrant recalls, background check and driver's license status.

[4] Limited Case Management: Fathers meet with only the Employment Case Manager a minimum of once a week during the 4 weeks. They receive one year of follow-up after completing the curriculum until training or employment is obtained, meeting once every two weeks. This case management pertains to employment activities only.

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

Component	Education and initial training of staff	Ongoing training of staff
Intervention		
Family Formation Program	The minimum educational requirement for facilitators is a high school diploma or GED. New facilitators receive a full-day of training in the curriculum.	Facilitators receive annual trainings in the intervention's curricula. New facilitators are observed by experienced facilitators regularly over the six-week program implementing curriculum topics.
Counterfactual		
Economic Stability Program	The minimum educational requirement for facilitators is a high school diploma or GED. New facilitators receive a full-day of training in the curriculum.	Facilitators receive annual trainings in the intervention's curricula. New facilitators are observed by experienced facilitators regularly over the four-week program implementing curriculum topics.

B. Description of counterfactual condition as intended

Intended components: Fathers enrolled in ES receive economic stability curriculum only. They do not receive responsible parenting and healthy relationships curricula. They also receive limited case management related to employment only for up to one year (e.g., job retention skills training, mock interviews, resume writing, employment placement services, transportation services) and limited legal services (i.e., warrant recalls, background check, and driver's license status).

Intended content: ES content includes the FFSC *Employment Readiness and Retention Training*. *Employment Readiness and Retention* is an FFSC-developed curriculum designed to focus on helping fathers who have not worked or have had poor work experiences to identify areas of growth to improve their ability to function and succeed in a work environment.

Intended dosage: ES is a four week, 80-hour program, with sessions occurring 5 days a week for 4 hours per session.

Intended delivery, target population and staff education and training were the same as described above for the intervention program. The attendance and graduation policy for ES was the same as FF: three unexcused missed days resulted in being dropped from the program and to graduate from the program, you must have attended at least 75% of the program sessions. ES implementation schedule is included in Appendix A.

C. Research Questions about the Intervention and Counterfactual Conditions as Implemented

Implementation research questions will examine dosage, acceptability and appropriateness as perceived by the participating fathers, and benefits and barriers to program implementation of the comprehensive FF (intervention group) in comparison to fathers receiving ES (comparison group). Specific implementation research questions include:

- How many families were *reached* by each of the conditions (attended, graduated, dropped out)?
- Was there *variability in retention* rates of fathers across conditions?
- Did fathers in the intervention and comparison conditions find the intervention to be *acceptable and appropriate*?
- What *benefits and barriers* exist to successful implementation of father-focused family support programs?

III. Study design

An RCT was conducted to test the Family Formation program (FF; intervention group) compared to the Economic Stability program (ES; comparison group) on identified outcomes. This section describes the recruitment, eligibility criteria, consent process, and random assignment process. In addition, this section describes the data collection approaches for both the implementation evaluation and impact evaluation including methods and measures.

A. Sample formation and research design

Recruitment. FFSC was responsible for recruiting program participants for each cohort. FFSC utilized its referral network of community based organizations in St. Louis and previous study experience to recruit fathers to the study (Avellar, Covington, Moore, Patnaik & Wu, 2018). Many enrollees heard about the program from past participants. Staff members also participated in outreach activities such as community events and distributed brochures, and spoke with members of the community about their services. Active recruitment efforts included commercials on a local TV channel and four radio stations, along with advertisements on billboards, bus shelters, and buses. To enroll, interested fathers could either walk into a participating facility or call FFSC's main office to make an appointment.

Eligibility Criteria. The eligibility criteria for men to participate in the evaluation were: (a) being a father (biological or adoptive) with at least one child 16 years old or younger; (b) no presence of a restraining order from any of his children or the mother(s) of his children; (c) not currently incarcerated (in prison); (d) currently unemployed or underemployed (i.e., working part-time but would prefer full-time, does not feel skills are fully utilized in current position, and/or does not feel that he is paid enough for his skills); (e) did not self- identify as being homeless; and (f) being at least 18 years of age. A father who was on parole or living in transitional housing (i.e., a halfway house) was eligible to participate in the study.

Consent Process. Evaluation staff obtained written consent when in-person (cohort 1) or verbal informed consent via phone (cohorts 2 – 14) for all data collection activities from fathers before administering the baseline survey and conducting random assignment. The change from enrolling in-person to via phone was to better utilize FFSC space for programmatic services. The study design and data collection plans were approved by Washington University's Human Research Protection Office for Institutional Review Board (IRB).

Random Assignment Process. After the fathers completed the baseline survey, evaluation staff randomly assigned the father to either the intervention or comparison group. Program assignment was conveyed by evaluation staff directly to the father and FFSC intake specialist. Random assignment occurred at the unit of analysis: the fathers. Fathers were randomized using an SPSS computer-generated schedule with random block sizes of four and a one-to-one allocation ratio. Following the introduction of multiple program sites in Cohort 4, randomization was stratified by program location. North County, MET Center, and SLATE locations were added during Cohort 4 to expand FFSC's services in the St Louis region and meet enrollment goals. Block

randomization was used to ensure a near balance of participants to the intervention and comparison arms throughout the study and that fathers assigned to the intervention condition were equivalent to those assigned to the comparison condition.

There was equal probability of being assigned to either the intervention or comparison group. Assignment was implemented through a Google Sheet electronic interface that concealed group assignment until each individual who consented and completed the baseline survey. To reveal the group assignment, evaluation staff first selected the worksheet tab associated with the participant's requested program location, then highlighted the next available blocked row on that sheet and selected reset from the fill color icon dropdown menu. This removed the shading and revealed the participant's program assignment, either FF or ES. Group assignment could not be changed by evaluation staff. Once revealed, evaluation staff entered the program time (day or evening class) and this with the location and program assignment generated the Participant ID # which was used in the Google sheets instead of participant names.

B. Data collection

1. Implementation analysis

Evaluation staff collected qualitative data at the 3-month follow-up survey to assess barriers to program participation and through participant focus groups to assess acceptability and appropriateness of the program for both FF and ES participants. Additionally, qualitative data were collected through staff focus groups to assess benefits and barriers to program implementation. Twelve participant focus groups were conducted (6 FF, 6 ES) with 59 fathers (23 FF, 36 ES) and 3 staff focus groups were conducted with program staff (11 – 14 staff per focus group). Attendance data were also collected and analyzed to assess dosage. Key features of data collection for the implementation analysis are summarized per implementation element and research question in Appendix B, Table B.1.

2. Impact analysis

Evaluation staff collected quantitative data for both FF and ES participants at three time points: at program enrollment and 3 and 12months post intervention. All surveys collected by the evaluation staff were administered with fathers by trained interviewers. The interviewers entered responses electronically via an online Qualtrics survey. With the exception of the Cohort 1 baseline survey that was conducted in-person, all subsequent (Cohorts 2-14) baseline and follow-up surveys were administered via phone. In addition to surveys collected by the evaluation staff, the nFORM applicant characteristics survey was administered in-person at enrollment by FFSC staff and was used for participant demographics. Key features of data collection for the impact analysis are summarized for each study group in Appendix B, Table B.2.

IV. Analysis methods

In this section we present the construction of the samples used for primary and secondary research questions, the specific measures used for each outcome, and the baseline equivalence between our intervention group (FF) and comparison group (ES) for each analytic sample. Several analytic samples were used due to the large number of outcomes analyzed at two different time points (five primary outcomes assessed at 12 months post-intervention, and five secondary outcomes assessed at 3 months post-intervention) as well as varying subsets of participants who were eligible to be assessed for each specific outcome.

A. Analytic sample

Of the 931 individuals screened for the evaluation, 214 did not meet eligibility criteria for one or more reasons. Specifically, 28 had no children 16 years or younger, 40 had a restraining order against them, 120 were not un- or under-employed, and 49 self-identified as homeless. In addition, 25 of the screened individuals declined participation. The remaining 692 fathers completed the baseline survey and were randomly assigned, 350 to FF (intervention) and 342 to ES (comparison). Baseline data collection and random assignment occurred between June 2016 and September 2018. A CONSORT diagram is included in Appendix B, Figure B1.

Our primary outcomes were based on twelve-month follow-up surveys, which began in June 2017 and concluded in December 2019. Overall, 55% (378/692) of fathers completed the 12-month follow-up survey. This includes 57% (198/350) of the FF fathers and 53% (180/342) of the ES fathers. Secondary outcomes were based on three-month follow-up surveys, which were conducted between October 2016 and March 2019. Overall, 60% (415/692) of fathers completed the 3-month follow-up survey. This includes 62% (217/350) of the FF fathers and 58% (198/342) of the ES fathers. Non-response reasons are noted in Appendix B, Figure B1.

Among the 342 fathers randomized to the ES group, two fathers mistakenly attended FF. Both of the crossovers were included in the 12-month analytic sample for primary outcomes with their original group assignment, accounting for 0.5% of the sample. Only one of these crossovers was included in the 3-month analytic sample for secondary outcomes with his original group assignment, accounting for 0.2% of the sample; the other crossover did not complete the 3-month follow-up survey.

Table IV.1 presents detailed response rates for primary outcomes measured at the 12-month follow-up survey, accounting for item non-response and restrictions due to item eligibility. For example, some items or scales were only asked among fathers who lived with or had seen their child in the past 6 months or year, or if the child was at least a certain age. Eligibility for each of the outcomes measures are clearly identified in the table footnotes. Each sample includes fathers who were eligible to answer the specific items or scales for each outcome at baseline, and who had complete outcome data at both baseline and the follow-up survey.

Overall attrition for primary outcomes ranged from 47% to 54%, and differential attrition (difference in attrition between FF and ES groups) ranged from <1 to 4 percentage points. Four of the five primary outcomes did not meet the What Works Clearinghouse (2017) guidelines for low attrition using the cautious sets of assumptions. The only primary outcome that had low attrition was *Parenting Skills* (overall attrition 50%, differential attrition 0.9 percentage points).

Because the large majority of outcomes in this RCT had high attrition, complete case analysis was used for all outcomes, baseline equivalence was assessed and statistical adjustments in analysis were made as needed following WWC recommendations (see Section C for additional details regarding the assessment of baseline equivalence).

Appendix B Table B.3 presents response rates for secondary outcomes. All five of the secondary outcomes had high attrition according to What Works Clearinghouse (2017) guidelines. The same approach for analysis was used for secondary outcomes as for primary outcomes, including using complete case analysis for all outcomes, assessing baseline equivalence, and statistical adjustments as needed.

Table IV.1. Individual sample sizes by intervention status for primary outcomes at 12-month follow-up survey

Number of individuals	Intervention (FF) sample size (numerator / denominator)	Comparison (ES) sample size (numerator / denominator)	Total sample size (numerator / denominator)	Intervention (FF) response rate	Comparison (ES) response rate	Total response rate
Assigned to condition	**350**	**342**	**692**	**n.a.**	**n.a.**	**n.a.**
Contributed a baseline survey	**350**	**342**	**692**	**100%**	**100%**	**100%**
Contributed to second follow-up survey (12 months relative to baseline)	**198 / 350**	**180 / 342**	**378 / 692**	**56.6%**	**52.6%**	**54.6%**
Contributed to second follow-up (12 months) for Father involvement: FRPN Father Engagement caregiving/play subscale	193 / 350	178 / 342	371 / 692	55.1%	52.0%	53.6%
Contributed to second follow-up (12 months) for Healthy (Co-parenting) Relationship Skills – FRPN co-parenting Alliance subscale a	178 / 347	157 / 334	335 / 681	51.3%	47.0%	49.2%
Contributed to second follow-up (12 months) for Parenting Skills: CTSPC Psychological Aggression subscale b	160 / 316	154 / 310	314 / 626	50.6%	49.7%	50.2%
Contributed to second follow-up (12 months) for Father Well-Being: Mental Health Composite Scale (MCS)	192 / 350	174 / 342	366 / 692	54.9%	50.9%	52.9%
Contributed to second follow-up (12 months) for Child Well-Being: CBCL Aggressive Behavior subscale c	109 / 229	103 / 225	212 / 454	47.6%	45.8%	46.7%

n.a. = not applicable. ES=Economic Stability. FF=Family Formation. FRPN= Father Research & Practice Network. CTSPC= Conflict Tactics Scale-Parent-Child. CBCL= Child Behavior Checklist.

[a] Co-parenting Relationship Skills were only asked if the mother was not deceased. The denominator includes fathers that met this criterion at baseline.

[b] Parenting Skills were only asked if the father lived with the child or saw the child in the past year. The denominator includes fathers that met these criteria at baseline.

[c] Child Well-Being items were only asked if the father lived with the child or saw the child in the past 6 months and the child was at least 1.5 years old. The denominator includes fathers that met these criteria at baseline.

B. Outcome measures

This study included five primary outcomes assessed at 12 months post-intervention, including *Father Involvement*, *Healthy Relationship Skills*, *Parenting Skills*, *Father Well-Being* and *Child Well-Being*. We also examined five secondary outcomes assessed at 3 months post-intervention. Three of the secondary outcomes were identical to primary outcomes (*Father Involvement*, *Healthy Relationship Skills*, *Parenting Skills*), but were assessed at 3 months post-intervention. The two additional secondary outcomes were *Financial Responsibility* and *Financial Stability*. All of these outcomes are described in more detail below.

Primary outcome measures (assessed at 12 months post-intervention)

All primary outcome measures pertaining to children (*Father Involvement*, *Healthy Relationship Skills*, *Parenting Skills*, and *Child Well-Being*) were assessed for the father's youngest child. For multi-item scales, all items must have been completed to calculate a summary score. The primary outcome measures are presented in Table IV.2, which also includes scale internal consistency measures based on data collected on fathers in this study.

Father Involvement was assessed using the Father Research & Practice Network (FRPN) caregiving/play subscale. This brief instrument is designed to assess fathers' engagement with children at different ages (Dyer, Kaufman, Fagan, Pearson, & Cabrera, 2018a). Specifically, caregiving/play sub-scale items (e.g., "How often have you praised [name of child]?") were scored on a scale of 0 to 4, 0 indicating "*never*" and 4 indicating "*every day or almost every day*." Age-specific scores were summed and then standardized into z-scores so that scores could be combined and compared over time. All age-specific caregiving/play sub-scales have been found by the scale developers to have good reliability ($\alpha \geq .92$; maximal reliability $\geq .95$) and validity (Dyer, Kaufman, Fagan, Pearson, & Cabrera, 2018a).

Healthy Relationship Skills were measured with the FRPN co-parenting relationship scale. This 11-item measure was designed to assess fathers' co-parenting relationships with the mother of their children and has shown good reliability and validity (Dyer, Fagan, Kaufman, Pearson, & Cabrera, 2018b). We used the five-item alliance subscale (mean maximal reliability found by scale developers = .94) to assess the co-parents level of collaboration. Items such as "*The mother of [name of child] and I try to understand where each other is coming from*" were scored on a scale of 1 ("*strongly disagree*") to 5 ("*strongly agree*"), and then averaged.

Parenting Skills were assessed using the Conflict Tactics Scale-Parent-Child (CTSPC), a well-established instrument used to assess physical and psychological maltreatment (Straus, Hamby, Finkelhor, Moore, & Runyan, 1998). For our primary outcome, we used the five-item psychological aggression subscale to assess the frequency of behaviors that are psychologically damaging (e.g., "*Shouted, yelled, or screamed at your child*"). The midpoint of item frequency categories (ranged from "*Never happened/not since last interview*" to "*More than 20 times*") were summed to provide a total frequency of psychological aggression incidents that occurred in

the past year (assessed at baseline) and since the last interview (assessed at 3-month and 12-month follow-up surveys).

Father Well-Being was measured with the SF-12v2 Health Survey. The 12-item nationally-normed health survey measures functional health and well-being from the patient's perspective with questions such as, "How much time in the past four weeks have you felt…calm and peaceful?" (Maruish, 2012). The SF-12v2 yields scores for eight domains of health that can be further aggregated into two component summary measures: the Physical Component Summary (PCS) and Mental Component Summary (MCS). Our primary outcome used the normalized MCS score (reliability α =.88) which ranges from 0 to 100 (worst to best mental health). MCS was calculated using licensed instrument scoring software available from Quality Metric.

Child Well-Being was assessed using the Child Behavior Checklist. This standardized instrument is based on national norms and assesses child behavioral and emotional problems (Achenbach & Rescorla, 2000; 2001). For our primary outcome, we used the age- and gender-specific aggressive behavior subscale which asks the participant to describe to what degree an item (e.g., defiant, hits others) describes his child. Total score was calculated by summing individual item scores of 0 ("*Not True*"), 1 ("*Somewhat or Sometimes True*"), or 2 ("*Very True or Often True*"). The age- and gender-specific scores were then standardized into z-scores so that scores could be combined across ages and genders and compared over time. The aggressive behavior subscale has been previously found to have good reliability (α =.92 for pre-school version; α =.94 for school-age version) and validity (Achenbach & Rescorla, 2001; Angrist, Imbens & Rubin, 1996).

Secondary outcome measures (assessed at 3 months post-intervention)

Appendix B Table B.4 describes outcome measures to answer secondary research questions. These measures examine *Father Involvement, Healthy Relationship Skills*, and *Parenting Skills* 3-months post-intervention. These three measures were calculated using the scales described in the primary outcome measure section above and were assessed regarding the father's youngest child. In addition, *Financial Responsibility* and *Financial Stability* were assessed at 3 months post-intervention. *Financial Responsibility* was measured using a single survey item asking fathers who are court-ordered to pay child support, "*Are you paying towards the child support order*?". This item was also assessed in relation to the father's youngest child. *Financial Stability* was assessed using the survey question, "*What is your current employment status?*"

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

Outcome measure	Description of the outcome measure	Timing of measure
Father Involvement	Father involvement was assessed using the Father Research & Practice Network (FRPN) Father Engagement Scale's caregiving/play subscale (Dyer, Kaufman, Fagan, Pearson, & Cabrera, 2018a). The measure's age-specific scale items (value range 0 to 4) were summed and then standardized into z-scores, which indicate the number of standard deviations each individual score is away from the mean for the relevant age group. Using z-scores allowed us to combine scores across ages and compare scores over time. Cronbach's alpha: All age groups ≥0.95	At enrollment and 12 months after end of program
Healthy (Co-parenting) Relationship Skills	Healthy relationship skills were assessed with the FRPN co-parenting measure (Dyer, Fagan, Kaufman, Pearson, & Cabrera, 2018b).The 5-item alliance subscale (measuring co-parents level of collaboration) was used. This outcome score was calculated by summing the 5 alliance subscale items and dividing by 5 to create a mean or average (value range 1 meaning *strongly disagree* to 5 meaning *strongly agree*). Cronbach's alpha: 0.93	At enrollment and 12 months after end of program
Parenting Skills	The Conflict Tactics Scale-Parent Child (CTSPC) was used to assess parenting skills/parental discipline style (Straus, Hamby, Finkelhor, Moore, & Runyan, 1998). The psychological aggression subscale was used to determine the self-reported frequency of any of the subscale's behaviors (e.g., *"Shouted, yelled, or screamed at your child"*). The midpoint of item frequency categories were summed to provide a total frequency of psychological aggression incidents, which could theoretically range from 0 to 125. Cronbach's alpha: 0.58 (Note: CTSPC items have previously been documented with suboptimal internal consistency reliability (Straus et al. 1998, Cao & Maguire-Jack 2016). This is because some of the behaviors in the scale are rare events and high alpha is not necessarily expected.)	At enrollment and 12 months after end of program
Father Well-being	The SF-12v2 Health Survey was used to measure functional health & well-being (Maruish, 2012). The Mental Health Composite Scale score (MCS) was applied and computed using licensed Quality Metric software. Scores range from 0 to 100 (worst to best health). Internal consistency: 0.85	At enrollment and 12 months after end of program
Child Well-being	Child well-being was assessed using the Child Behavior Checklist (CBCL) (Achenbach & Rescorla, 2000; 2001). The aggressive behavior subscale was used and total score was calculated by summing individual scale items (values 0="*Not True*", 1="*Somewhat or sometimes true*", 2="*Very True or Often True*"). The age- and gender-specific scores were then standardized into z-scores. Cronbach's alpha: Preschool 0.84; School-age 0.83	At enrollment and 12 months after end of program

Note: Source for all items included baseline and 12-month follow-up surveys. Cronbach's alpha was calculated for each multi-item scale using the baseline values for the analytic sample for that outcome. Internal consistency for the MCS of the SF-12V2 was calculated using the internal consistency method outlined in the SF-12v2 user's manual (Maruish, 2012; Nunnally & Bernstein, 1994).

C. Baseline equivalence and sample characteristics

Because the large majority of our outcomes had high attrition, baseline equivalence for key demographic variables was assessed for each analytic sample following WWC guidelines and best practices (What Works Clearinghouse, 2017). The analytic sample for each outcome included fathers with complete data on all baseline covariates in Table V.1, as well as complete data on the outcome at baseline and follow-up. Baseline differences were measured in standard deviations/effect sizes using Hedge's g for continuous outcomes and the Cox index for dichotomous outcomes (What Works Clearinghouse, 2017). Equivalence was examined for key demographic characteristics and baseline measures of the primary and secondary outcomes variables.

Baseline characteristics and equivalence across study groups for the *primary* outcomes are presented in Tables IV.4a through IV.4e. Here we describe the characteristics of fathers for the largest analytic sample (for the outcome measure *Father Involvement* at 12 months post-intervention (n=329), shown in Table IV.4a), but details on characteristics of the other primary outcome analytic samples can be found in Tables IV.4b through IV.4e. Almost all fathers were non-White (94% African American, 4% other/mixed race); thus, baseline equivalence was not assessed for race due to insufficient variation (only n=5 white fathers). Fathers were on average about 34 years old with two children. The average age of the father's youngest child was 6 years old. Approximately two-thirds of the fathers had never been married, and approximately three-quarters had at least a high school diploma. Around half were employed (either full-time or part-time). Approximately one-third of the fathers lived with their youngest child. Over 40% had a child support order for their youngest child. Approximately 17% had a substance use problem. There were no significant differences in these baseline characteristics between the study groups, but the effect sizes for many of the characteristics were large enough (absolute effect size >0.05 and ≤0.25) to warrant inclusion as a covariate in subsequent analyses per WWC recommendations (What Works Clearinghouse, 2017). For example, for the *Father Involvement* at 12 months post-intervention analytic sample, these necessary covariates included nearly all of the baseline characteristics, with the exception of number of children and substance use problems (effect size ≤0.05). Because most of the covariates had a baseline equivalence effect size >0.05 and ≤0.25 and all covariates were chosen because they were likely to influence the outcomes of interest, all characteristics in Table V.1 were included as covariates in the impact analysis models.

Table IV.4a. Summary statistics of key baseline measures and baseline equivalence across study groups, for fathers completing the Father Engagement Caregiving/Play subscale in the follow-up survey at 12 months post-intervention (n=329)

Baseline measure	Intervention (FF) mean (standard deviation) or %	Comparison (ES) mean (standard deviation) or %	Intervention versus comparison mean difference (p-value of difference)	Effect size
Demographics				
Father's age (years)	35.2 (9.2)	33.9 (8.4)	1.3 (0.164)	0.153
Number of children	2.4 (1.6)	2.4 (1.5)	0.0 (0.902)	-0.014
Child's age	6.2 (4.5)	5.7 (4.4)	0.5 (0.369)	0.099
Marital status (% Never married)	64%	70%	-6% (0.368)	-0.145
Education level (% ≥High school diploma/GED)	76%	74%	2% (0.726)	0.074
Employment status (% Employed)	46%	50%	-4% (0.483)	-0.109
Child support order (% yes)	41%	49%	-8% (0.178)	-0.196
Lives with child (% yes)	31%	37%	-6% (0.275)	-0.171
Substance use				
Alcohol or drug problem (% moderate to high)	17%	17%	0% (1.000)	0.021
Primary outcome				
Father involvement: FRPN Father Engagement caregiving/play subscale (standardized – Z scores)	-0.09 (0.99)	0.04 (0.99)	-0.13 (0.247)	-0.128
Sample size	168	161	n.a.	n.a.

n.a. = not applicable. ES=Economic Stability. FF=Family Formation. FRPN= Father Research & Practice Network.

Notes: Independent samples t-tests were used to compare continuous variables across groups. Pearson chi-square tests (with the Yates continuity correction for 2X2 tables) were used to compare nominal variables across groups. p-values are included in parentheses. Effect sizes are Hedge's g for continuous variables and the Cox index for nominal variables.

Race was not assessed for baseline equivalence as over 98% of the sample was non-white (94% African American, 4% other/mixed race); the sample included only 5 white fathers.

Table IV.4b. Summary statistics of key baseline measures and baseline equivalence across study groups, for fathers completing the Co-Parenting Relationship Skills Alliance subscale in the follow-up survey at 12 months post-intervention (n=300)

Baseline measure	Intervention (FF) mean (standard deviation) or %	Comparison (ES) mean (standard deviation) or %	Intervention versus comparison mean difference (p-value of difference)	Effect size
Demographics				
Father's age (years)	34.9 (9.0)	33.5 (8.0)	1.4 (0.140)	0.170
Number of children	2.4 (1.7)	2.4 (1.6)	0.0 (0.938)	-0.009
Child's age	6.0 (4.5)	5.7 (4.6)	0.3 (0.488)	0.080
Marital status (% Never married)	68%	72%	-4% (0.517)	-0.118
Education level (% ≥High school diploma/GED)	76%	75%	1% (1.000)	0.020
Employment status (% Employed)	46%	54%	-8% (0.203)	-0.195
Child support order (% yes)	41%	49%	-8% (0.158)	-0.216
Lives with child (% yes)	32%	35%	-3% (0.595)	-0.097
Substance use				
Alcohol or drug problem (% moderate to high)	17%	18%	-1% (1.000)	-0.022
Primary outcome				
Healthy (Co-parenting) Relationship Skills: FRPN co-parenting Alliance subscale (Range 1 to 5)	3.44 (1.22)	3.36 (1.25)	0.08 (0.588)	0.063
Sample size	158	142	n.a.	n.a.

n.a. = not applicable. ES=Economic Stability. FF=Family Formation. FRPN= Father Research & Practice Network.

Notes: Independent samples t-tests were used to compare continuous variables across groups. Pearson chi-square tests (with the Yates continuity correction for 2X2 tables) were used to compare nominal variables across groups. p-values are included in parentheses. Effect sizes are Hedge's g for continuous variables and the Cox index for nominal variables. Race was not assessed for baseline equivalence as 99% of the sample was non-white (94% African American, 5% other/mixed race); the sample included only 3 white fathers.

Table IV.4c. Summary statistics of key baseline measures and baseline equivalence across study groups, for fathers completing the Parenting Skills Psychological Aggression subscale in the follow-up survey at 12 months post-intervention (n=280)

Baseline measure	Intervention (FF) mean (standard deviation) or %	Comparison (ES) mean (standard deviation) or %	Intervention versus comparison mean difference (p-value of difference)	Effect size
Demographics				
Father's age (years)	34.8 (9.5)	33.4 (7.8)	1.4 (0.200)	0.153
Number of children	2.4 (1.6)	2.5 (1.5)	-0.1 (0.674)	-0.050
Child's age	5.7 (4.3)	5.3 (4.4)	0.4 (0.514)	0.078
Marital status (% Never married)	68%	71%	-3% (0.698)	-0.081
Education level (% ≥High school diploma/GED)	76%	74%	2% (0.782)	0.070
Employment status (% Employed)	47%	51%	-3% (0.550)	-0.104
Child support order (% yes)	38%	44%	-6% (0.331)	-0.161
Lives with child (% yes)	36%	42%	-6% (0.392)	-0.145
Substance use				
Alcohol or drug problem (% moderate to high)	16%	16%	0% (1.000)	0.032
Primary outcome				
Parenting Skills: CTSPC Psychological Aggression subscale (Frequency in past year)	9.99 (14.72)	8.18 (12.80)	1.81 (0.272)	0.131
Sample size	140	140	n.a.	n.a.

n.a. = not applicable. ES=Economic Stability. FF=Family Formation. CTSPC= Conflict Tactics Scale-Parent-Child.

Notes: Independent samples t-tests were used to compare continuous variables across groups. Pearson chi-square tests (with the Yates continuity correction for 2X2 tables) were used to compare nominal variables across groups. p-values are included in parentheses. Effect sizes are Hedge's g for continuous variables and the Cox index for nominal variables.
Race was not assessed for baseline equivalence as 99% of the sample was non-white (94% African American, 5% other/mixed race); the sample included only 4 white fathers.

Table IV.4d. Summary statistics of key baseline measures and baseline equivalence across study groups, for fathers completing the Father Well-Being Mental Health Composite Scale in the follow-up survey at 12 months post-intervention (n=326)

Baseline measure	Intervention (FF) mean (standard deviation) or %	Comparison (ES) mean (standard deviation) or %	Intervention versus comparison mean difference (p-value of difference)	Effect size
Demographics				
Father's age (years)	35.2 (9.2)	34.1 (8.3)	1.1 (0.230)	0.133
Number of children	2.4 (1.6)	2.4 (1.5)	0.0 (0.988)	0.002
Child's age	6.1 (4.6)	5.8 (4.5)	0.3 (0.534)	0.069
Marital status (% Never married)	65%	69%	-4% (0.578)	-0.097
Education level (% ≥High school diploma/GED)	77%	75%	2% (0.660)	0.090
Employment status (% Employed)	47%	52%	-5% (0.442)	-0.118
Child support order (% yes)	40%	49%	-9% (0.133)	-0.218
Lives with child (% yes)	32%	37%	-5% (0.325)	-0.156
Substance use				
Alcohol or drug problem (% moderate to high)	17%	17%	0% (1.000)	-0.018
Primary outcome				
Father Well-being: SF-12v2 Health Survey, Mental Health Composite Scale score (MCS) (Range 0 to 100)	48.26 (11.84)	48.52 (11.27)	-0.26 (0.840)	-0.022
Sample size	168	158	n.a.	n.a.

n.a. = not applicable. ES=Economic Stability. FF=Family Formation.

Notes: Independent samples t-tests were used to compare continuous variables across groups. Pearson chi-square tests (with the Yates continuity correction for 2X2 tables) were used to compare nominal variables across groups. p-values are included in parentheses. Effect sizes are Hedge's g for continuous variables and the Cox index for nominal variables.
Race was not assessed for baseline equivalence as 99% of the sample was non-white (94% African American, 5% other/mixed race); the sample included only 5 white fathers.

Table IV.4e. Summary statistics of key baseline measures and baseline equivalence across study groups, for fathers completing the Child Well-Being Aggressive Behavior subscale in the follow-up survey at 12 months post-intervention (n=190)

Baseline measure	Intervention (FF) mean (standard deviation) or %	Comparison (ES) mean (standard deviation) or %	Intervention versus comparison mean difference (p-value of difference)	Effect size
Demographics				
Father's age (years)	35.9 (9.1)	35.4 (7.9)	0.5 (0.671)	0.061
Number of children	2.3 (1.5)	2.3 (1.5)	0.0 (0.798)	0.037
Child's age	6.7 (4.0)	6.7 (4.2)	0.0 (0.974)	0.005
Marital status (% Never married)	65%	68%	-3% (0.757)	-0.087
Education level (% ≥High school diploma/GED)	78%	75%	3% (0.808)	0.086
Employment status (% Employed)	48%	51%	-3% (0.775)	-0.076
Child support order (% yes)	41%	50%	-9% (0.261)	-0.225
Lives with child (% yes)	36%	45%	-9% (0.273)	-0.224
Substance use				
Alcohol or drug problem (% moderate to high)	15%	17%	-2% (0.848)	-0.093
Primary outcome				
Child Well-being: CBCL Aggressive Behavior subscale (Standardized – Z scores)	0.04 (0.97)	-0.02 (0.97)	0.06 (0.712)	0.053
Sample size	98	92	n.a.	n.a.

n.a. = not applicable. ES=Economic Stability. FF=Family Formation. CBCL= Child Behavior Checklist.

Notes: Independent samples t-tests were used to compare continuous variables across groups. Pearson chi-square tests (with the Yates continuity correction for 2X2 tables) were used to compare nominal variables across groups. p-values are included in parentheses. Effect sizes are Hedge's g for continuous variables and the Cox index for nominal variables.
Race was not assessed for baseline equivalence as 99% of the sample was non-white (94% African American, 5% other/mixed race); the sample included only 2 white fathers.

V. Findings and estimation approach

A. Implementation evaluation

1. Key Findings

Overall, 38% of fathers attended at least 75% of program sessions and 37% graduated from the program. Both FF and ES fathers found the program to be acceptable and appropriate. There were several benefits and barriers to program implementation. Examples of benefits included exceeding fathers' expectations of the program, the brotherhood formed between fathers, and shared/similar backgrounds of fathers and their facilitators. Examples of barriers included employment challenges with program participation, staff turnover and training, and disappointment in program assignment for ES fathers.

Measures and Analyses. Evaluators collected and analyzed program performance data from FFSC's database for tracking program participation, to assess actual dosage received (i.e., number of families reached by each condition; the extent of retention rates variability of fathers across conditions). To assess barriers to program participation, fathers were asked at the 3-month follow up survey about any barriers to participating in their respective program (FF or ES) using an open-ended question. For analysis, a codebook was developed and responses were coded by the evaluators. Participant and staff focus groups were conducted to assess the acceptability and appropriateness of the FF and ES programs and to identify benefits and barriers to successful implementation of the programs. A thematic content analysis of the focus group data was conducted by the evaluators, identifying major themes with supporting quotes across groups per focus group question and summarizing findings (Hsieh and Shannon, 2005). For more details on focus group participation, see data collection under Section III, (*Study Design).*

Dosage. Overall, of the 692 fathers assigned to the FF and ES programs, 69% attended at least one session. Attendance of at least one session was higher among FF fathers (72%) versus ES fathers (66%), but this difference did not reach statistical significance (p=0.078). Approximately 38% of both groups attended at least 75% of program sessions, and differences by program assignment were not significant (FF 40% vs ES 35%, p=0.239). Approximately 37% graduated from the program, and differences were not significant by program assignment (FF 39% vs ES 35%, p=0.306).

Fathers received the intervention and comparison curricula as described in Section II(*Intervention and Counterfactual Conditions)*. Lessons were collapsed into topic areas, and attendance in at least one lesson for each topic area is shown in Appendix G Table G.1. Around 60% of FF and ES fathers attended at least one lesson related to job readiness or at least one lesson related to money management, and these did not differ significantly by program assignment. Significantly more FF fathers (over half) attended at least one lesson related to communication, goals/future planning, stress/anger management or health compared to ES fathers (around 1/3 to half of ES) (all $p<0.05$). Notably, ES fathers were taught these topics in the context of employment readiness. Finally, over half of FF fathers attended at least one lesson

related to parenting/fathering, relationships, making decisions and manhood. While lessons related to these topics were only offered for FF fathers, up to two ES fathers mistakenly attended some FF lessons (these crossovers were previously mentioned in Section IV.A).

Fathers who did not complete the program identified the following home or work barriers (collected at the 3-month follow-up survey) that prevented them from completing their programs (n=181, FF 95, ES 86): (a) an existing job schedule or a job schedule change conflicted with the program time (FF 32% vs ES 34%); (b) other responses (e.g., probation, family matters, mental health) (FF 17% vs ES 17%); and (c) started a new job (FF 15% vs ES 12%). About 12% of fathers from both groups experienced no home or work barriers preventing them from attending the program. There were no significant differences in types of barriers reported between FF and ES groups (all $p>0.05$).

Participant & Staff Focus Groups

The following highlight key themes from the participant and staff focus groups.

Acceptability & Appropriateness. FF and ES fathers agreed that the program met or exceeded their expectations. The content was appropriate and met their needs. Similarly, program staff felt that FFSC often meets or exceeds fathers' expectations by how they treat their participants and by the services they offer.

> *"I came to the program in hopes of getting child support modification and visitation for my son… I didn't realize all the other subjects that they help you with. It's [the program] teaching you how to be a better man, a better father. From nutrition to finances and even open house with therapy. I didn't realize all that took place with this program, and I'm definitely very happy that I've decided to come here." - FF Program Participant*

> *"What I expected was it to be solely about jobs and job readiness, but it was actually way more helpful than just that… the facilitators… broke down the whole aspect and brought it down to a level where we would know it's economic stability and life stability. And how it all coincides with each other." - ES Program Participant*

> *"FFSC, with the exception of housing, is a full service organization…When they come in with mental challenges, we have a family therapist that can assist them. When they come in with employment challenges, we have employment that can assist them…So, it makes a lot smoother transition for those guys. We are able to address a lot more issues for the client than most social service agencies." - ES Program Staff*

Fathers from both the FF and ES programs found the curriculum helpful and relevant to their lives. They gained valuable knowledge, skills and tools related to employment, financial stability, health, healthy relationships and parenting.

"The most helpful to me, I would say the budgeting sheet… I really needed that information to show me how to put my money in places where it should be, not just splurging. You spent $100 a week on food, just eating out, you can cut back on that." - ES Program Participant

"The nutritionist, she comes in and makes this food, stuff I've never had, and she takes the time to help us learn about, you know, what's good and bad and how to eat better and everything." - FF Program Participant

"[FF program staff] taught me how to actually have activities with my kids without spending money. I didn't realize like playing tag, playing Monopoly, and playing Uno … all this time we got this stuff… but we going out spending money… keep a weekend in house, and that's what I've been doing. And it turns out that I've really started learning a lot about my kids that I really did not know. Just by spending quality time with them." - FF Program Participant

Benefits and Impact of Program Participation. FF and ES fathers felt that the best part of participating in the program was the brotherhood that was formed. Fathers held each other accountable towards change. Program staff recognized the importance of the new sense of community that is formed within the classroom. The brotherhood provided accountability and moral support beyond the role of the facilitators.

"I would have to say it was my group. We went from 10 complete strangers to family. These guys are my brothers, and I love them to death… This is an environment that I wish I would have grew up in. My life probably would have been on track a long time ago, if I'd grew up in an environment like this in my own house..." - FF Program Participant

"We came together and decided we were going to treat each other like brothers, not friends. Your friend may tell you something you want to hear, your brother is going to tell you what you need. So we stayed on each other because it's hard to push a man out of his comfort zone… we can all grow, and I think that's something you don't get every day." - ES Program Participant

"… it's always 'the facilitator was really helpful,' and they develop these bonds with the men... To be able to have a support system, they are really satisfied by the end of it because usually they're calling each other brothers. The greatest thing that they get is that support system because a lot of them…might not have people who look out for them or someone to call when they are in a crisis." - FF Program Staff

Overall, FF and ES fathers appreciated that they shared similar backgrounds with their program facilitator. This helped fathers feel comfortable with opening up and is an effective model for serving this population. Fathers benefited from hearing facilitators' own stories and struggles.

"It was the facilitators giving us an ear, giving us the time…not judging us and letting us know they are normal people too and they know what we are going through, and they

may have been through it too. Once you get some people around you like that… you feel comfortable opening up to ask for help or even to receive the help." - FF Program Participant

"My facilitator made me realize that I wasn't the only one with problems… he came from where I came from. He was once also in the program and that really made me feel good… it was more than just about economic stability, it was also about parenting and about being a father… after being in the group for a week it made me reach out to my kids who I haven't talked to in a couple of years..." - ES Program Participant

"I think it goes to the facilitators coming from the same background as a lot of the clients are in. They (clients) look at them (facilitators) now and see they're employed, not doing drugs. They have relationships with their children. They can see where they came from. In some respects, that helps too." - FF Program Staff

Most FF and ES fathers gained an improved outlook on life and self-confidence, a positive change in their relationships and improved management of emotions and communication at home and work.

"I've always been fairly confident in myself but I feel like Fathers Support has given me the biggest confidence to move forward with my life and to do what I need to do for not just me but for my daughter. They just opened my eyes to a lot of stuff that I didn't even think of before I was here." - FF Program Participant

Challenges to Program Implementation. Overall, FF and ES fathers were satisfied with their program and had very few things they disliked about their program experience. Experiences did not appear to differ by program group. Most dislikes for both groups were related to program length, classroom distractions, and not receiving program incentives (gift cards) on time.

"I didn't like the fact that the program went so fast. It probably should be longer than just the four weeks" - ES Program Participant

"If I had any complaints, it's when they say they going to do something, it don't happen… The incentives that they give, they supposed to come on Monday, we don't get them until Friday." - FF Program Participant

"It's just a lot of things that's going on in our class. Maybe our facilitator could be a little stern… we've got a couple guys in our class… they ain't ready for this [playing around] the facilitator needs to get them out." - FF Program Participant

Both FF and ES fathers faced employment challenges. It was almost impossible to work full-time and participate in the program. Some fathers had to choose one or the other. Program staff recognized the importance of employment and the challenge to program participation.

"Sometimes some of them just have to get a job and it runs into the program time, and so it runs into classroom hours. So, they miss days and they get dropped." – FF Program Staff

"I'm stressed sometimes by the lack of attendance. A lot of our activities require group participation. When you have one person who shows up that day, it's pretty hard to do what you need to do with the curriculum." – ES Program Staff

Program staff turnover left a gap in services and in comradery of the team. Top reasons for staff leaving were feeling underappreciated and the pay. A lack of sufficient training was a common reason for job dissatisfaction resulting in inconsistency in program implementation.

"A lot of them I ask why they are choosing to move on. The number one thing on their list is that they don't feel appreciated… Also, the second one is pay. We have a lot of young individuals who are between the ages of 23 and 30 who end up getting their master's degrees and they do good work. But their pay doesn't increase. And so, they feel like FFSC doesn't appreciate them, isn't trying to give them a little more for their new degree." – FF Program Staff

"It's really beneficial if everyone is trained the same way so that all the services will be delivered the same way or how it should be. If you give all 3 of us the same curriculum but never tell us anything about it, our classes will get different things." – ES Program Staff

Some fathers randomly assigned to ES seemed to be disappointed. Program staff shared the fathers were not satisfied with the limited access to resources compared to the FF program. Program advertisement and recruitment information presented to participants did not have clear information about the study and ES and may have affected enrollment.

"They expect they will get all the benefits of the FFP program. But once they find out that they don't, usually more than half decide not to come or not to come at all. I can say they're usually disappointed because they're not getting that family therapy, parenting class, full access to the legal department. They're not getting everything that the program advertised." – ES Program Staff

"I just think it's doing more harm to those coming in looking for something and then getting denied. So, they're either not gonna get anything at all or they're gonna get limited access to those things. That's gonna affect the numbers, too. We could have served twice as many people if they didn't have the study in place." – ES Program Staff

Lessons Learned from the Implementation Evaluation & Key Limitations

Fathers found both the ES and FF programs to be acceptable and appropriate. However, there were several barriers to program implementation including employment challenges with program participation, staff turnover and training and, recruitment and advertisement messaging about the study and ES. Limitations for the implementation evaluation include: (a) program fidelity (fidelity measurement), (b) program retention and (c) response bias.

Program fidelity is important to determine if any unsuccessful outcomes are due to the model itself (FF or ES program) or due to failure to implement the program as originally designed.

Based on findings from the staff focus groups, the evaluators learned that due to staff turnover, it is possible not all new staff hired had enough training to adequately implement the FF and ES programs as originally designed. In response, FFSC implemented continuous quality improvement strategies to address program implementation and staffs' concerns.

Comparison of the ES and FF curriculum content show that fathers assigned to ES were not supposed to receive content on responsible parenting and healthy relationships as this content is specific to FF. However, findings from focus groups conducted with fathers suggest that parenting and father involvement were discussed in the ES condition. Because independent rating of fidelity to the two program models was outside of the scope of this evaluation, we do not have a measure of the extent to which this content was included in ES. It is possible that this may have affected study outcomes.

Employment may have affected program retention. We understand from the 3-month follow-up survey and staff focus groups that some fathers dropped out of the program or did not start because they had to work. Although FFSC provides both day and evening classes, employment continued to be a challenge to program participation. It is also important to note that recruitment and advertisement messages about the study and ES or the lack thereof may have played a role too. We learned from the staff focus groups that FFSC is a well-known organization in the St. Louis community, and many fathers come to FFSC expecting program elements of FF, for example, legal services. It is no surprise that some fathers may have been disappointed by their program assignment (ES program) and chose not to attend.

Because of the limited participation of fathers in focus groups and the 3-month follow-up survey, these fathers were not representative of all fathers who participated. Although attempts were made to recruit fathers who did not complete their program, participant focus groups were limited to a convenience sample of fathers who completed and or graduated from their program. Therefore focus group findings are biased, likely limited to those who had a more favorable view of their program experience. Not surprisingly, we only heard from a small sample of fathers who did not complete the program at the 3-month follow-up survey about their program experience, thus limiting our understanding of barriers to program completion.

B. Primary impact evaluation

1. Key findings

Our primary research questions addressed whether *Father Involvement*, *Healthy Co-Parenting Relationship Skills*, *Parenting Skills*, *Father Well-Being*, and *Child Well-Being* differed between study groups (FF vs ES) at 12 months post-intervention. While both groups experienced improvement in all primary outcomes (further described in Section D), only *Parenting Skills* differed by study group at 12 months post-intervention, with a higher number of psychological aggression occurrences since the last interview among FF fathers than ES fathers.

Adjusted mean scores for each primary outcome at 12 months post-intervention for each study group, along with adjusted mean differences between groups, are presented in Table V.2. There were no significant differences in *Father Involvement*, *Healthy Co-Parenting Relationship Skills*, *Father Well-Being*, and *Child Well-Being* between study groups at 12 months post-intervention. *Parenting Skills* did differ significantly between study groups; frequency of psychological aggression occurrences since the last interview was significantly higher in the FF group than the ES group (adjusted mean 4.3 vs 2.4, p=0.011).

An intention-to-treat (ITT) framework was used for the primary analyses, including fathers randomized to their study group regardless of their program attendance or compliance. Our analyses included fathers who provided 12-month outcome data in the impact analysis, even if they did not complete services.

To estimate intervention impacts for the primary research questions, linear regression models were used for the continuous outcomes of *Father Involvement*, *Healthy Co-parenting Relationship Skills*, *Father Well-Being*, and *Child Well-Being*. Negative binomial regression was used for the *Parenting Skills* primary outcome (frequency of psychological aggression occurrences) because this outcome was a count of occurrences and was over-dispersed (variance was greater than the mean). The dependent variable for these models was the primary outcome reported at 12-months post-intervention. Study group (FF versus ES) was included as the main independent variable of interest, and all covariates described in Table V.1 were included in the models as well. Only fathers with complete data on the outcome of interest and all baseline covariates were included in the analyses. Findings were considered statistically significant if intervention group effects had a p value <0.05. Data were analyzed using IBM SPSS Statistics and R. Equations for estimating impacts are included in Appendix E.

Table V.1. Covariates included in impact analyses[a]

Covariate	Description of the covariate
Father's age (years)	Father's age (in years) at baseline
Number of children	Number of children at baseline
Child's age	Child's age (in years) at baseline
Marital status	Father's marital status (1=never married; 0=ever married) at baseline
Education level	Father's education (1=HS grad; 0=not HS grad) at baseline
Employment status	Father's employment status (1=employed; 0=not employed) at baseline
Child support order	Father has court order for paying child support (1=yes, 0=no) at baseline
Lives with child	Father's resident status (1=lives with child; 0= does not live with child) at baseline
Substance use problem[b]	Moderate to high alcohol or drug problem (1=yes, 0=no) at baseline
Primary/secondary outcome	Baseline measure of relevant outcome

[a] All covariates were included in the model for each analytic sample.

[b] Alcohol use was assessed using the Alcohol Use Disorders Identification Test (AUDIT) (Saunders, Aasland, Babor, de la Fuente, & Grant, 1993). Drug abuse was assessed using the Drug Abuse Screening Test (DAST-10) (Skinner, 1982).

Table V.2. Post-intervention estimated effects using data from survey 12 months after the program to address the primary research questions

Outcome measure	Intervention (FF) n	Intervention (FF) Intervention mean (standard error)[a]	Comparison (ES) n	Comparison (ES) Comparison mean (standard error) [a]	Intervention compared to comparison mean difference – Adjusted b (p-value of difference)
Father Involvement: FRPN Father Engagement caregiving/play subscale (standardized – Z scores)	168	0.14 (0.06)	161	0.20 (0.06)	β = -0.06 (*p* = 0.434)
Healthy (Co-parenting) Relationship Skills: FRPN co-parenting Alliance subscale (Range 1 to 5)	158	3.63 (0.08)	142	3.64 (0.08)	β = -0.01 (*p* =0.928)
Parenting Skills: CTSPC Psychological Aggression subscale (Frequency since last interview)	140	4.29 (0.68)	140	2.36 (0.39)	β = 0.60 (*p* =0.011)
Father Well-being: SF-12v2 Health Survey, Mental Health Composite Scale score (MCS) (Range 0 to 100)	168	52.84 (0.73)	158	52.70 (0.75)	β = 0.14 (*p* = 0.894)
Child Well-being: CBCL Aggressive Behavior subscale (Standardized – Z scores)	98	-0.28 (0.09)	92	-0.34 (0.10)	β = 0.06 (*p* = 0.666)

Source: Second follow-up surveys administered 12 months after the program.

[a] Means and standard errors (not standard deviations) estimated from the regression models adjusting for covariates are presented.

[b] Difference in means between groups are reported for continuous outcomes (*Father Involvement*, *Co-parenting Relationship Skills*, *Father Well-being*, and *Child Well-being*). Difference in log count is shown for *Parenting Skills* frequency of psychological aggression. p-values are included in parentheses. All models include covariates shown in Table V.1. See Table IV.2 for a more detailed description of each measure and Chapters IV and V.B for a description of the impact estimation methods.

C. Sensitivity analyses

1. Key findings

To examine the robustness of our findings, we also present results of models that included only study group as an independent variable, excluding covariates listed in Table V.1. These results (shown in Appendix F Table F.1) indicate similar findings. The only significant difference observed between the two study groups was for *Parenting Skills*. Similar to when adjusting for covariates, when *not* adjusting for covariates the frequency of psychological aggression occurrences was higher in the FF group than the ES group (mean 6.5 vs 3.6, p=0.016).

To further explore the association between study group and psychological aggression, we examined potential influential cases. Exclusion of sixteen potentially influential cases (Cook's distance >4/n) yielded similar results (adjusted means FF 2.6, ES 1.5, β=0.552, p=0.012). We also examined this outcome as a dichotomous indicator (engaged in any psychological aggression behaviors since the last interview vs did not engage in any of these behaviors) using a logistic regression model, including all covariates. Results indicated a significantly higher probability of reporting any psychological aggression occurrences in the FF group than the ES group, holding all other covariates constant at their mean (probability 0.60 for FF group vs 0.42 for ES group, p=0.012).

D. Additional analyses

1. Key findings

Secondary Outcomes

Secondary research questions addressed whether *Father Involvement*, *Healthy Co-Parenting Relationship Skills*, *Parenting Skills*, *Financial Responsibility*, and *Financial Stability* differed between study groups at 3 months post-intervention. Mean scores for secondary outcomes at 3 months post-intervention for each study group, as well as adjusted mean differences between groups, are presented in Table V.3. None of the secondary outcomes differed significantly between the FF and ES groups at 3 months post intervention, regardless of whether baseline covariates were included in models (Table V.3) or not (Appendix G Table G.2).

Methods for analyzing secondary outcomes were similar to those for the primary outcomes. Baseline characteristics across study groups for the *secondary* outcomes are presented in Appendix C Tables C.1 through C.5. The effect sizes for many characteristics were large enough (absolute effect size >0.05 and ≤0.25) to warrant inclusion as covariates in models. Similar to our primary outcomes, we decided to include all covariates listed in Table V.1 in our models to estimate program impact on our secondary outcomes. Notably, the analytic samples for *Parenting Skills*, *Financial Responsibility*, and *Employment* as assessed at 3 months post-intervention did not meet baseline equivalence requirements. One baseline characteristic for each of these analytic samples differed between study groups with an effect size >0.25 (child support order for *Parenting Skills* and *Employment* found in Tables C.3 and C.5, baseline Financial Responsibility measure for this outcome found in Table C.4). Thus, for these outcomes, propensity score methods were used to create a matched set of fathers in the ES and FF groups which did meet baseline equivalence requirements. These matched samples were used for the impact analyses for these secondary outcomes. The propensity score matching methods are further described in Appendix C, and the baseline equivalence of the matched samples are shown in Tables C.6 through C.8.

Table V.3. Post-intervention estimated effects using data from survey 3 months after the program to address the secondary research questions

Outcome measure	Intervention (FF) n	Intervention mean (standard error)[a] or %	Comparison (ES) n	Comparison mean (standard error)[a] or %	Intervention compared to comparison mean difference – Adjusted[b] (p-value of difference)
Father Involvement: FRPN Father Engagement caregiving/play subscale (standardized – Z scores)	189	0.07 (0.05)	173	0.03 (0.05)	β = 0.04 (*p* = 0.518)
Healthy (Co-parenting) Relationship Skills: FRPN co-parenting Alliance subscale (Range 1 to 5)	177	3.55 (0.06)	159	3.62 (0.07)	β = -0.07 (*p* = 0.407)
Parenting Skills: CTSPC Psychological Aggression subscale (Frequency since last interview)[c]	131	2.38 (0.40)	131	2.41 (0.41)	β = -0.01 (*p* = 0.962)
Financial Responsibility: Among those court ordered to pay child support, % paying[c]	65	56 (86%)	65	54 (83%)	β = 0.211(*p* = 0.693)
Financial Stability: Employment status (% Employed) [c]	158	117 (74%)	158	117 (74%)	β = 0.070 (*p* = 0.803)

Source: First follow-up surveys administered 3 months after the program.

[a] Means and standard errors (not standard deviations) estimated from the regression models adjusting for covariates are presented.

[b] Difference in means between groups are reported for continuous outcomes (Father Involvement, Co-parenting Relationship Skills). Difference in log count is shown for Parenting Skills frequency of psychological aggression. Difference in log odds is shown for Financial Responsibility and Financial Stability. p-values are included in parentheses. All models include covariates shown in Table V.1. See Appendix B Table B.4 for a more detailed description of each measure and Chapters IV and V.B for a description of the impact estimation methods.

[c] Propensity score matched samples used. For more details, see Appendix C.

Treatment-on-Treated Findings

The main findings presented above for primary and secondary outcomes are based on ITT analysis. We also conducted a treatment-on-treated (TOT) analysis. Specifically, these analyses explored program effects among fathers who actually participated in FF by attending either 75% or more of FF sessions or 50% or more of FF sessions. These analyses did not show statistically significant differences between fathers with high attendance of FF sessions versus those with lower attendance for most primary outcomes. Similar to the ITT analysis, psychological aggression occurrences (Parenting Skills domain) remained significantly higher among those in the ≥75% FF attendance and ≥50% FF attendance than those attending less than 75% or 50% of FF sessions.

To conduct the TOT analysis, a complier average causal effect (CACE) of FF intervention was estimated using two-stage least squares (2SLS) regression, which uses an instrumental variable (randomized assignment to the intervention or comparison group) and jointly models the process of participation in the intervention as well as the outcome (Angrist, Imbens, & Rubin, 1996). Compliance was defined as attending ≥75% of FF sessions, which is required for graduation. For more details on the methodology see Appendix G. For the 12-month follow-up analytic sample, 50% of FF fathers attended ≥75% of the sessions. We also examined attendance of ≥50% of sessions. Approximately 57% of FF fathers attended ≥50% of sessions. Detailed results for the TOT analyses (for both ≥75% attendance and ≥50% attendance of FF sessions) can be found in Appendix G Tables G.3 and G.4.

Changes Over Time

Figures G1 and G2 in Appendix G visually depict mean scores for primary and secondary outcomes by study group and time point. Table V.4 below reports primary outcome mean scores by study group and time as well as results of mixed models assessing effects of group, time, and a group by time interaction. All primary outcomes improved over time (all time effects $p<0.01$), but these changes over time did not differ significantly between groups (all group*time effects $p>0.05$).

Appendix G Table G.5 presents results of mixed models for secondary outcomes. Similarly, most secondary outcomes improved over time. The exception was Financial Responsibility (time effect $p=0.1$), but trends showed improvement, for both groups and this analysis was limited by small sample size (only 65 per group in the propensity matched sample). Changes over time did not differ between groups for any of the secondary outcomes (all group*time effects $p>0.05$).

To examine whether primary and secondary outcomes changed over time, and whether changes over time differed between FF and ES groups, we used linear and generalized linear mixed models. For each outcome, we assessed the effects of group (FF vs ES), time (baseline to follow-up survey), and the group by time interaction. A significant group X time interaction would indicate that changes over time in the outcome differed by intervention group. For example, if the group X time interaction effect was significant for Father Involvement, it would mean that

the change in Father Involvement between baseline and the follow-up time point was different for the two groups; one of the groups would have a larger improvement over time than the other group. First, models included only group, time and group X time interaction as independent variables. Next, baseline covariates listed in Table V.1 were added to the models.

Table V.4. Mixed model results examining change in scores over time and between groups using data from baseline and survey 12 months after the program

Outcome measure	Intervention (FF)		Comparison (ES)		*Unadjusted Model Results*[a]				*Adjusted Model Results*[b]			
	T1 mean (sd)	T2 mean (sd)	T1 mean (sd)	T2 mean (sd)	n	*Group p value*	*Time p value*	*Group*Time p value*	*n*	*Group p value*	*Time p value*	*Group*Time p value*
Father involvement – FRPN Father Engagement caregiving/play subscale (standardized – Z scores)	-0.09 (0.99)	0.10 (0.91)	0.04 (0.99)	0.24 (0.98)	329	0.157	<0.001	0.804	329	0.185	<0.001	0.804
Healthy (Co-parenting) Relationship Skills – FRPN co-parenting Alliance subscale (Range 1 to 5)	3.44 (1.22)	3.66 (1.16)	3.36 (1.25)	3.61 (1.17)	300	0.617	<0.001	0.804	300	0.955	<0.001	0.804
Parenting Skills – CTSPC Psychological Aggression subscale (Frequency since last interview)	9.99 (14.72)	6.51 (12.27)	8.18 (12.80)	3.58 (6.67)	280	0.510	<0.001	0.266	280	0.542	<0.001	0.198
Father Well-being – SF-12v2 Health Survey, Mental Health Composite Scale score (MCS) (Range 0 to 100)	48.26 (11.84)	52.80 (10.79)	48.52 (11.27)	52.73 (10.29)	326	0.930	<0.001	0.795	326	0.940	<0.001	0.795
Child Well Being - CBCL Aggressive Behavior subscale (Standardized – Z scores)	0.04 (0.97)	-0.28 (1.09)	-0.02 (0.97)	-0.35 (0.99)	190	0.631	<0.001	0.913	190	0.623	<0.001	0.913

Source: Baseline survey and second follow-up survey administered 12 months after the program.

[a] Unadjusted models only include group, time, and group by time interaction.

[b] Adjusted models additionally include all covariates listed in Table V.1.

Note: See Table IV.2 for a more detailed description of each measure and Chapters IV and V.D for a description of the impact estimation methods.

VI. Discussion

The aim of the FFSC NPFF Impact Evaluation was to assess the added benefit of parenting, father-child engagement, and father well-being curriculum compared to course content containing economic stability material only on several outcomes: financial stability (employment) and responsibility (paying child support), involvement with their children, co-parenting and parenting skills. This evaluation sought to build upon the earlier Parents and Children Together (PACT) evaluation, which compared FF to a control group with no services offered (Avellar, et al., 2018). This earlier study found that compared to fathers who did not receive any services, fathers in the FF group achieved better parenting and employment outcomes at follow up (i.e., engaged in more nurturing and age appropriate activities with their children, and were employed for a longer length of time). The current study aimed to further build the evidence in support of FF by comparing it to a much more robust comparison condition – ES. Although we hypothesized that fathers and children in the FF group would demonstrate better outcomes than the ES group across multiple domains, in the current evaluation outcomes between the FF and ES groups, in general, were not significantly different.

Although a direct comparison cannot be made because different measures were used in the PACT and NPFF evaluations, some important information can be surmised when comparing outcomes of the two studies. Participants in the same program, FF, had significantly better outcomes than fathers receiving no services (in the PACT evaluation). However, when FF was compared to ES in the current NPFF study very few significant differences were found. This suggests that training in economic stability alone – without the additional fathering-focused content –may have an impact on parenting outcomes. Hence, we saw no significant differences between the two groups on father involvement outcomes. Likewise, the unique lessons included in FF may not contribute to economic stability and therefore we saw no significant differences in financial stability outcomes. The shorter ES training, therefore, may be the more economical and efficient strategy to improve outcomes across parenting and economic stability.

The one outcome with significantly different outcomes was the parenting skill of psychological aggression. Both groups experienced reductions in psychological aggression incidents after the intervention, but the frequency was significantly higher among fathers in the FF group than the ES group at 12 months post-intervention. This finding is in an unexpected direction. The increased emphasis on responsible fatherhood and healthy relationships in the FF curriculum may have caused these fathers to be more aware of these negative parenting behaviors, leading to increased reporting among this group. Additional research is warranted to better understand this relationship.

Despite the finding of no difference in most outcomes between the intervention and comparison conditions, the evaluation did yield some other important results. Findings from the qualitative interviews indicated that both FF and ES fathers had positive experiences with the program. Fathers appreciated the empathetic staff who understood their struggles and felt that the

programs provided a supporting community of fathers as well as valuable knowledge and skills to help gain employment and improve their relationships and parenting skills.

An important observation is that fathers in both groups demonstrated gains in employment, increasing from less than half employed at baseline to approximately 3/4 employed at three months post-intervention. We are unable to say whether program participation (in ES or FF) caused the employment gains because the evaluation design did not compare program participants to a control group that did not receive any program services. However, this finding is promising. Teaching skills for gaining employment was a core component of the programming for both groups. It is possible that financial stability could have influenced improvements in other fatherhood outcomes. For example, psychological well-being is known to be higher among employed versus unemployed individuals (McKeen-Ryan, Song & Kinicki, 2005). Coley and Morris (2002) found that employed low-income minority fathers had higher levels of involvement with their children. Furthermore, Bronte-Tinkew and Horowitz (2010) documented that a more supportive coparental relationship is perceived among nonresident fathers who are employed. In addition, prior research also suggests that fathers with stable employment may be more likely to pay child support, which could impact their relationship with the child. Turner and Waller (2017) found that nonresident fathers with child support arrears saw their children less and engaged in activities with them less frequently. Given the simultaneous improvements in many outcomes along with employment status in this study, disentangling these relationships warrants further exploration. In addition, the lack of differences between groups in parenting and co-parenting outcomes could have been influenced by some of the employment readiness and retention components of the economic stability curriculum, such as communication and active listening, and stress and anger management.

Several challenges were encountered during the implementation of the program. Program completion among fathers in both groups was relatively low. Some fathers indicated that it was difficult to attend sessions if they conflicted with their work schedules. Some fathers were dropped from the program due to the program's attendance policy (three unexcused missed days resulted in being dropped from the program). Others might have dropped out of the program when they were able to obtain employment, thus not feeling the need to complete the program or not having enough time or the social support to work, parent and complete the program. Some fathers may have dropped out of the program due to the ES program assignment which may not meet fathers' specific needs. Child-related legal services (child support modification, visitation and or custody) was an expressed need for some fathers assigned to ES which only offered limited legal services. Other participant characteristics, such as lower education and lower income, could have also contributed to lower program completion (Laxman, Higginbothman, & Bradford, 2019). In addition, staff-related challenges might have impacted program implementation and fidelity. Staff turnover, common among non-profit organizations, and insufficient training were issues that could have contributed to gaps in services and inconsistencies in program implementation.

Results and lessons learned from this implementation and impact evaluation can help inform FFSC's programming and services to support low-income fathers. Programming for both groups focused on employment and included some level of legal services, but differed in the dose (length) of programming with a greater emphasis on responsible fatherhood, healthy relationships, and case management services in the FF group. That there was no significant differences in outcomes between the two groups along with the challenges in program completion suggest that shortening the length of FF could be explored. However, given that legal services and case management services are often vital services sought by fathers, these components should likely remain in the program. Alternative modes for programming could be explored such as online or make-up sessions where fathers can complete their respective program at their own leisure. This allows for flexibility for employed fathers or fathers with less social support to complete the program. In addition, exploring alternative program incentives (e.g., increasing dollar amount or providing cash instead of gift-cards) for their participation may also increase program completion, allowing for unemployed fathers to have the means necessary to provide for themselves and their families while completing the 4 – 6 week program. To address staff turnover and program fidelity, it may be helpful to explore employee retention strategies for social service agency workers to ensure retention of strong and committed workers to fatherhood programs.

Several limitations should be considered when interpreting the results of this study. First, in order to identify the key promising element(s) of responsible fatherhood programs, this study included an active control group rather than a "no treatment" control. It was judicious to provide key employment-focused programming to all fathers, regardless of study group, and test whether receiving a higher dose of programming and components focused on responsible fatherhood and healthy relationships would have an additional impact on outcomes. As discussed above, similarity in programming with a focus on employment in both groups could have positively impacted outcomes for both groups, contributing to the lack of differences in outcomes between groups. Notably, a prior study of the FFSC programming did find statistically significant effects on parenting, co-parenting, and well-being by comparing fathers receiving services to a "no treatment" control group (Avellar, Covington, Moore, Patnaik, & Wu, 2018). High attrition in follow-up surveys was another limitation of the current study. While in-person follow-up surveys (i.e., meeting with the father at FFSC or in their home versus by phone), which might have helped reduce attrition, were considered, in-person survey administration was not feasible due to funding constraints. While statistical procedures were conducted to help address possible biases due to high attrition per WWC recommendations (What Works Clearinghouse, 2017), these procedures do not eliminate all bias. High attrition also contributed to diminished power to detect differences between study groups. With only 378 fathers at 12 month follow-up, this study had 80% power to detect an effect size of 0.3 or larger in the difference in means between study groups (Faul, Erdfelder, Lang, & Buchner, 2007). Finally, while the perceived quality of programming was assessed via qualitative interviews with fathers and staff, there were limited data addressing the actual fidelity of FF and ES program implementation. There is also some

evidence in our qualitative data that parenting content was discussed in the ES group; these discussions very likely had an impact on evaluation findings.

Although both groups experienced improvements in outcomes over time, FF fathers did not experience better outcomes than ES fathers. The lack of difference in outcomes could be due to a similar core focus on employment-related curriculum for both groups. It is possible that gaining financial stability contributed to positive improvements in other fatherhood domains, and future research should focus on further understanding the links between fathers gaining employment and their parenting skills, child engagement, and well-being. In addition, future research should continue to explore which specific components of responsible fatherhood programs are most needed and provide the greatest benefit for fathers and children.

VII. References

Achenbach, T.M., & Rescorla, L.A. (2000). *Manual for the ASEBA Preschool Forms and Profiles.* Burlington, VT: University of Vermont, Research Center for Children, Youth & Families.

Achenbach, T.M., & Rescorla, L.A. (2001). *Manual for the ASEBA School-age Forms and Profiles.* Burlington, VT: University of Vermont, Research Center for Children, Youth, & Families.

Angrist J. D., Imbens G. W., & Rubin, D. B. (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association, 91,* 444–455

Austin, P. C. (2011). Optimal caliper widths for propensity-score matching when estimating differences in means and differences in proportions in observational studies. *Pharm Stat, 10*(2), 150-161. doi:10.1002/pst.433.

Avellar, S., Covington, R., Moore, Q., Patnaik, A., & Wu, A. (2018). Parents and Children Together: Effects of four responsible fatherhood programs for low-income fathers. *Mathematica Policy Research.*

Bronte-Tinkew, J., & Horowitz, A. (2010). Factors Associated With Unmarried, Nonresident Fathers' Perceptions of Their Coparenting. *Journal of Family Issues, 31*(1), 31- 65. doi: 10.1177/0192513X09342866

Bronte-Tinkew, J., Burkhauser, M., & Metz, A. J. R. (2012). Elements of promising practices in fatherhood programs: Evidence-based research findings on interventions for fathers. *Fathering, 10*(1), 6-30. doi:10.3149/fth.1001.6

Cao Y, Maguire-Jack K. Interactions with community members and institutions: Preventive pathways for child maltreatment. Child Abuse & Neglect 62 (2016) 111–121. http://dx.doi.org/10.1016/j.chiabu.2016.10.012

Centers for Disease Control and Prevention (CDC). (2018a). Sexually Transmitted Disease Surveillance Table 9. Chlamydia – Reported Cases and Rates of Reported Cases in Counties and Independent Cities Ranked by Number of Reported Cases. Atlanta: U.S. Department of Health and Human Services; 2019. doi: 10.15620/cdc.79370

Centers for Disease Control and Prevention (CDC). (2018b). Sexually Transmitted Disease Surveillance Table 20. Gonorrhea – Reported Cases and Rates of Reported Cases in Counties and Independent Cities Ranked by Number of Reported Cases. Atlanta: U.S. Department of Health and Human Services; 2019. doi: 10.15620/cdc.79370

Coley, R. L., & Morris, J. E. (2002). Comparing Fathers and Mother Reports of Father Involvement Among Low-Income Minority Families. *Journal of Marriage and Family, 64*(4), 982-997. doi: 10.1111/j.1741-3737.2002.00982.x

Cowan PA, Cowan CP, Pruett MK, Pruett K, Wong JJ. (2009). Promoting fathers' engagement with children: preventive interventions for low-income families. *Journal of Marriage & Family, 71*, 663–79. doi: 10.1111/cdev.12744

Dyer, W. J., Kauffman, R., Fagan, J., Pearson, J. and Cabrera, N. (2018a). Measures of Father Engagement for Nonresident Fathers. *Family Relations: An Interdisciplinary Journal of Applied Family Studies 67*(3)*,* 381-398. doi:10.1111/fare.12317

Dyer, W. J., Fagan, J., Kaufman, R., Pearson, J. and Cabrera, N. (2018b). Self-perceived Coparenting of Nonresident Fathers: Scale Development and Validation. *Family Process Institute, 57,* 927-946. doi:10.1111/famp.12331

Elze, M. C., Gregson, J., Baber, U., Williamson, E., Sartori, S., Mehran, R., . . . Pocock, S. J. (2017). Comparison of propensity score methods and covariate adjustment: Evaluation in 4 cardiovascular studies. *Journal of the American College of Cardiology, 69*(3), 345-357.

Fagan, J., & Iglesias, A. (1999). Father involvement program effects on fathers, father figures, and their head start children: A quasi-experimental study. *Early Childhood Research Quarterly*, *14*(2), 243-269. http://dx.doi.org/10.1016/S0885-2006(99)00008-3

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*, 175- 191. doi:10.3758/BF03193146

Guo, S., & Fraser, M. W. (2015). *Propensity score analysis: Statistical methods and applications* (Vol. 11). Thousand Oaks, CA: SAGE Publications.

Hsieh, H., & Shannon S. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research*. 15(9):1277-1288.

Jiang, Y. & Small, D. (2014). ivpack: Instrumental Variable Estimation. R package version 1.2. Retrieved from https://CRAN.R-project.org/package=ivpack

Barthelemy, J. J., & Coakley, T. M. (2017). Fathering attitudes and behaviors among low-income fathers, Journal of Family Social *Work*, 399- 415. doi: 10.1080/10522158.2017.1302379

Laxman, D. J., Higginbotham, B. J., & Bradford, K. (2019) Predictors of attrition and attendance in a fatherhood education program. *Children and Youth Services Review, 103*, 287- 297. doi:10.1016/j.childyouth.2020.104886

Lundahl, B., Risser, H.J., & Lovejoy, M.C. (2006). A meta-analysis of parent training: Moderators and follow-up effects. *Clinical Psychology Review, 26,* 86-104. doi: 10.1016/j.cpr.2005.07.004

Maruish, M. E. (Ed.). (2012). *User's manual for the SF-12v2 Health Survey* (3rd ed.). Lincoln, RI: QualityMetric Incorporated.

Maruish, M. E. (Ed.). (2012). *User's manual for the SF-12v2 Health Survey* (3rd ed.). Lincoln, RI: Quality Metric Incorporated.

McKee-Ryan, F., Song, Z., Wanberg, C., & Kinicki, Angelo. (2005). Psychological and Physical Well-Being During Unemployment: A Meta-Analytic Study. *The Journal of Applied Psychology, 90*(1)*,* 53-76. doi:10.1037/0021-9010.90.1.53

Missouri Department of Health and Senior Services (DHSS). (2018). Death MICA.MOPHIMS (Missouri Public Health Information Management Systems). Retrieved from https://healthapps.dhss.mo.gov/MoPhims/QueryBuilder?qbc=FM&q=1&m=1

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York: McGraw-Hill.

Rienks, S.L., Wadsworth, M.E., Markman, H.J., Einhorn, L., & Etter, E. M. (2011). Father involvement in urban low-income fathers: Baseline associations and changes resulting from preventive intervention. *Family Relations*, *60*(2), 194-204. http://dx.doi.org/10.1111/j.1741-3729.2010.00642.x

Robbers, M. P. (2009). Facilitating fatherhood: A longitudinal examination of father involvement among young minority fathers. *Child & Adolescent Social Work Journal*, *26*, 121–134. doi:10.1007/s10560-008-0157-6

Saunders JB, Aasland OG, Babor TF, de la Fuente JR and Grant M. (1993) Development of the Alcohol Use Disorders Identification Test (AUDIT): WHO Collaborative Project on Early Detection of Persons with Harmful Alcohol Consumption II. *Addiction 88*(6), 791-804. doi:10.1111/j.1360-0443.1993.tb02093.x

Sjolander, A., Dahlqwist, E., & Martinussen, T. (2020). ivtools: Instrumental Variables. R package version 2.3.0. Retrieved from https://CRAN.R-project.org/package=ivtools

Skinner, H. A. (1982). The Drug Abuse Screening Test. *Addictive Behavior*, *7*(4), 363–371. doi: 10.1016/0306-4603(82)90005-3

Straus, M.A., Hamby, S.L., Finkelhor, D., Moore, D.W., & Runyan, D. (1998). Identification of child maltreatment with the parent-child conflict tactics scales: Development and psychometric data for a national sample of American parents. *Child Abuse & Neglect, 22*(4), 249-270. http://dx.doi.org/10.1016/S0145-2134(97)00174-9

Stuart, E. A., King, G., Imai, K., & Ho, D. (2011). MatchIt: nonparametric preprocessing for parametric causal inference. *Journal of statistical software*.

Turner, K. J., & Waller, M. R. (2017). Indebted Relationships: Child Support Arrears and Nonresident Fathers' Involvement with Children. *Journal of Marriage and Family, 79*(1), 24- 43. https://doi.org/10.1111/jomf.12361

United States (U.S.) Census Bureau. (2014-2018a). American Community Survey 5-Year Estimates. Table S1201; generated using data census.gov < https://data.census.gov/cedsci/>; (11 May 2020).

United States (U.S.) Census Bureau. (2014-2018b) American Community Survey 5-Year Estimates. Table S1701; generated using data.census.gov < https://data.census.gov/cedsci/>; (11 May 2020).

United States (U.S.) Census Bureau. (2014-2018c) American Community Survey 5-Year Estimates. Table S2301; generated using data.census.gov <https://data.census.gov/cedsci/ >; (11 May 2020).

United States (U.S.) Department of Justice (DOJ), Federal Bureau of Investigation. (2019, September). Crime in the United States, 2018. Table 6. Retrieved from https://ucr.fbi.gov/crime-in-the-u.s/2018/crime-in-the-u.s.-2018/topic-pages/tables/table-6

What Works Clearinghouse, Institute of Education Sciences, U.S. Department of Education. (2017, October). *What Works Clearinghouse: Standards Handbook (Version 4.0)*. Retrieved from http://whatworks.ed.gov

VIII. APPENDIX TABLES AND FIGURES TO SUPPLEMENT FINAL IMPACT REPORT

Appendix A: Logic Model & Implementation Schedules

Logic Model: Fathers & Families Support Center St. Louis, Inc. – Family Formation (FF) Program

Need: Low levels of healthy family functioning, adult and child well-being and economic stability, and high rates of poverty

Program Goal	***Activities***	***Outputs/Impl. Objectives Each Year 2-5***	***Short-term Outcomes***	***Long-Term Outcomes***
Implement an integrated parenting, healthy marriage and economic stability and mobility program for 580 econo- mically disadvantaged fathers in Saint Louis Missouri to: strengthen positive father-child engagement, improve social and economic outcomes to increase the well-being and economic stability of adults and children. ***Inputs*** • **FFSC**: BOD, CEO, PD, PC, SW, ES, FAC • **Partners**: MM, FWCA, SLATE, MET • Experienced and well-trained **staff** • Social marketing, recruitment, referrals • Policies, procedures • **Evaluation**: Washington University, AMTC **Budget:** $2,000,000	**Curricula** • *Family Formation (includes parenting and co-parenting support)* • *Money Smarts* • *Empl. Readiness, Retn. Within My Reach* **Employment** HiSet/ GED, job training, job placement, retention, 12 months follow up **Case Management** • Foster parenting, co-parenting, job and relationship, skills • Assess for trauma, ACE • *Helping Men Recover* • Family Mediation **Program Processes** monthly reports, data & evaluation tracking & collection; staff observations, ongoing improvements.	**Obj. 1:** 350 fathers recv. *Family Formation* (FF) **Obj. 2**: 75 young fathers receive abbrev. FFC. **Obj. 3:** 75 dislocated fathers receive abbr. FFC. **Obj. 4:** 80 incarcerated fathers receive abbr. FFC. **Obj. 5:** 80 incar. fathers receive WMR **Obj. 6:** 40 incar. fathers receive CM, mediation **Obj. 7:** 425 fathers (obj. 1) assessed for trauma **Obj. 8:** 425 fathers-objs. 1-3- recv. CM/ mediatn. **Obj. 9:** 425 fathers - objs. 1-3- recv. employ. readiness services /skills **Obj. 10** 345 job placems. **Obj. 11.** 50 fathers dev. co-parent. agr. w/mother	***Relationship / Marriage*** *1. increase in co-parenting skills; alliance* ***Parenting*** *2. increase in father involvement; caregiving/play* *3. decrease in psych. aggression* ***Econ. Stability, Mobility*** *4. increase in financial responsibility; paying towards child support* *5. increase in employment; gaining employment*	1. improved family functioning 2. improved father and child well-being ***Assumptions*** 1. Target parents have positive attitude re: improving couple relationships, parenting and co-parenting. 2. Every father is a responsible father committed to cohesive family relationships. 3. Target parents are open to improving their and child's well-being. 4. Target parents understand that in order to be a responsible parent, they must obtain and maintain economic stability and growth.

Inputs: **FFSC**: BOD – Board of Directors, CEO – Chief Executive Officer, PD – Program Director, SW – Social Worker/Case Manager, ES – Employment Stability Case Manager, FAC – Facilitators; **Partners**: MM – Manasseh Ministries, FWCA – Family Workforce Centers of America, SLATE – St. Louis Agency on Training and Employment, MET – Metropolitan Education and Training Center


Fathers & Families
SUPPORT CENTER
HELP FOR FATHERS. HOPE FOR FAMILIES.

Appendix B: Data, Sample, and Measures

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

Implementation element	Research question	Data source	Timing/frequency of data collection	Party responsible for data collection
Dosage	How many families were reached by each of the conditions? Was there variability in retention rates of fathers across conditions?	Workshop sessions in nFORM and Apricot; attendance logs	All sessions delivered, entered by FSC Facilitators	FSC & AMTC
Acceptability and Appropriateness	Did fathers in the intervention and comparison conditions find the intervention to be acceptable and appropriate?	Participant focus groups; transcripts Telephone Survey; program participation barriers	Starting with Cohort 4, every other cohort per condition, per year (2017, 2018), a total of 12 participant focus groups, 59 participants, conducted by the Evaluator, 90 minutes each at FSC 3-months after the intervention	WashU
Benefits and Barriers	What benefits and barriers exist to successful implementation of father-focused family support programs?	Staff focus groups; transcripts	One focus group per year (2016, 2017, 2019), a total of 3 staff focus groups, approx.18 staff members, conducted by the Evaluator, 90 minutes each at FSC	AMTC

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

	Data source	Timing of data collection	Mode of data collection	Party responsible for data collection	Start and end date of data collection
Intervention	Intervention group fathers	• Enrollment (baseline) • 3 months after the end of program • 12 months after the end of the program	Telephone survey	Evaluation staff	June 2016 through December 2019
		Enrollment	In-person online survey (nFORM)	Program staff	June 2016 through September 2018
Counterfactual	Comparison group fathers	• Enrollment (baseline) • 3 months after the end of program • 12 months after the end of the program	Telephone survey	Evaluation staff	June 2016 through December 2019
		Enrollment	In-person online survey (nFORM)	Program staff	June 2016 through September 2018

Figure B1. CONSORT diagram

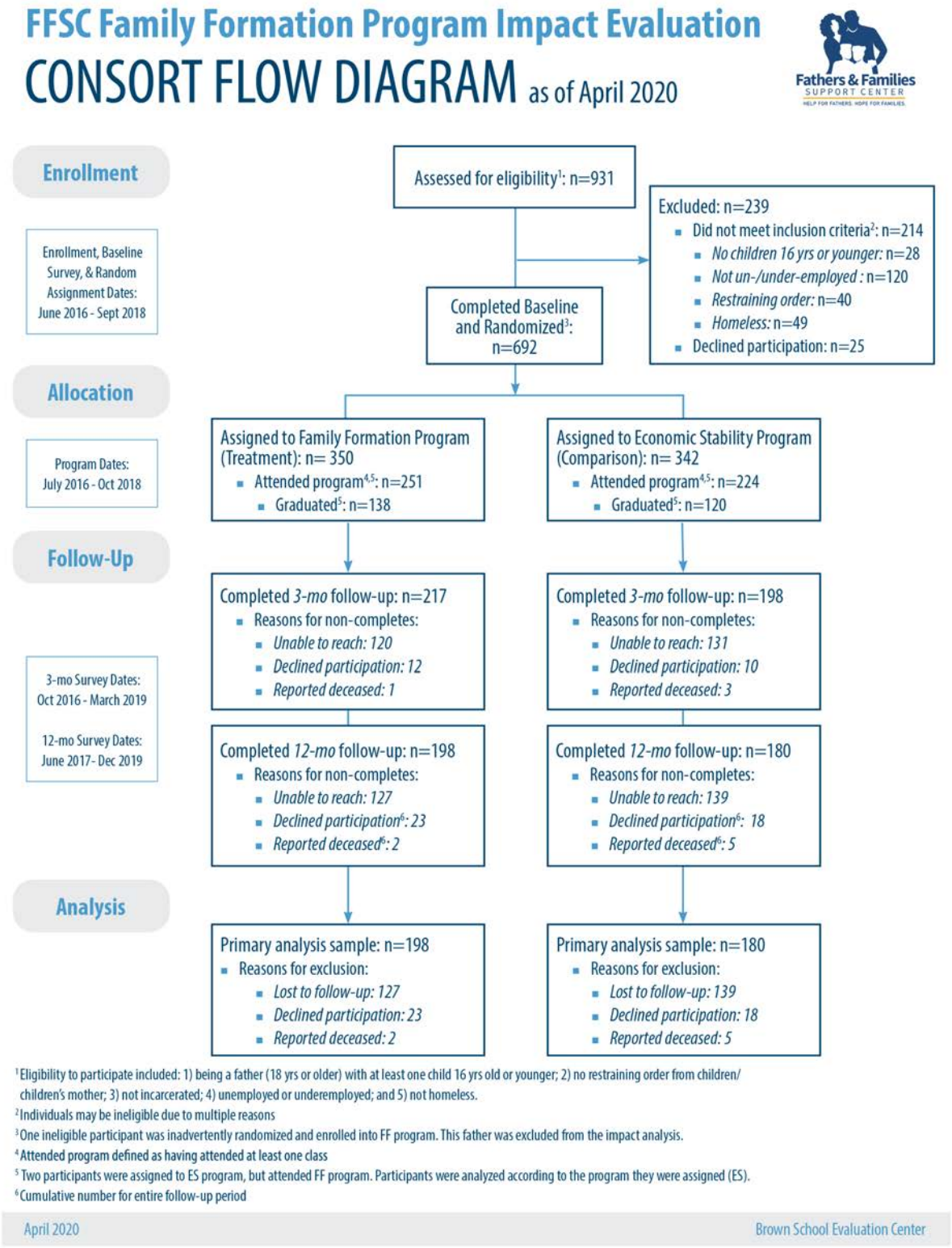

Table B.3. Individual sample sizes by intervention status for secondary outcomes

Number of individuals	Intervention (FF) sample size (numerator / denominator)	Comparison (ES) sample size (numerator / denominator)	Total sample size (numerator / denominator)	Intervention (FF) response rate	Comparison (ES) response rate	Total response rate
Assigned to condition	**350**	**342**	**692**	**n.a.**	**n.a.**	**n.a.**
Contributed a baseline survey	**350**	**342**	**692**	**100%**	**100%**	**100%**
Contributed to first follow-up survey (3 months relative to baseline)	**217 / 350**	**198 / 342**	**415 / 692**	**62.0%**	**57.9%**	**60.0%**
Contributed to first follow-up (3 months) for Father involvement: FRPN Father Engagement caregiving/play subscale	217 / 350	194 / 342	411 / 692	62.0%	56.7%	59.4%
Contributed to first follow-up (3 months) for Healthy (Co-parenting) Relationship Skills – FRPN co-parenting Alliance subscale [a]	200 / 347	177 / 334	377 / 681	57.6%	53.0%	55.4%
Contributed to first follow-up (3 months) for Parenting Skills: CTSPC Psychological Aggression subscale [b]	177 / 316	167 / 310	344 / 626	56.0%	53.9%	55.0%
Contributed to first follow-up (3 months) for Financial Responsibility: Paying court ordered child support [c]	80 / 140	87 / 146	167 / 286	57.1%	59.6%	58.4%
Contributed to first follow-up (3 months) for Financial Stability: Employment status	210 / 350	195 / 342	405 / 692	60.0%	57.0%	58.5%

n.a. = not applicable. ES=Economic Stability. FF=Family Formation. FRPN= Father Research & Practice Network. CTSPC= Conflict Tactics Scale-Parent-Child.

[a] Co-parenting Relationship Skills were only asked if the mother was not deceased. The denominator includes fathers that met this criterion at baseline.

[b] Parenting Skills were only asked if the father lived with the child or saw the child in the past year. The denominator includes fathers that met these criteria at baseline.

[c] Financial Responsibility (paying child support) was only asked among fathers who were court ordered to pay child support. The denominator includes fathers that met this criterion at baseline.

Table B.4. Outcome measures used for secondary impact analyses research questions

Outcome measure	Description of outcome measure	Timing of measure
Father Involvement	Father involvement was assessed using the Father Research & Practice Network (FRPN) Father Engagement Scale's caregiving/play subscale (Dyer, Kaufman, Fagan, Pearson, & Cabrera, 2018a). The measure's age-specific scale items (value range 0 to 4) were summed and then standardized into z-scores, which indicate the number of standard deviations each individual score is away from the mean for the relevant age group. Using z-scores allowed us to combine scores across ages and compare scores over time. Cronbach's alpha: All age groups ≥0.94	At enrollment and 3 months after end of program
Healthy (Co-parenting) Relationship Skills	Healthy relationship skills were assessed with the FRPN co-parenting measure (Dyer, Fagan, Kaufman, Pearson, & Cabrera, 2018b).The 5-item alliance subscale (measuring co-parents level of collaboration) was used. This outcome was calculated by summing the 5 alliance subscale items and dividing by 5 to create a mean or average (value range 1 meaning *strongly disagree* to 5 meaning *strongly agree*). Cronbach's alpha: 0.93	At enrollment and 3 months after end of program
Parenting Skills	The Conflict Tactics Scale-Parent Child (CTSPC) was used to assess parenting skills/parental discipline style (Straus, Hamby, Finkelhor, Moore, & Runyan, 1998). The psychological aggression subscale was used to determine the self-reported frequency of any of the subscale's behaviors (e.g., *"Shouted, yelled, or screamed at your child"*). The midpoint of item frequency categories were summed to provide a total frequency of psychological aggression incidents, which could theoretically range from 0 to 125. Cronbach's alpha: 0.62 (Note: CTSPC items have previously been documented with suboptimal internal consistency reliability (Straus et al. 1998, Cao & Maguire-Jack 2016). This is because some of the behaviors in the scale are rare events and high alpha is not necessarily expected.)	At enrollment and 3 months after end of program
Financial Responsibility	Survey items related to child support and other support were used to evaluate this construct. The outcome measure was a yes/no response taken directly from the question, "Are you paying towards the child support order?" (Yes=1, No=0)	At enrollment and 3 months after end of program
Financial Stability	Applicant characteristic and 3-month survey items related to employment status were used to evaluate this construct. The outcome measure was taken directly from the question, "What is your current employment status?" Responses were then dichotomized into 1=Employed (including Full-time 35 or more hours a week, Part-time 1-34 hours a week, Employed but number of hours changes from week to week, and Temporary, occasional, or seasonal employment), and 0=Not currently employed	At enrollment and 3 months after end of program

Note: Source for all items included baseline and 3-month follow-up surveys. Cronbach's alpha was calculated for each multi-item scale using the baseline values for the analytic sample for that outcome.

Appendix C: Baseline Equivalence of the RCT Design for Secondary Outcomes

Baseline Equivalence for Secondary Outcomes

Baseline equivalence was tested for all secondary outcome analytic samples (Tables C.1 through C.5). While all baseline covariates for *Father Involvement* and *Co-Parenting Skills* analytic samples met baseline equivalence requirements (difference between groups effect size ≤0.25), this was not true for *Parenting Skills*, *Financial Responsibility*, and *Employment*. One baseline characteristic for each of these analytic samples differed between study groups with an effect size >0.25 (child support order for *Parenting Skills* and *Employment*, baseline *Financial Responsibility* measure for this outcome) (see Tables C3, C4, and C5).

Table C.1. Summary statistics of key baseline measures and baseline equivalence across study groups, for fathers completing the Father Engagement Caregiving/Play subscale in the follow-up survey at 3 months post-intervention (n=362)

Baseline measure	Intervention (FF) mean (standard deviation) or %	Comparison (ES) mean (standard deviation) or %	Intervention versus comparison mean difference (p-value of difference)	Effect size
Demographics				
Father's age (years)	34.8 (9.2)	34.2 (8.4)	0.6 (0.480)	0.074
Number of children	2.3 (1.6)	2.5 (1.6)	-0.2 (0.453)	-0.079
Child's age	6.1 (4.5)	5.8 (4.6)	0.3 (0.466)	0.077
Marital status (% Never married)	65%	69%	-4% (0.524)	-0.102
Education level (% ≥High school diploma/GED)	73%	75%	-2% (0.830)	-0.049
Employment status (% Employed)	48%	49%	-1% (0.935)	-0.024
Child support order (% yes)	42%	52%	-10% (0.082)	-0.237
Lives with child (% yes)	35%	38%	-3% (0.597)	-0.084
Substance use				
Alcohol or drug problem (% moderate to high)	14%	13%	1% (0.904)	0.051
Secondary outcome				
Father involvement: FRPN Father Engagement caregiving/play subscale (standardized – Z scores)	-0.07 (1.00)	-0.03 (0.97)	-0.04 (0.688)	-0.042
Sample size	189	173	n.a.	n.a.

n.a. = not applicable. ES=Economic Stability. FF=Family Formation. FRPN= Father Research & Practice Network.

Notes: Independent samples t-tests were used to compare continuous variables across groups. Pearson chi-square tests (with the Yates continuity correction for 2X2 tables) were used to compare nominal variables across groups. p-values are included in parentheses. Effect sizes are Hedge's g for continuous variables and the Cox index for nominal variables.
Race was not assessed for baseline equivalence as 98% of the sample was non-white (94% African American, 4% other/mixed race); the sample included only 6 white fathers.

Table C.2. Summary statistics of key baseline measures and baseline equivalence across study groups, for fathers completing the Co-Parenting Relationship Skills Alliance subscale in the follow-up survey at 3 months post-intervention (n=336)

Baseline measure	Intervention (FF) mean (standard deviation) or %	Comparison (ES) mean (standard deviation) or %	Intervention versus comparison mean difference (p-value of difference)	Effect size
Demographics				
Father's age (years)	34.9 (9.2)	33.7 (8.2)	1.2 (0.207)	0.137
Number of children	2.3 (1.6)	2.5 (1.6)	-0.2 (0.481)	-0.077
Child's age	6.1 (4.5)	5.7 (4.7)	0.4 (0.350)	0.102
Marital status (% Never married)	66%	71%	-5% (0.389)	-0.140
Education level (% ≥High school diploma/GED)	72%	74%	-2% (0.890)	-0.039
Employment status (% Employed)	47%	50%	-3% (0.765)	-0.054
Child support order (% yes)	42%	52%	-10% (0.115)	-0.224
Lives with child (% yes)	34%	38%	-4% (0.536)	-0.101
Substance use				
Alcohol or drug problem (% moderate to high)	15%	13%	2% (0.815)	0.075
Secondary outcome				
Healthy (Co-parenting) Relationship Skills: FRPN co-parenting Alliance subscale (Range 1 to 5)	3.38 (1.20)	3.40 (1.23)	-0.02 (0.863)	-0.019
Sample size	177	159	n.a.	n.a.

n.a. = not applicable. ES=Economic Stability. FF=Family Formation. FRPN= Father Research & Practice Network.

Notes: Independent samples t-tests were used to compare continuous variables across groups. Pearson chi-square tests (with the Yates continuity correction for 2X2 tables) were used to compare nominal variables across groups. p-values are included in parentheses. Effect sizes are Hedge's g for continuous variables and the Cox index for nominal variables.

Race was not assessed for baseline equivalence as 99% of the sample was non-white (94% African American, 5% other/mixed race); the sample included only 5 white fathers.

Table C.3. Summary statistics of key baseline measures and baseline equivalence across study groups, for fathers completing the Parenting Skills Psychological Aggression subscale in the follow-up survey at 3 months post-intervention (n=304)

Baseline measure	Intervention (FF) mean (standard deviation) or %	Comparison (ES) mean (standard deviation) or %	Intervention versus comparison mean difference (p-value of difference)	Effect size
Demographics				
Father's age (years)	34.4 (9.5)	34.2 (8.6)	0.2 (0.877)	0.018
Number of children	2.4 (1.6)	2.5 (1.6)	-0.1 (0.804)	-0.028
Child's age	5.7 (4.6)	5.5 (4.5)	0.2 (0.689)	0.046
Marital status (% Never married)	67%	69%	-2% (0.773)	-0.061
Education level (% ≥High school diploma/GED)	73%	75%	-2% (0.744)	-0.073
Employment status (% Employed)	50%	50%	0% (1.000)	0.016
Child support order (% yes)	38%	48%	-10% (0.085)	-0.259 *
Lives with child (% yes)	42%	44%	-2% (0.921)	-0.030
Substance use				
Alcohol or drug problem (% moderate to high)	14%	13%	1% (1.000)	0.025
Secondary outcome				
Parenting Skills: CTSPC Psychological Aggression subscale (Frequency in past year)	10.56 (15.22)	8.36 (14.04)	2.20 (0.192)	0.150
Sample size	153	151	n.a.	n.a.

n.a. = not applicable. ES=Economic Stability. FF=Family Formation. CTSPC= Conflict Tactics Scale-Parent-Child.

* Does not meet baseline equivalence requirements because effect size>0.25

Notes: Independent samples t-tests were used to compare continuous variables across groups. Pearson chi-square tests (with the Yates continuity correction for 2X2 tables) were used to compare nominal variables across groups. p-values are included in parentheses. Effect sizes are Hedge's g for continuous variables and the Cox index for nominal variables.
Race was not assessed for baseline equivalence as 99% of the sample was non-white (95% African American, 4% other/mixed race); the sample included only 4 white fathers.

Table C.4. Summary statistics of key baseline measures and baseline equivalence across study groups, for fathers completing the Financial Responsibility item in the follow-up survey at 3 months post-intervention (n=156)

Baseline measure	Intervention (FF) mean (standard deviation) or %	Comparison (ES) mean (standard deviation) or %	Intervention versus comparison mean difference (p-value of difference)	Effect size
Demographics				
Father's age (years)	35.9 (9.2)	36.0 (8.7)	-0.1 (0.967)	-0.007
Number of children	2.2 (1.6)	2.2 (1.5)	0.0 (0.963)	-0.008
Child's age	7.8 (4.4)	8.0 (4.6)	-0.2 (0.878)	-0.025
Marital status (% Never married)	67%	73%	-6% (0.527)	-0.171
Education level (% ≥High school diploma/GED)	71%	74%	-3% (0.814)	-0.090
Employment status (% Employed)	51%	52%	-1% (1.000)	-0.024
Lives with child (% yes)	19%	17%	2% (0.809)	0.114
Substance use				
Alcohol or drug problem (% moderate to high)	17%	20%	-3% (0.715)	-0.144
Secondary outcome				
Financial Responsibility: Among those court ordered to pay child support, % paying	75%	64%	11% (0.204)	0.310 *
Sample size	72	84	n.a.	n.a.

n.a. = not applicable. ES=Economic Stability. FF=Family Formation.

* Does not meet baseline equivalence requirements because effect size>0.25

Notes: Independent samples t-tests were used to compare continuous variables across groups. Pearson chi-square tests (with the Yates continuity correction for 2X2 tables) were used to compare nominal variables across groups. p-values are included in parentheses. Effect sizes are Hedge's g for continuous variables and the Cox index for nominal variables.
Race was not assessed for baseline equivalence as 97% of the sample was non-white (93% African American, 4% other/mixed race); the sample included only 4 white fathers.

Table C.5 Summary statistics of key baseline measures and baseline equivalence across study groups, for fathers completing the Employment item in the follow-up survey at 3 months post-intervention (n=359)

Baseline measure	Intervention (FF) mean (standard deviation) or %	Comparison (ES) mean (standard deviation) or %	Intervention versus comparison mean difference (p-value of difference)	Effect size
Demographics				
Father's age (years)	35.0 (9.3)	34.3 (8.5)	0.7 (0.474)	0.075
Number of children	2.3 (1.5)	2.4 (1.6)	-0.1 (0.511)	-0.069
Child's age	6.2 (4.6)	5.9 (4.7)	0.3 (0.554)	0.062
Marital status (% Never married)	66%	69%	-3% (0.650)	-0.077
Education level (% ≥High school diploma/GED)	73%	75%	-2% (0.841)	-0.047
Child support order (% yes)	42%	54%	-12% (0.041)	-0.276 *
Lives with child (% yes)	35%	38%	-3% (0.719)	-0.062
Substance use				
Alcohol or drug problem (% moderate to high)	14%	14%	0% (1.000)	0.021
Secondary outcome				
Financial Stability: Employment status (% Employed)	48%	49%	-1% (0.886)	-0.032
Sample size	184	175	n.a.	n.a.

n.a. = not applicable. ES=Economic Stability. FF=Family Formation.

* Does not meet baseline equivalence requirements because effect size>0.25

Notes: Independent samples t-tests were used to compare continuous variables across groups. Pearson chi-square tests (with the Yates continuity correction for 2X2 tables) were used to compare nominal variables across groups. p-values are included in parentheses. Effect sizes are Hedge's g for continuous variables and the Cox index for nominal variables.
Race was not assessed for baseline equivalence as 98% of the sample was non-white (94% African American, 4% other/mixed race); the sample included only 6 white fathers.

Propensity Score Matching

Propensity score matching was used to create a matched set of fathers in the ES and FF groups for the secondary outcomes assessed at 3 months post-intervention which did not reach baseline equivalence requirements: *Parenting Skills*, *Financial Responsibility*, and *Employment*.

The method of nearest neighboring matching with a caliper was chosen as it has been widely used due to its reliability and ability to provide excellent covariate balance (Elze et al., 2017; Guo & Fraser, 2015). Per recommendations from Austin (2011) on the optimal caliper width, we used a caliper width of 0.2 with nearest neighboring matching. All baseline characteristics included in Table V.1 (father's age, number of children, child's age, marital status, education level, living with the child, child support order, substance abuse, employment status, and the baseline assessment of the outcome of interest) were used to create the propensity score of intervention assignment (FF vs. ES). The analyses were conducted with R using the MatchIt package (Stuart, King, Imai, & Ho, 2011).

By applying the propensity matching method, we created three sets of matched fathers in the FF and ES groups for the *Parenting Skills* analytic sample (n = 131 in each group), *Financial Responsibility* analytic sample (n = 65 in each group), and *Employment* analytic sample (n = 158 in each group). We conducted baseline equivalence tests with the matched samples using Cox Index and Hedge's g and found that all effect sizes met baseline equivalence requirements (effect size$\leq$ 0.25). Baseline equivalence of the propensity matched samples are shown in Tables C.6 through C.8. The propensity matched samples were used for the impact analyses for these three analytic samples.

Table C.6. Summary statistics of key baseline measures and baseline equivalence across propensity score matched study groups, for fathers completing the Parenting Skills Psychological Aggression subscale in the follow-up survey at 3 months post-intervention (n=262)

Baseline measure	Intervention (FF) mean (standard deviation) or %	Comparison (ES) mean (standard deviation) or %	Intervention versus comparison mean difference (p-value of difference)	Effect size
Demographics				
Father's age (years)	33.7 (9.0)	34.4 (8.8)	-0.7 (0.542)	-0.075
Number of children	2.5 (1.7)	2.4 (1.5)	0.1 (0.619)	0.061
Child's age	5.3 (4.4)	5.7 (4.6)	-0.4 (0.522)	-0.079
Marital status (% Never married)	69%	68%	1% (1.000)	0.021
Education level (% ≥High school diploma/GED)	73%	77%	-4% (0.477)	-0.148
Employment status (% employed)	50%	52%	-2% (0.902)	-0.037
Child support order (% yes)	43%	48%	-5% (0.457)	-0.131
Lives with child (% yes)	44%	40%	4% (0.707)	0.076
Substance use				
Alcohol or drug problem (% moderate to high)	13%	13%	0% (1.000)	0.000
Secondary outcome				
Parenting Skills: CTSPC Psychological Aggression subscale (Frequency in past year)	9.11 (13.60)	8.79 (14.51)	0.32 (0.850)	0.023
Sample size	131	131	n.a.	n.a.

n.a. = not applicable. ES=Economic Stability. FF=Family Formation. CTSPC= Conflict Tactics Scale-Parent-Child.

Notes: Independent samples t-tests were used to compare continuous variables across groups. Pearson chi-square tests (with the Yates continuity correction for 2X2 tables) were used to compare nominal variables across groups. p-values are included in parentheses. Effect sizes are Hedge's g for continuous variables and the Cox index for nominal variables.

Table C.7. Summary statistics of key baseline measures and baseline equivalence across propensity score matched study groups, for fathers completing the Financial Responsibility item in the follow-up survey at 3 months post-intervention (n=130)

Baseline measure	Intervention (FF) mean (standard deviation) or %	Comparison (ES) mean (standard deviation) or %	Intervention versus comparison mean difference (p-value of difference)	Effect size
Demographics				
Father's age (years)	35.9 (9.2)	35.2 (8.5)	0.7 (0.685)	0.071
Number of children	2.2 (1.7)	2.3 (1.6)	-0.1 (0.704)	-0.066
Child's age	8.1 (4.5)	7.6 (4.6)	0.5 (0.550)	0.105
Marital status (% Never married)	69%	74%	-5% (0.698)	-0.138
Education level (% ≥High school diploma/GED)	72%	71%	1% (1.000)	0.046
Employment status (% Employed)	55%	52%	3% (0.860)	0.075
Lives with child (% yes)	18%	14%	4% (0.634)	0.208
Substance use				
Alcohol or drug problem (% moderate to high)	17%	18%	-1% (1.000)	-0.064
Secondary outcome				
Financial Responsibility: Among those court ordered to pay child support, % paying	75%	74%	1% (1.000)	0.049
Sample size	65	65	n.a.	n.a.

n.a. = not applicable. ES=Economic Stability. FF=Family Formation.

Notes: Independent samples t-tests were used to compare continuous variables across groups. Pearson chi-square tests (with the Yates continuity correction for 2X2 tables) were used to compare nominal variables across groups. p-values are included in parentheses. Effect sizes are Hedge's g for continuous variables and the Cox index for nominal variables.

Table C.8. Summary statistics of key baseline measures and baseline equivalence across propensity matched study groups, for fathers completing the Employment item in the follow-up survey at 3 months post-intervention (n=316)

Baseline measure	Intervention (FF) mean (standard deviation) or %	Comparison (ES) mean (standard deviation) or %	Intervention versus comparison mean difference (p-value of difference)	Effect size
Demographics				
Father's age (years)	34.7 (9.1)	34.5 (8.7)	0.2 (0.885)	0.016
Number of children	2.3 (1.5)	2.4 (1.6)	-0.1 (0.614)	-0.057
Child's age	6.0 (4.6)	5.9 (4.7)	0.1 (0.830)	0.024
Marital status (% Never married)	67%	67%	0% (1.000)	0.000
Education level (% ≥High school diploma/GED)	73%	74%	-1% (0.899)	-0.039
Child support order (% yes)	47%	52%	-5% (0.500)	-0.107
Lives with child (% yes)	39%	36%	3% (0.727)	0.066
Substance use				
Alcohol or drug problem (% moderate to high)	14%	13%	1% (0.868)	0.067
Secondary outcome				
Financial Stability: Employment status (% Employed)	46%	48%	-2% (0.735)	-0.062
Sample size	158	158	n.a.	n.a.

n.a. = not applicable. ES=Economic Stability. FF=Family Formation.

Notes: Independent samples t-tests were used to compare continuous variables across groups. Pearson chi-square tests (with the Yates continuity correction for 2X2 tables) were used to compare nominal variables across groups. p-values are included in parentheses. Effect sizes are Hedge's g for continuous variables and the Cox index for nominal variables.

Appendix D: Data Preparation

Data monitoring was conducted throughout the time of data collection to ensure quality. Surveys were reviewed routinely for data quality to identify item- and scale-level inconsistencies. Data-cleaning included identifying and addressing: missing data, duplicate data, incorrect values, and/or outliers. When combining data from multiple datasets, we merged data based on a unique participant ID number. For multi-item scales, all items must have been completed to calculate a summary score. All analytic samples included fathers who were eligible to answer the specific items or scales for each outcome at baseline, had complete outcome data at both baseline and the follow-up survey, and had complete data on key baseline characteristics (listed in Table V.1).

Appendix E: Impact Estimation

Baseline equivalence effect size

Effect sizes for baseline equivalence were calculated using Hedges' g and the Cox Index (What Works Clearinghouse, 2017). The effect size for continuous measures across groups was calculated using Hedges' g, shown below.

$$g = \frac{y_i - y_c}{S}$$

$$S = \sqrt{\frac{(n_i - 1)s_i^2 + (n_c - 1)s_c^2}{n_i + n_c - 2}}$$

In the above equations, y_i and y_c represent the means of the continuous outcomes for the intervention (FF) and comparison (ES) groups, respectively. The sample sizes for the intervention and comparison groups are indicated by n_i and n_c, respectively. The standard deviations for the intervention and comparison groups are indicated by s_i and s_c, and S indicates the pooled within-group standard deviation of the outcome.

The effect size for baseline equivalence of dichotomous variables across groups was calculated using the Cox Index, which is shown below.

$$d_{Cox} = \frac{LOR}{1.65}$$

In the above equation, LOR is the difference between the log odds for the intervention (FF) and comparison (ES) groups ($\ln(Odds_i) - \ln(Odds_c)$).

Impact analysis

The differences in primary and secondary outcomes between the intervention (FF) and comparison group (ES) at 12 months post-intervention (primary outcomes) and 3 months post-intervention (secondary outcomes) were assessed using linear and generalized linear regression models. Linear regression models were used for the continuous outcomes (i.e. *Father Involvement*, *Healthy Co-Parenting Relationship Skills*, *Father Well-Being*, or *Child Well-Being*). The basic equation for the linear regression models is shown below.

$$Y_i = \beta_0 + \beta_1 T_{1i} + \beta_2 x_{2i} + \ldots + \beta_k x_{ki} + \varepsilon_i$$

Y_i represents the value of the continuous dependent variable (e.g., *Father Involvement*), β_0 is the intercept, β_1 is the regression coefficient for the treatment variable, T_{1i} is the value of the

treatment variable (1 for FF, 0 for ES), β_2 through β_k are regression coefficients for the covariates, and ε_i represents the random error.

Logistic regression models were used for the secondary outcomes *Financial Responsibility* and *Financial Stability*. The structure of the model equations were similar to that described above, except the dependent variable was in the form of the log odds of the outcome $\left(\ln\left(\frac{P}{1-P}\right)\right.$ and there is no error term. Negative binomial regression was used for the outcome *Parenting Skills* (count of psychological aggression occurrences) because this outcome reflected counts of occurrences since the last interview and was over-dispersed (overdispersion test: dispersion estimate 17.7, $z=5.3$, $p<0.001$). The structure of the model equation was also similar to the linear model described above, but the dependent variable was in the form of the log of the count outcome.

Appendix F: Sensitivity Analyses and Alternative Model Specifications

Table F.1. Post-intervention estimated effects using data from survey 12 months after the program to address the primary research questions, not adjusting for baseline covariates

	Intervention (FF)		Comparison (ES)		
Outcome measure	n	Intervention mean (standard deviation)	n	Comparison mean (standard deviation)	Intervention compared to comparison mean difference – Unadjusted [a] (p-value of difference)
Father Involvement: FRPN Father Engagement caregiving/play subscale (standardized – Z scores)	168	0.10 (0.91)	161	0.24 (0.98)	β = -0.15 (*p* = 0.156)
Healthy (Co-parenting) Relationship Skills: FRPN co-parenting Alliance subscale (Range 1 to 5)	158	3.66 (1.16)	142	3.61 (1.17)	β = 0.05 (*p* =0.735)
Parenting Skills: CTSPC Psychological Aggression subscale (Frequency since last interview)	140	6.51 (12.27)	140	3.58 (6.67)	β = 0.60 (*p* =0.016)
Father Well-being: SF-12v2 Health Survey, Mental Health Composite Scale score (MCS) (Range 0 to 100)	168	52.80 (10.79)	158	52.73 (10.29)	β = 0.08 (*p* = 0.949)
Child Well-being: CBCL Aggressive Behavior subscale (Standardized – Z scores)	98	-0.28 (1.090)	92	-0.35 (0.978)	β = 0.07 (*p* = 0.651)

[a] Difference in means between groups are reported for continuous outcomes (*Father Involvement*, *Co-parenting Relationship Skills*, *Father Well-being*, and *Child Well-being*). Difference in log count is shown for *Parenting Skills* frequency of psychological aggression. p-values are included in parentheses. See Table IV.2 for a more detailed description of each measure and Chapters IV and V.B for a description of the impact estimation methods.

Appendix G: Additional Analyses

Table G.1. Lesson topics attended by program assignment

Topic[a]	FF (n=349) [b] % attended ≥ one lesson	ES (n=341) [c] % attended ≥ one lesson	p [d]
Economic stability topics offered to both groups			
Job readiness	60%	60%	0.949
Money management	61%	56%	0.207
Other topics offered to both groups (but taught in context of economic stability for ES group)			
Communication	59%	38%	<0.001
Goals/Future planning	59%	51%	0.029
Stress/Anger management	55%	46%	0.015
Health/Substance abuse	54%	36%	<0.001
Topics offered to FF only			
Parenting/Fathering	64%	<1%	<0.001
Relationships	61%	<1%	<0.001
Making decisions	55%	<1%	<0.001
Manhood	52%	<1%	<0.001
Child support	44%	<1%	<0.001

[a] Job readiness included 9 lessons offered to FF and 17 lessons offered to ES. Money management included 7 lessons offered to FF and 7 lessons offered to ES. Communication included 5 lessons offered to FF and 2 lessons offered to ES. Stress/Anger management included 2 lessons offered to FF and 4 lessons offered to ES.Goals/Future planning included 4 lessons offered to FF and 2 lessons offered to ES. Health/Substance abuse includes 2 lessons offered to FF and 2 lessons offered to ES.
Parenting/Fathering (11 lessons), Relationships (9 lessons), Making decisions (4 lessons), Manhood (2 lessons), and Child Support (1 lesson) were only offered to FF. However, up to two ES individuals attended some FF courses, accounting for the <1% ES attendance for these topics.
[b] Excludes one ineligible participant who was inadvertently randomized and enrolled into FF program.
[c] Excludes one ES participant with missing data on specific lessons attended.
[d] P value shown for Pearson chi-square tests.

Table G.2. Post-intervention estimated effects using data from survey 3 months after the program to address the secondary research questions, not adjusting for baseline covariates

	Intervention (FF)		Comparison (ES)		
Outcome measure	n	Intervention mean (standard deviation) or %	n	Comparison mean (standard deviation) or %	Intervention compared to comparison mean difference – Unadjusted [a] (p-value of difference)
Father Involvement: FRPN Father Engagement caregiving/play subscale (standardized – Z scores)	189	0.05 (0.99)	173	0.05 (0.97)	β <0.01 (*p* = 0.993)
Healthy (Co-parenting) Relationship Skills: FRPN co-parenting Alliance subscale (Range 1 to 5)	177	3.54 (1.17)	159	3.63 (1.15)	β = -0.09 (*p* = 0.472)
Parenting Skills: CTSPC Psychological Aggression subscale (Frequency since last interview)[b]	131	4.14 (8.92)	131	4.40 (10.42)	β = -0.06 (*p* = 0.823)
Financial Responsibility: Among those court ordered to pay child support, % paying [b]	65	56 (86%)	65	54 (83%)	β = 0.24 (*p* = 0.627)
Financial Stability: Employment status (% Employed) [b]	158	117 (74%)	158	117 (74%)	β = 0.00 (*p* = 1.000)

Source: First follow-up surveys administered 3 months after the program.

[a] Difference in means between groups are reported for continuous outcomes (Father Involvement, Co-parenting Relationship Skills). Difference in log count is shown for Parenting Skills frequency of psychological aggression. Difference in log odds is shown for Financial Responsibility and Financial Stability. p-values are included in parentheses. See Appendix B Table B.4 for a more detailed description of each measure and Chapters IV and V.B for a description of the impact estimation methods.

[b] Propensity score matched sample was used. For more details, see Appendix C.

Estimation of complier average causal effect (CACE)

Two-stage least squares (2SLS) regression, with randomization to the intervention (FF) or comparison group (ES) as the instrumental variable, was used to estimate the complier average causal effect (CACE) of FF intervention (Angrist, Imbens & Rubin, 1996). With 2SLS regression two models are jointly estimated, the first of which predicts participation in the intervention, and the second of which predicts the outcome given participation in the intervention. Randomization to intervention group was used as the instrumental variable. In the 2SLS regression, the first stage predicts attendance in the intervention (0= <75% of FF sessions,

1= ≥75% of FF sessions) as a function of randomized assignment to FF or ES. The second stage predicts the outcome given participation in FF , and the two models are estimated jointly.

Stage 1 regresses the intervention received, FF (i.e., 1 if attended ≥75% of FF sessions, 0 if attended <75% of sessions), on the instrumental variable (randomization to the FF or the ES group) denoted by Z (1 if FF, 0 if ES).

$$FF_i = \alpha_0 + \alpha_1 Z_i + \varepsilon_i$$

Predicted values of $\widehat{FF}$ given Z are obtained. Then in Stage 2, the outcome Y is regressed on predicted values of $\widehat{FF}$ obtained from Stage 1.

$$Y_i = \beta_0 + \beta_1 \widehat{FF}_i + \varepsilon_i$$

β_1 estimates the causal effect of the intervention. The above represents models used for our continuous outcomes (*Father Involvement*, *Co-Parenting Relationship Skills*, *Father Well-Being*, and *Child Well-Being*) which were performed using the ivpack package in R (Jiang & Small, 2014). The distribution of the outcome in second stage models was changed for our non-continuous outcomes using the ivtools package in R (Sjolander, Dahlgwist, & Martinussen, 2020). The binomial distribution was used for our dichotomous outcomes (*Financial Responsibility* and *Financial Stability*), and a "quasipoisson" distribution was used for the count outcome for *Parenting Skills* (psychological aggression occurrences).

Table G.3. Differences in means between intervention and comparison groups for primary outcomes estimated using treatment-on-treated approach

Outcome	ITT (from Table F.1)	CACE: 75% of FF sessions attended	CACE: 50% of FF sessions attended
Primary outcomes			
Father Involvement: FRPN Father Engagement caregiving/play subscale (standardized – Z scores)	β = -0.15 (*p* = 0.156)	β = -0.30 (p=0.156)	β = -0.26 (p=0.156)
Healthy (Co-parenting) Relationship Skills: FRPN co-parenting Alliance subscale (Range 1 to 5)	β = 0.05 (*p* = 0.735)	β = 0.09 (p=0.736)	β = 0.08 (p=0.735)
Parenting Skills: CTSPC Psychological Aggression subscale (Frequency since last interview)	β = 0.60 (p=0.016)*	β = 1.28 (p=0.009)**	β = 1.12 (p=0.009)**
Father Well-being: SF-12v2 Health Survey, Mental Health Composite Scale score (MCS) (Range 0 to 100)	β = 0.08 (*p* = 0.949)	β = 0.15 (p=0.949)	β = 0.13 (p=0.949)
Child Well-being: CBCL Aggressive Behavior subscale (Standardized – Z scores)	β = 0.07 (*p* = 0.651)	β = 0.15 (p=0.649)	β = 0.13 (p=0.649)

Source: Second follow-up surveys administered 12 months after the program.

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

Notes: Intention-to-treat (ITT) and complier average causal effect (CACE) estimates are presented, unadjusted for covariates. For CACE estimates, two-stage least squares (2SLS) regression analysis was used.

Table G.4. Differences in means between intervention and comparison groups for secondary outcomes using treatment-on-treated approach

Outcome	ITT (from Table G.2)	CACE: 75% of FF sessions attended	CACE: 50% of FF sessions attended
Secondary outcomes			
Father Involvement: FRPN Father Engagement caregiving/play subscale (standardized – Z scores)	β < 0.01 (*p* = 0.993)	β < 0.01 (*p* = 0.993)	β < 0.01 (*p* = 0.993)
Healthy (Co-parenting) Relationship Skills: FRPN co-parenting Alliance subscale (Range 1 to 5)	β = -0.09 (*p* = 0.472)	β = -0.18 (*p* = 0.473)	β = -0.16 (*p* = 0.473)
Parenting Skills: CTSPC Psychological Aggression subscale (Frequency since last interview)	β = -0.06 (*p* = 0.823)	β = -0.16 (*p* = 0.772)	β = -0.15 (*p* = 0.749)
Financial Responsibility: Among those court ordered to pay child support, % paying	β = 0.24 (*p* = 0.627)	β = 0.43 (*p* = 0.613)	β = 0.37 (*p* = 0.628)
Financial Stability: Employment status (% Employed)	β = 0.00 (*p* = 1.000)	β = 0.02 (*p* = 0.977)	β = 0.04 (*p* = 0.937)

Source: First follow-up surveys administered 3 months after the program.

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

Notes: Intention-to-treat (ITT) and complier average causal effect (CACE) estimates are presented, unadjusted for covariates. For CACE estimates, two-stage least squares (2SLS) regression analysis was used.

Change in outcomes over time and by group

To examine whether primary and secondary outcomes changed over time, and whether these changes over time differed between FF and ES groups, we used linear and generalized linear mixed effects models which allow for correlation among repeated measures within individuals. Below is the general form of our random-intercept linear mixed effects models, which were used for our continuous outcomes (i.e. *Father Involvement*, *Co-Parenting Relationship Skills*, *Father Well-Being*, or *Child Well-Being*).

$$Y_{ij} = \gamma_{00} + \gamma_{10}\text{Time}_{ij} + \gamma_{01}\text{Group}_j + \gamma_{11}\text{Group}_j\text{*Time}_{ij} + \gamma_{02}X_j \ldots + \gamma_{0k}X_j + u_{0j} + r_{ij}$$

Where Y_{ij} is the outcome for j^{th} father at time i, γ_{00} is the fixed intercept, γ_{10} is the coefficient for the fixed effect of time (follow-up survey vs baseline survey), γ_{01} is the coefficient for the fixed effect of intervention group (FF vs ES), γ_{11} is the coefficient for the interaction of intervention group and time, γ_{02} through γ_{0k} are the coefficients for father characteristics at baseline. u_{0j} is the error, or unmodeled variability, between fathers, and r_{ij} is the residual error within fathers. If the coefficient for the interaction between intervention group and time is significant, the change over time in the outcome differs by intervention group.

Generalized linear mixed models were used for non-continuous outcomes and had similar structure to the above equation, but with a different form of the dependent variable. The logistic link function was used for the dichotomous dependent variables (*Financial Responsibility* and *Financial Stability*) and the log link function for the count outcome (*Parenting Skills* count of psychological aggression occurrences). IBM SPSS and R were used for this analysis. A p value <0.05 was considered statistically significant.

Table G.5. Mixed model results examining change in scores over time and between groups using data from baseline and survey 3 months after the program to address the secondary research questions

Outcome measure	Intervention (FF)		Comparison (ES)		*Unadjusted Model Results*[a]				*Adjusted Model Results*[b]			
	T1 mean (sd) or %	T2 mean (sd) or %	T1 mean (sd) or %	T2 mean (sd) or %	*n*	*Group p value*	*Time p value*	*Group*Time p value*	*n*	*Group p value*	*Time p value*	*Group*Time p value*
Father involvement – FRPN Father Engagement caregiving/play subscale (standardized – Z scores)	-0.07 (1.00)	0.05 (1.00)	-0.03 (0.97)	0.05 (0.97)	362	0.833	0.006	0.559	362	0.947	0.006	0.559
Healthy (Co-parenting) Relationship Skills – FRPN co-parenting Alliance subscale (Range 1 to 5)	3.38 (1.20)	3.54 (1.17)	3.40 (1.23)	3.63 (1.15)	336	0.634	<0.001	0.490	336	0.379	<0.001	0.490
Parenting Skills – CTSPC Psychological Aggression subscale (Frequency since last interview) [c]	9.11 (13.60)	4.14 (8.92)	8.79 (14.51)	4.40 (10.42)	262	0.499	<0.001	0.795	262	0.452	<0.001	0.843
Financial Responsibility: Among those court ordered to pay child support, % paying [c]	75%	86%	74%	83%	130	0.885	0.128	0.823	130	0.985	0.118	0.848
Financial Stability: Employment status (% Employed) [c]	46%	74%	48%	74%	316	0.646	<0.001	0.764	316	0.747	<0.001	0.758

Source: Baseline survey and first follow-up survey administered 3 months after the program.

Note: See Appendix B Table B.4 for a more detailed description of each measure and Chapters IV and V.D for a description of the impact estimation methods.

[a] Unadjusted models only include group, time, and group by time interaction.

[b] Adjusted models additionally include all covariates listed in Table V.1.

[c] Propensity score matched sample used. For more details see Appendix C.

Figure G1. Unadjusted primary outcomes by time and intervention group

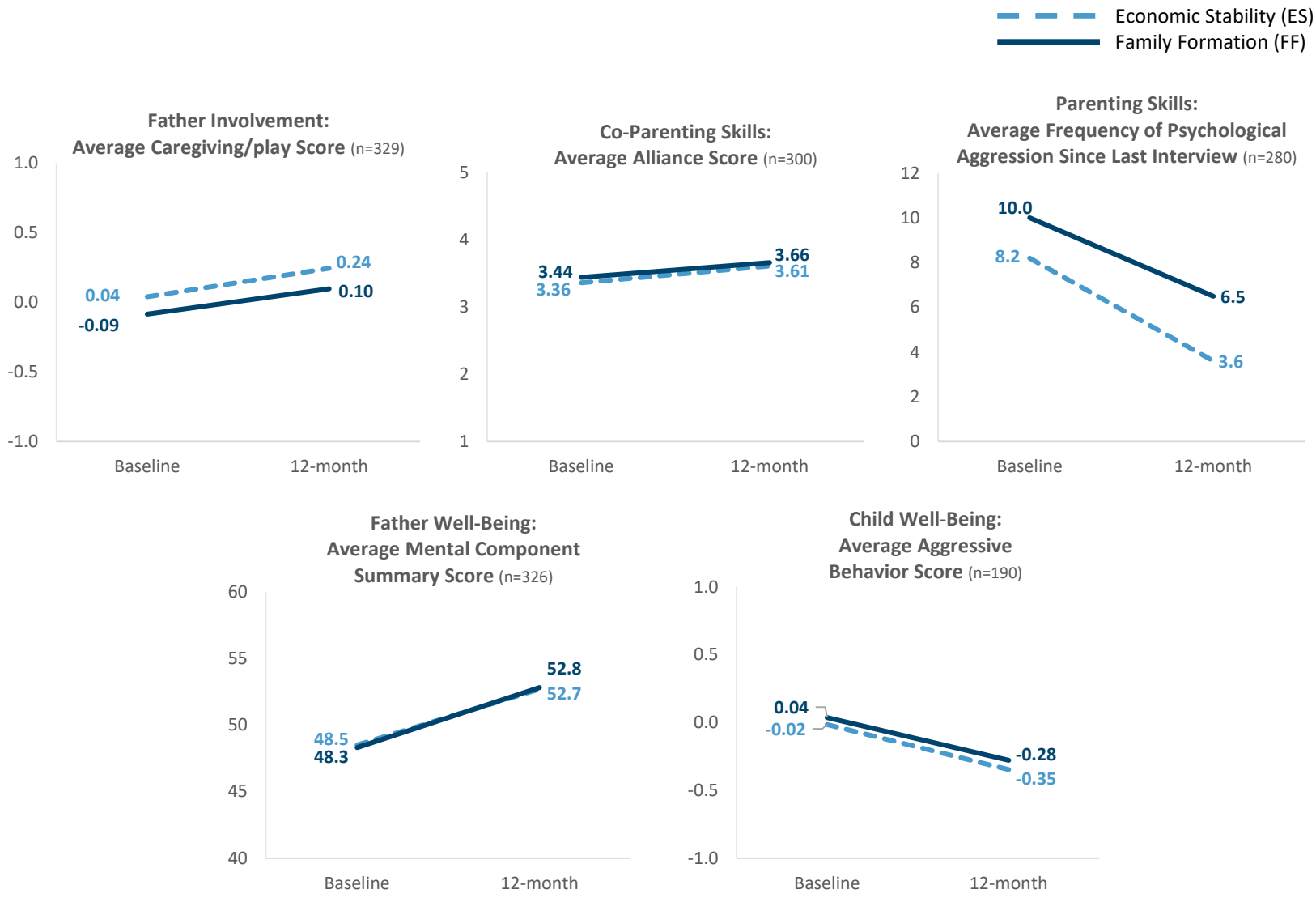

Figure G2. Unadjusted secondary outcomes by time and intervention group