Impact Evaluation of The Pathways for Fathers and Families Program in Milwaukee, WI

Final Impact Evaluation Report for Milwaukee County Child Support Services

07/15/2020

Prepared by

Razia Azen, University of Wisconsin – Milwaukee,

Samuel Maurice, University of Wisconsin – Milwaukee

Sara Woods, AMTC & Associates

Ryan Adomavich, AMTC & Associates

Recommended Citation:

Azen, R., Maurice, S. A., Woods, S., & Adomavich, R. (2020). Pathways for Fathers and Families Impact Evaluation. OPRE Report Number ###. Washington, DC: Office of Family Assistance, Administration for Children and Families, U.S. Department of Health and Human Services.

Acknowledgements:

The authors would like to acknowledge a number of persons critical to making this evaluation and program implementation possible, including Luciana Cançado, Cindy Walker, and Sonja Sprandel (University of Wisconsin - Milwaukee); Angela Turner (AMTC & Associates); Jeanette Stevens, Sandra Stevens, and JeTaunne Richardson (Milwaukee County Child Support Services). We also wish to thank the staff at My Father's House for their dedication to the implementation of this study.

Disclaimer:

The authors have no conflict of interest, financial or otherwise.

This publication was prepared under Grant Number 90FK0093 from the Office of Family Assistance (OFA) within the Administration for Children and Families (ACF), U. S. Department of Health & Human Services (HHS). The views expressed in this report are those of the authors and do not necessarily represent the views or policies of HHS, ACF, or OFA.

Structured Abstract: "The Evaluation of the Pathways for Fathers and Families Program in Milwaukee, WI"

Objective. The main goal of this study was to evaluate the impact of the Pathways for Fathers and Families (PFF) program in changing attitudes, beliefs, and behaviors regarding positive parenting and healthy relationships among non-custodial fathers in Milwaukee, WI.

Study design. The study utilized a randomized control trial design. Participants (n = 482) enrolled in a fatherhood and healthy relationship program at one community agency in Milwaukee, WI. Prior to the onset of the PFF program, fathers were randomly assigned to either a treatment group, which received the full fatherhood program, or a control group that received a subset of the services provided in the full program. Data collection occurred prior to the onset of training, immediately after the completion of training, and six months following the completion of training.

Results. No significant changes in positive parenting skills or attitudes were found at the completion of the PFF program. Fathers in the treatment group (receiving the full program) did not demonstrate significantly better outcomes than those in the control group. Secondary analyses found that six months after the completion of the training program participants in the treatment group did show significantly more positive views on healthy co-parenting, better communication with their co-parent, increased conflict resolution skills, more positive views on healthy parenting, and more engagement with their child, relative to participants in the control group.

Conclusion. The full PFF program did not appear to significantly improve fathers' parenting, conflict resolution, and communication skills immediately after completion of the program. While statistically significant improvements were detected at six months after the completion of the program, these analyses were exploratory and their results should be interpreted with caution.

Contents

Tables

Impact Evaluation of the Pathways for Fathers and Families Program in Milwaukee, WI

I. Introduction

A. Introduction and study overview

Over the past three decades, researchers have devoted substantial time and energy to investigating the causes and effects of non-custodial fatherhood (Garfinkle & Oellerich, 1989; Threfall & Kohl 2015; Finzi-Dottan & Cohen, 2019). For some time, most studies focused on examining issues related to child support non-payment (Del Boca & Flinn, 1996; Sorensen, 1997). Recently, however, there has been a shift away from the sole emphasis on child support issues. Researchers have begun asking more complex and intriguing questions about the various factors that affect non-custodial fathers. Studies have examined the degree to which fathers find support programs socially valid (Meyer, Kim, & Cancian 2019), the efficacy of programs for improving fathers' parenting skills (Self-Brown et al., 2017), and even ways to help support non-custodial fathers with multiple risk factors such as homelessness (Ferguson & Morley, 2011).

One of the many benefits of this new research is the emerging consensus that non-custodial fatherhood cannot be simply conceptualized by considering only issues related to child support payment (Gordon et al., 2012). Instead, the non-custodial father should be seen as a holistic person, someone who may have experienced trauma or mental health challenges, who may be struggling with poverty, who may never have had the opportunity to learn healthy parenting skills, and who may need support in many facets of life. Indeed, research appears to be showing that the most effective way to support children, co-parents, and communities is to support the non-custodial fathers themselves (Solomon-Fears, Falk, & Fernandes, 2013).

This hypothesis is not only being supported empirically but is also theoretically sound. By addressing root causes of psychological, interpersonal, and economic risk factors, experts can facilitate real and permanent change for both fathers and the various ecological systems around them (Hamer, 1998; Campbell, Howard, Rayford, & Gordon, 2015). Indeed, if one takes only the narrow example of increasing non-custodial fathers' physical contact with their children, researchers have found a variety of factors at work including fathers' satisfaction with parenting and their belief that they can positively influence their children's lives (McKenry, Price, Fine, & Serovich, 1992). These types of factors are clearly psychological in nature and are not likely to be influenced through simple economic reinforcement or consequences alone.

This corpus of research led directly to the present study. In September of 2006, the Office of Family Assistance (OFA) awarded demonstration grants to 226 organizations to promote the overall well-being of children and families, including the implementation of Responsible

Fatherhood programs. The Responsible Fatherhood program was designed to help fathers overcome barriers that impede them from becoming effective and nurturing parents, while also helping them improve their relationships with their children.

In 2010, Milwaukee County Child Support Services (MCCSS) had the opportunity to convene a city-wide coalition to provide parenting, healthy relationship and marriage education, as well as economic stability services, in a variety of community-based organizations and correctional facilities. Fathers could pick and choose one or more services, and it was not required that a father receive an integrated package of all three program components. With this being a novel model in Milwaukee, rigorous empirical evaluation was not a part of the 5-year project. In 2015, OFA presented an opportunity for MCCSS to further its work by implementing a multi-component program for fathers, and it was the opportune time to add a rigorous evaluation to the new program design. Subsequently, MCCSS selected My Father's House (MFH), a non-profit agency with which they had an established relationship, to participate in the impact evaluation. MFH had been providing fatherhood services in Milwaukee since 1999 and had worked closely with MCCSS to overcome implementation challenges in the 2010-2015 OFA-funded fatherhood program.

The present impact evaluation was specifically designed to determine if adding Fatherhood program components to education on financial responsibility would yield positive outcomes for non-custodial fathers in the areas of maintaining a heathy relationship with the mother of their child(ren) and being more involved in the lives of their child(ren). This evaluation contributes to the base of research as there has been no previous empirical study on the Pathways for Fathers and Families (PFF) intervention using a randomized control trial (RCT) design. The Parents and Children Together (PACT) evaluation, which was conducted by Mathematica from 2011-2019, is a large-scale, multi-component project intended to broaden the understanding of Responsible Fatherhood programs (Avellar, Covington, Moore, Patnaik, & Wu, 2019). One program studied in the PACT evaluation was provided by the Father's Support Center in St. Louis, MO. That program is called Family Formation and it provided a relevant framework for the current study in terms of both the curricula offered and demographics of the study target population. The PACT evaluation was a cohort-based model that offered both economic stability services and healthy parenting/fatherhood curricula. This research provided the impetus for MCCSS to examine the chosen intervention, an integrated model of parenting, healthy relationship, and economic stability services, in Milwaukee, WI.

The present impact evaluation utilized an RCT design to provide a rigorous and credible examination of the efficacy of the curricula and services provided. The RCT design is best suited to answer the research questions because it allows the evaluators to determine the impact of programming by comparing, on relevant outcomes, non-custodial fathers who were offered the additional curricula in the areas of parenting and healthy relationships with a comparable group of fathers who were not offered these components. After consenting to be part of the study and providing baseline data, individual fathers were randomly assigned to either a treatment or control group. Fathers in the treatment group received the healthy relationship curriculum and the responsible parenting curriculum, as well as child support services, education to promote

economic stability, and case management services. Fathers in the control group received only the child support services, education to promote economic stability, and case management services provided to all participants regardless of group. For both groups, data was collected at baseline (before starting the program), immediately on completion of the program (one or two months after baseline data collection, depending on group assignment), and six months after completion of the program.

The remainder of this section lists the study's research questions. In Section II, we provide descriptions of the intervention and counterfactual conditions, as well as the implementation evaluation research questions. Section III describes the study design, including the sample of participants and the data collection processes. Section IV details the analytic methods, including the formation of analytic samples, the outcome measures, and the examination of baseline equivalence. Section V presents the results pertaining to the implementation evaluation, the primary and additional research questions of the impact evaluation, and sensitivity analyses. Section VI concludes with a discussion of the findings.

B. Primary research question(s)

Six primary research questions were developed to test hypotheses concerning the efficacy of the PFF program. It was hypothesized that the effect of the program would be most apparent immediately at the conclusion of the program. Thus, the primary research questions focus on evaluating the intended outcomes of the intervention program **immediately upon completion** of the program. Skills, attitudes, and beliefs were measured using six researcher-created scales, and the primary research questions were determined based on the various outcomes of interest as measured by these scales; for example, to evaluate changes in conflict resolution skills, the scale of conflict resolution skills was used as the measured outcome.

1. Primary research questions

1. What is the impact of the Milwaukee PFF program on fathers' views on healthy co-parenting at the completion of the intervention?
2. What is the impact of the Milwaukee PFF program on fathers' attitudes towards communicating with their co-parent at the completion of the intervention?
3. What is the impact of the Milwaukee PFF program on fathers' ability to resolve conflict at the completion of the intervention?
4. What is the impact of the Milwaukee PFF program on fathers' beliefs about healthy parenting at the completion of the intervention?
5. What is the impact of the Milwaukee PFF program on fathers' engagement with their child at the completion of the intervention?
6. What is the impact of the Milwaukee PFF program on fathers' conflicts with their co-parent at the completion of the intervention?

C. Secondary research question(s)

Secondary or additional research questions were developed to evaluate the outcomes of interest at the **six-month Follow-Up** time point. The first six additional research questions were designed to measure a specific skill, attitude, or belief, **six months after program completion**, and were evaluated using the same researcher-created scales that were used for the primary research questions. The last additional research question was intended to examine whether fathers in the treatment group demonstrated greater economic stability relative to those in the control group at the **completion** of the training. Economic stability was only measured at the program exit and not at the six-month follow-up time point. This outcome was considered additional (i.e., secondary) because both groups received training on economic stability.

2. Secondary/additional research questions

1. What is the impact of the Milwaukee PFF program on fathers' views on healthy co-parenting six months after the completion of the intervention?
2. What is the impact of the Milwaukee PFF program on fathers' attitudes towards communicating with their co-parent six months after the completion of the intervention?
3. What is the impact of the Milwaukee PFF program on fathers' ability to resolve conflict six months after the completion of the intervention?
4. What is the impact of the Milwaukee PFF program on fathers' beliefs about healthy parenting six months after the completion of the intervention?
5. What is the impact of the Milwaukee PFF program on fathers' engagement with their child six months after the completion of the intervention?
6. What is the impact of the Milwaukee PFF program on fathers' conflicts with their co-parent six months after the completion of the intervention?
7. What is the impact of the Milwaukee PFF program on fathers' economic skills at the completion of the intervention?

II. Intervention and counterfactual conditions

This section provides a description of the PFF Program. All participants received economic stability workshops, case management, and child support services. Participants in the intervention condition (i.e., treatment group) additionally received training in healthy relationships, positive parenting, and conflict resolution skills. The program is intended for non-custodial fathers who were at least 18 years of age, with at least one child between the ages of 0 and 21, and living in Milwaukee County, at the time of enrollment.

A. Description of program as intended

1. Purpose

The purpose of the study was to examine the primary objectives of the PFF program, which are to (1) strengthen positive engagement between fathers and their children, (2) develop the skills to form and sustain healthy relationships, and (3) improve economic stability. The logic model guiding the evaluation can be found in Appendix A.

2. Program components

This was a multi-component intervention, in which fathers in the treatment group participated in workshops and services to improve parenting and relationship skills and increase economic stability. The main components of the program included training in healthy relationships, positive parenting, and conflict resolution skills, as well as economic stability programming, case management, and child support services. Program components are summarized in Table II.1.

3. Program content

The intervention consisted of the Within My Reach healthy marriage relationship curriculum (12 lessons), the 24/7 Dad A.M. responsible parenting curriculum (lessons 3, 5, 7, 8, 10, 11), and the Fatherhood Development responsible parenting curriculum (lessons 2, 3, 6, 7, 8, 11, 12, 13, 15). A sample curriculum implementation plan is included in Appendix B (Table B.1). All curricula education was provided by My Father's House (MFH) staff.

Fathers in the treatment group also participated in economic stability workshops on budgeting, credit cards, loans, savings and investing (provided by MFH), an orientation to career and workforce services (provided by Employ Milwaukee [EM)]), the local workforce development board), an orientation to child support services and screening for service needs (provided by MCCSS), driver's license education and screening for services (provided by Wisconsin Community Services Center for Driver's License Recovery), and an orientation and screening for legal services (provided by Centro Legal, a local non-profit legal services provider).

4. Intended implementation

Location or setting: MFH implemented the community-based impact evaluation at their organization location.

Duration and dosage: Each cohort series was open to up to 30 fathers, and the entire study was designed to have 22 cohorts/workshops. The treatment group received the healthy relationship curriculum followed by the responsible parenting curriculum, as well as child support services, workforce preparation and economic stability services, and other support services. The program was delivered through 2-hour classes that occurred 3 times per week over a 7-week period, for a total of 42 hours of training/19 sessions.

Staffing and staff training: MFH employed two dedicated Father Advocates to deliver curriculum education and provide case management services. To accommodate fathers, programming was delivered twice a day. One Advocate implemented the program services during the morning cohort series and the other Advocate implemented the program services during the evening cohort series. All program implementation staff completed a comprehensive training plan prior to enrolling fathers and providing services. Curricula trainings for Within My Reach, Fatherhood Development and 24/7 Dad A.M. were provided by the curriculum developers. Motivational Interviewing (MI) training was provided by experts from Alma Center, a local organization specializing in trauma and domestic violence services for fathers. Child Support training was provided by MCCSS. An orientation of the workforce services available in Milwaukee was provided by EM, the local workforce development board, and effective messaging training was provided by a local organizational development consultant. The staff education and training is summarized in Table II.2.

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

Component	Curriculum and content	Dosage and schedule	Delivery	Target Population
Intervention				
Relationship skills workshops	Within My Reach: Healthy relationship curriculum: understanding partner's perspectives; avoiding destructive conflict; and communicating effectively	15 hours, with 2-hour sessions occurring three times a week	Group lessons provided at the intervention's facilities by one trained facilitator in every session (option of a day or night session)	Non-Custodial Parent (NCP) within Milwaukee (MKE) County (Co)
Parenting workshops	Fatherhood Development & 24/7 Dad A.M.: Comprehensive fatherhood program designed to help men improve their parenting skills and fathering knowledge	The following 6 select lessons from the 24/7 Dad A.M. curriculum: Lessons 3, 5, 7, 8, 10, and 11; and the following 9 select lessons from the Fatherhood Development curriculum: Lessons 2, 3, 6, 7, 8, 11, 12, 13, and 15. Delivered in 2-hour sessions occurring three times a week	Group lessons provided at the intervention's facilities by one trained facilitator in every session (option of a day or night session)	NCP within MKE Co.
Economic stability workshops	Budgeting, credit cards, car loans, saving and investing	Weekly 2-hour workshops	Workshops are provided by one MFH facilitator	NCP within MKE Co.
Workforce services	Orientation to career and workforce preparation services	1-hour session	Session provided by EM Milwaukee's workforce development board	NCP within MKE Co.
Access to Child Support Services	Overview of the State child support system, including orientation to services provided by the Child Support Office	1-hour session	Session provided by Child Support Services (MCCSS)	NCP within MKE Co.
Case management	Create individual service plans, provide coaching, resources, and referrals to comprehensive support services network	Ongoing	MFH Case Manager	NCP within MKE Co.

Component	Curriculum and content	Dosage and schedule	Delivery	Target Population
Counterfactual				
Economic stability workshops	Budgeting, credit cards, car loans, saving and investing	Weekly 2-hour workshops	Workshops are provided by one MFH facilitator	NCP within MKE Co.
Workforce services	Orientation to career and workforce preparation services	1-hour session	Session provided by Employ Milwaukee	NCP within MKE Co.
Access to Child Support Services	Overview of the State child support system, including orientation to Child Support Services	1-hour session	Session provided by Child Support Services (MCCSS)	NCP within MKE Co.
Case management	Create individual service plans, provide coaching, resources, and referrals to comprehensive support services network	Ongoing	MFH Case Manager	NCP within MKE Co.

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

Component	Education and initial training of staff	Ongoing training of staff
Intervention		
Relationship skills and parenting workshops	MFH utilized male facilitators trained by the curriculum publisher. Facilitators hold a Bachelor's Degree upon hire, have past experience in facilitating fatherhood or relationship education, and received comprehensive initial training and certification by the curriculum developer.	Facilitators receive a half-day of annual refresher training in the policies and procedures of the program and refresher trainings for the two fatherhood curricula.
Economic stability workshops	MFH utilized trained male and female facilitators, as did EM, the workforce development board. Staff hold a Bachelor's Degree upon hire and have past experience providing financial literacy services.	Facilitators receive a half-day of annual refresher training in the policies and procedures of the program.
Workforce services	MFH utilized trained male and female facilitators, as did EM, the workforce development board. Facilitators hold a Bachelor's Degree upon hire and are certified in a variety of workforce and career development tools.	EM semi-annually provided workforce development training to all partner staff.
Access to Child Support Services	CSS utilized a trained male facilitator. Facilitators hold a Bachelor's Degree upon hire and have over 5 years of experience working as a child support professional.	Facilitators participated in ongoing staff training and development provided by MCCSS.
Case management	Case management services were provided by male and female facilitators who also served as curriculum facilitators. Case Managers hold a Bachelor's Degree upon hire and have past experience in case management. Initial case manager training was provided by Alma Center.	Case Managers participated in ongoing training provided by Alma Center.
Counterfactual		
Economic stability workshops	MFH utilized trained male and female facilitators, as did EM, the workforce development board. Staff typically hold a Bachelor's Degree upon hire and have past experience providing financial literacy services	Facilitators receive a half-day of annual refresher training in the policies and procedures of the program.
Workforce services	MFH utilized trained male and female facilitators, as did EM, the workforce development board. Facilitators hold a Bachelor's Degree upon hire and are certified in a variety of workforce and career development tools.	EM semi-annually provided workforce development training to all partner staff.
Access to Child Support Services	CSS utilized a trained male facilitator. Facilitators hold a Bachelor's Degree upon hire and have over 5 years of experience working as a child support professional.	Facilitators participated in ongoing staff training and development provided by MCCSS.

Component	Education and initial training of staff	Ongoing training of staff
Case management	Case management services were provided by male facilitators who also served as curriculum facilitators. Case Managers hold a Bachelor's Degree upon hire and have past experience in case management. Initial case manager training was provided by Alma Center.	Case Managers participated in ongoing training provided by Alma Center.

B. Description of counterfactual condition as intended

1. Program content

The counterfactual group received a defined program of services that consisted of a subset of services the intervention group received. Fathers in the counterfactual group received economic stability education, child support services, workforce services, and other support services such as case management. MFH, MCCSS, and EM delivered these services. See Table II.1 for a summary of the counterfactual group components.

2. Intended implementation

This training for the counterfactual group was delivered through 2-hour classes that took place once per week for three weeks, and a 1-hour class in the fourth week, for a total of seven hours of training. MFH delivered the curriculum and MCCSS and EM delivered the support services within the MFH organization. The counterfactual group had one educator delivering the lessons (aside from week three, when MCCSS and EM were facilitating their respective sessions). A sample curriculum implementation plan is included in Appendix B (Table B.2) as an example. Each workshop was designed to consist of 25-30 fathers. The entire study was designed to provide services to 22 cohorts of fathers.

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

Three Implementation Evaluation research questions were jointly developed by AMTC & Associates, MCCSS, and MFH. The questions were:

1. Were all intended intervention components offered and for the expected duration?
2. Did the intended persons deliver services to fathers?
3. What was the program completion rate for fathers?

These three questions were selected to evaluate whether the program was delivered as intended. The findings from these questions were designed to assess program fidelity, staffing, and dosage, to inform the evaluators of any potential issues with program delivery or the survey data, and to inform future program planning/delivery. For example, if the program sessions were not delivered, not delivered in order, or a father did not receive 75% of the sessions, this could lead to smaller entrance-to-exit survey impacts of the program than would be expected.

Each implementation element (Fidelity, Staffing, and Dosage) and accompanying research question was analyzed by two measures. The fidelity implementation element was measured by percentages of cohorts that delivered all intended sessions and percentage of sessions delivered in order. The staffing implementation element was measured by the same staff delivering content to fathers and the percentage of staff trained. The dosage implementation element was measured by clients attending (1) the first session, (2) at least one session, (3) 75% of sessions, and (4)

100% of sessions. These implementation elements, evaluation questions, and measures are summarized in Table II.3. A summary of how program completion was computed is also provided in Table II.3. See Section III.B (on Data Collection) for a detailed description of how each measure was utilized for each research question.

Table II.3. Measures for addressing implementation research questions

Implementation element	Research question	Measures
Fidelity	Were all the intended intervention components offered and for the expected duration?	• Percentage of cohorts that were delivered to all intended sessions (i.e., 19 sessions for treatment group and 4 for the counterfactual group) within the workshop • Percentage of sessions delivered in order of intended implementation plans
Staffing	Did the intended staff deliver services to fathers and receive training?	• Same staff delivering content to fathers across sessions • Percentage of staff trained
Dosage	What was the program completion rate for fathers?	• Number and percentage of sessions clients attended (fathers can be counted in any one or all of the categories below): – first session – at least one session – 75% of sessions – 100% of sessions • Percentage of the sample that did not attend any sessions

III. Study design

This study utilized a randomized control trial (RCT) research design, and this section provides an overview of the formation of the sample of participants and the data collection procedures. Subsection A describes the sample formation, consent process, and random assignment of participants into groups in accordance with an RCT design. Subsection B describes the data collected for both the implementation and impact evaluations.

A. Sample formation

1. Sample recruitment

Participants for both the treatment and control groups were largely recruited through referrals to the program by the MCCSS office, word of mouth, and some community recruitment efforts. Near the end of the study, a professional recruiter was hired by MCCSS to assist with recruitment into the study. A screening form was developed and used to ensure that any potential participant met the inclusion criteria (i.e., non-custodial fathers who live in Milwaukee County, are not currently incarcerated, are at least 18 years of age, and have at least one child who is between the ages of 0 and 21). To encourage enrollment and participation, about halfway through the study an incentive (in the form of a gift card) was offered to those attending a study orientation session, and additional gift cards were made available to fathers who attended sessions and/or completed evaluation surveys.

2. Consent process

A comprehensive process, approved by an Institutional Review Board (IRB)[1], was used to obtain consent. Consent forms were distributed to all potential participants and reviewed at an orientation session led by staff from MFH, AMTC & Associates, and the University of Wisconsin – Milwaukee (UWM). Potential participants attended the orientation in either a larger-group or one-on-one setting, and reviewed the consent forms individually with a staff member from UWM or AMTC & Associates. Consent forms included information on the purpose of the study as well as a detailed description of the time commitments and information that will be required from participants. Several opportunities were provided for fathers to review the form and ask questions prior to consenting. The consent process was identical for all participants across the study.

3. Random assignment process

Participants were randomly assigned into a treatment or control group in accordance with an RCT design. The unit of randomization in this study was the individual. Individuals were

[1] IRB Committee #1 at the University of Wisconsin – Milwaukee approved the study design and data collection procedures used. Initial IRB approval occurred on October 20, 2016. Continuing annual IRB reviews were completed on October 19, 2017 and October 17, 2018. A post-approval IRB review—a comprehensive audit of study design, data collection procedures, and data security—was completed on May 3, 2019.

recruited and enrolled in the study in cohorts. Cohorts were formed once a month, creating about 12 cohorts per year (these ranged from 5 to 35 participants, with an average of about 19 participants per cohort). Within each cohort, individual fathers were randomly assigned to either the treatment or control condition (group) on the day of their orientation into the study, after giving consent to participate and completing the baseline UWM PFF surveys and Information, Family Outcomes, Reporting, and Management (nFORM) survey. The UWM surveys were developed by the evaluators to evaluate the outcomes of interest, and are described further in Subsections III.B.2 and IV.B. The nFORM survey was part of the client management system that the Office of Family Assistance (OFA) provided to 2015 grantees, and was used (in part) to collect demographic data prior to group assignment, log class attendance data, and collect some data about fathers' beliefs and behaviors.

At the end of the orientation session, and after participants provided their consent and completed the baseline surveys, a member of the evaluation staff at UWM used Qualtrics, a web-based survey software, to randomly generate a group assignment for each participant (the software was programmed such that each participant would have an equal probability of being assigned into either the control or treatment group). Program evaluators then informed each participant of their group assignment, so participants (and researchers) were unaware of group assignment prior to participants providing consent and completing baseline surveys. Although randomization was conducted at the individual level, it occurred on a rolling basis (i.e., monthly, within each cohort). No stratification or blocking was used, and no sub-sampling occurred after random assignment. The intended probability for group assignment was 0.50 (i.e., an equal chance of being assigned to either the treatment or the control group). Final enrollment data confirms that this was achieved, with 50.11% of fathers assigned to the treatment group and 49.89% of fathers assigned to the control group.

B. Data collection

This section summarizes the data collection processes. First, we discuss the data collection methods used to answer the three implementation evaluation questions, followed by a discussion of the data collection methods used for the impact evaluation. Both evaluations used and relied on the nFORM management system, a web-based platform that was used to collect baseline demographic information, attendance data, and some data about fathers' beliefs and behaviors (e.g., in regards to economic stability).

1. Implementation evaluation

Data sources for the implementation evaluation are described here by research question. This information is also summarized in Table B.3 in Appendix B.

Implementation Research Question 1: Were all intended intervention components offered and for the expected duration?

Measure: Percentage of cohorts that delivered all intended sessions within the workshop

Prior to implementing the evaluation, AMTC, UWM, and MFH worked together to develop a program implementation plan for each cohort of fathers. The schedule included the day and time of each session, the specific curricula and lessons to be facilitated or support service provider to facilitate each session, the length of time for each lesson or support service provider, and the total length of each session and the educator for each session. AMTC reviewed all workshops entered in nFORM for both treatment and counterfactual groups. AMTC committed to quarterly attendance reviews to ensure data was entered into nFORM accurately. This consisted of AMTC reviewing physical sign-in sheets for each session day/lesson and comparing these to data entry into nFORM. The original implementation plan consisted of 19 sessions for the treatment group and 4 sessions for the counterfactual group.

Measure: Percentage of sessions delivered in order of intended implementation plans

Each implementation plan was sent to and ultimately approved by AMTC prior to the start of each group/cohort. Curricula order and selection of lessons were approved by each curriculum author. Marline Pearson, PREP, approved the order and selection of designated relationship components and lessons. Jeffrey Johnson, NPCL, approved the order and selection of designated Fatherhood Development components and lessons. National Fatherhood Institute (NFI), approved the order and selection of designated 24/7 DAD components and lessons. Pearson, Johnson and NFI also approved the modification to add cohorts every four weeks that would alternate whether fatherhood or relationship education would start the treatment workshops. Additionally, at each series session, each father signed in on the attendance sheet for that day to confirm his attendance. The MFH staff collected and entered attendance for each session into the nFORM system within 72 hours of the session ending. On a quarterly basis, AMTC compared the hard copy attendance sheets, nFORM attendance records and program implementation plan to assess data entry quality and to determine if the sessions adhered to the planned workshop schedule. AMTC then worked with MFH staff to correct any discrepancies.

Implementation Research Question 2: Did the intended staff deliver services to fathers and receive training?

Measure: Same staff delivering content to fathers across sessions

MFH and MCCSS jointly agreed that MFH would be the program partner to implement the impact evaluation because of their extensive experience providing quality fatherhood services within Milwaukee for the past 10 years. The educator layout for all sessions is indicated in Table B.4 (Appendix B). Note that on Monday through Wednesday the same session was held both in the AM and PM for flexibility. AMTC reviewed the program implementation plan in paper form and again after it was entered into nFORM. These sources provided the name of the educator prior to the start of each session series and were used to confirm that the correct educator was facilitating the proper components of the study.

Measure: Percentage of staff trained

In year one AMTC reviewed hiring records to confirm staff credentials and curriculum training logs to confirm completion of required curriculum training. Throughout the evaluation period, two trained Father Advocates from MFH were expected to facilitate all parenting and healthy marriage/relationship education sessions and another Father Advocate facilitated the economic stability (banking and credit) education sessions. Support services were delivered by funded support service providers. A trained representative from Child Support facilitated a Child Support 101 session and screened fathers for individual services. A trained representative from Employ Milwaukee facilitated a workforce 101 session to ensure fathers understood the services and eligibility requirement for services funded by the workforce board, and a trained representative from Wisconsin Community Services facilitated a session and screened fathers for individual driver's license reinstatement services.

Implementation Research Question 3: What was the program completion rate for fathers?

Measure: Number (or percentage) of sessions attended (including % that did not attend sessions at all)

Dosage was measured by two standards: completing 75% and completing 100% of program sessions. Data was collected by AMTC after each session and used to monitor fathers' progress through the program. Data collection included hard copy attendance sheets from each session and nFORM data reports for each session. On a quarterly basis, AMTC worked with MFH to complete a quality assurance log, which consisted of a side-by-side comparison of each hard copy attendance sheet and daily nFORM session entry. Based on the outcome of these reviews, AMTC was able to verify the completion rate for fathers that participated in the program. It was determined that an nFORM attendance entry differed from the hard copy attendance sheet on two occasions over the course of the entire study. AMTC found that on these two occasions there were more fathers that signed in on the attendance sheet than what was entered into nFORM. AMTC worked closely with the MFH educator and determined each of the discrepancies were entered in error. Upon discovering the discrepancy, the error was brought to the attention of MFH and immediately manually corrected by AMTC staff.

2. Impact evaluation

Data for the impact evaluation were collected by the evaluation team using Qualtrics (for the UWM PFF surveys) and nFORM. Qualtrics is an online software platform that allows researchers to create and distribute surveys online. Survey responses are automatically saved to Qualtrics' secure servers and are accessible only to the researchers who created the survey. Although in this study many respondents choose to take the surveys on paper or via phone, these survey responses were subsequently entered into Qualtrics by the researchers so as to have a single repository of participant data. There was no indication that the groups (treatment and control) differed in terms of the survey modes utilized in this study.

Data were collected at three time points: (a) **Pre-Survey/baseline**, prior to the onset of services (Surveys Administered: UWM PFF Pre-Survey, nFORM Entrance and Applicant Characteristics Survey); (b) **Post-Survey**, upon completion of services (Surveys Administered: UWM PFF Post-Survey and nFORM Exit Survey); and (c) **Follow-Up Survey**, six months after completion of services (Survey Administered: UWM PFF Follow-Up Survey). The survey completion process utilized on-site computers or tablets (at MFH) to minimize any data collection problems. The Pre-Surveys were administered by UWM and AMTC staff prior to random assignment, following a study orientation session, approximately one week prior to the start of the PFF program.

The Post-Surveys were typically administered at the conclusion of the program, after the last regular classroom instruction. This purpose of this was to ensure that the groups would be comparable in the sense that participants responded to the Post-Surveys as close to program completion as possible. Because of the differences in programming for participants in the treatment and control groups—treatment group participants received 42 hours of programming while those in the control group received 8 hours—the timing of Post-Survey administration differed by group, depending on when programming was completed for each. Specifically, control group participants completed the Post-Survey approximately one month after the administration of the Pre-Survey while treatment group participants completed the Post-Survey approximately two months following the administration of the Pre-Survey.

For both groups, the Follow-Up survey was completed approximately six months following the completion of the Post-Survey (i.e., seven months after the Pre-Survey for the control group and eight months after the Pre-Survey for the treatment group). The timing for the Post- and Follow-Up surveys is approximate due to two reasons: (1) the length of the training program sometimes varied due to holidays, snow days, etc., and (2) participants had a month-long window to take their Follow-Up Survey, meaning they could take it up to two weeks before or after the exact six-month date.

At all three time points the UWM PFF surveys were available for participants to take online at MFH. An option to answer the UWM PFF Post- and Follow-Up surveys on paper at MFH was added in April 2017. An option to answer the UWM PFF Post- and Follow-Up surveys online via an individual link sent to the father's phone or email was added in May 2017. Finally, in December of 2017, participants were given the option to complete the UWM PFF Post- and Follow-Up surveys over the phone if desired. These different modes for survey completion were made available in an effort to encourage high completion rates; it was hypothesized that certain participants might be more likely to choose one mode over another due to convenience, comfort with technology (such as smart phones), and personal preferences. All modes, except for the phone option, were self-administered. For the phone option, two study personnel, trained in phone survey administration, read the survey questions to participants and entered the responses into the Qualtrics survey software. While it is always possible that phone surveys introduce increased social desirability bias, the only study personnel allowed to conduct phone surveys had extensive experience administering surveys in this fashion to ensure that any social desirability effects were reduced to the greatest degree possible. UWM's evaluation team, AMTC, and

MCCSS reviewed the numbers of surveys completed on a monthly basis. There was no indication that the survey modes differed by group. The data collection processes are summarized in Table B.5 in Appendix B.

IV. Analysis methods

A. Analytic sample

In total, 482 participants were enrolled in the study. Of the 482 participants enrolled, it was discovered that 11 individuals who enrolled and took the UWM PFF and nFORM Pre-Surveys then dropped out of the training program and re-enrolled sometime later without informing the researchers. These individuals were included in the overall sample but were not included in any analytic samples because their UWM PFF Post- and Follow-Up surveys were not completed in the same time range as all other participants. Participants were given a one-month grace period for filling out the surveys in order to reduce the likelihood that timing effects could confound the study results.

Of the 482 unique participants enrolled in the study and assigned to either the treatment or control condition, all contributed baseline surveys. Of this total study sample, 320 participants provided responses to the UWM PFF Post-Survey and nFORM Exit Survey, and 338 participants provided responses to the UWM PFF Follow-Up Survey six months after their training ended. Of interest is the fact that more participants provided responses to the UWM PFF Follow-Up Survey than the Post-Survey. This is somewhat unexpected, and may be due to the fact that the Post-Survey was administered immediately after the final session of the program; that is, some participants stopped attending the program before reaching the final session and did not take the Post-Survey, but over the subsequent six months they were reached by the evaluators and received several reminders to take the Follow-Up Survey. When a participant did not complete an assigned survey (Post- or Follow-Up), MFH or AMTC staff attempted to contact the participant to schedule a make-up time to take it. Participants who were contacted a total of three times without response were considered to have revoked their consent to participate in the study and were not contacted again (in accordance with approved IRB procedures).

Two analytic samples were created from the total pool of (482) participants who provided survey data. These analytic samples represent the actual samples that were subsequently used in the analyses and, in accordance with the Office of Planning, Research, and Evaluation (OPRE) and OFA guidelines, the analytic samples excluded participants with incomplete or missing data on relevant variables. A participant was considered to have **fully completed** a survey subscale if he completed at least 80% of the subscale items (for example, if he completed at least 8 of the 11 items on the Healthy Co-Parenting subscale). Thus, participants who partially completed the UWM PFF Post- or Follow-Up Surveys (i.e., did not respond to at least 80% of all items in each subscale), or who did not provide data on baseline covariates needed for the analyses, were not included in the corresponding analytic samples. Baseline covariates are further discussed in Section IV.C.

The first analytic sample consisted of participants who fully completed (as described above) all subscales on both the UWM PFF Pre- and **Post-**Surveys, and provided baseline nFORM information that could be used as covariates. Of the 320 participants who provided responses on these surveys at the post-study time point, **265 were included in the first analytic sample**, used

to examine the primary research questions and the seventh additional research question. This seventh additional research question evaluated the effect of treatment on economic stability at the post-study time point, while all other additional (secondary) research questions tested the effect of treatment at the six-month follow-up time point; the reason for this is that the economic stability survey was only administered at program exit.

The second analytic sample consisted of participants who fully completed all subscales on both the UWM PFF Pre- and Follow-Up surveys, and provided baseline information on nFORM that could be used as covariates. Of the 338 participants who provided responses on the relevant surveys, 301 were included in the second analytic sample. This analytic sample was used to answer the first six additional research questions, which all involved the follow-up time point.

The sample size information is summarized in Table IV.1. Considering the attrition from the full sample to the analytic samples, the overall attrition rate for the first analytic sample is 45% and the differential attrition rate (difference in attrition rates between the groups) is 6.3%. Based on this information, the sample did not qualify as a low-attrition randomized control trial (RCT) under the conservative threshold guidelines from the What Works Clearinghouse (WWC). We thus treat the study as a high-attrition RCT, and this necessitates that we adjust the analyses by controlling for any covariates that do not show baseline equivalence among the groups, in accordance with WWC guidelines. The assessment of baseline equivalence is discussed in detail in Subsection IV.C. We note that the second analytic sample did qualify as a low-attrition RCT but, for the sake of consistency, we assessed baseline equivalence and controlled for covariates that were in the adjustment range (i.e., did not meet baseline equivalence requirements) for both samples.

Table IV.1. Individual sample sizes by intervention status

Number of individuals	Intervention sample size	Comparison sample size	Total sample size	Total response rate	Intervention response rate	Comparison response rate
Assigned to condition	243	239	482	n.a.	n.a.	n.a.
Contributed a baseline survey	243	239	482	100%	100%	100%
Contributed to Post-Study Survey	126	139	265	55.0%	51.9%	58.2%
Contributed to Follow-Up Survey	152	149	301	62.4%	62.6%	62.3%

Notes: The numbers reported account for item non-response and any other analysis restrictions; n.a. = not applicable.

The baseline survey was administered prior to the start of the program for both groups; the Post-study survey was administered one month (control group) or two months (treatment group) following the baseline survey; and the Follow-Up survey was administered seven months (control group) or eight months (treatment group) following the baseline survey.

B. Outcome measures

All outcome measures were created from multiple items on the UWM PFF surveys. Detailed descriptions of the subscales used for the primary and additional research questions are included in Tables IV.2 and F.1 (in Appendix F), respectively. One score was created for each subscale (at each time point) by calculating the mean score of all items on that subscale for participants with complete data (as defined in Section IV.A). That is, subscale scores were calculated by summing the individual item responses on a subscale, then dividing this total by the number of items answered to obtain an average item score for each subscale. For example, suppose that a respondent answered 10 of the 11 items on the Co-parenting Relationship Scale, where each item is scored from 1 to 5. If the respondent's responses on the 10 items summed to 45, this respondent's average score on this subscale would be 45/10 = 4.5.

These average subscale scores were the outcome measures used to address the primary and additional research questions. We note that, for the Conflict with Co-Parent (Outcomes 6) and Economic Stability (Outcome 7) subscales, higher scores represent less desirable outcomes for this study (i.e., more conflict with co-parent and more negative beliefs about one's economic stability). For all other subscales (Outcomes 1 through 5), higher scores are indicative of more desirable outcomes for this study (i.e., more skills learned, higher endorsement of positive beliefs, etc.). We also note that, for items that referred to the father's child, the father was instructed to respond regarding his youngest child. Reliability results for each outcome, as measured by Cronbach's α, are also included in Tables IV.2 and F.1 (for the primary and additional research questions, respectively).

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

Outcome measure	Description of the outcome measure	Timing of measure
Outcome Measure 1: Healthy Co-Parenting	The outcome measure is the 11-item Co-parenting Relationship Scale adapted from Dyer, Fagan, Kaufman, Pearson, & Cabrera (2015). All items are on a five-point Likert scale ranging from "Strongly disagree" to "Strongly agree". We reverse coded items 1, 2, 3, 9, 10, and 11. Mean outcome measure scores were derived by adding all completed items (at least 9 of 11 to meet the 80% analytic sample threshold) together then dividing by the number of items completed. Responses were coded such that higher scores are considered to indicate better co-parenting skills. Cronbach's α: .940	Immediately following the completion of the intervention
Outcome Measure 2: Communication with Co-Parent	The outcome measure is the 11-item Communication with Co-parent Scale created by the local evaluators. All items on the scale are on a five-point Likert scale ranging from "Strongly disagree" to "Strongly agree". Mean outcome measure scores were derived by adding all completed items (at least 9 of 11 to meet the 80% analytic sample threshold) together then dividing by the number of items completed. Responses were coded such that higher scores indicate better communication skills Cronbach's α: .974	Immediately following the completion of the intervention
Outcome Measure 3: Conflict Resolution Skills	The outcome measure is the 8-item Resolving Conflict with Co-parent Scale created by the local evaluators. All items are on a five-point Likert scale ranging from "Not at all confident" to "Completely confident". Mean outcome measure scores were derived by adding all completed items (at least 7 of 8 to meet the 80% analytic sample threshold) together then dividing by the number of items completed. Responses were coded such that higher scores indicate better conflict resolution skills. Cronbach's α: .965	Immediately following the completion of the intervention
Outcome Measure 4: Positive Parenting Skills	The outcome measure is the 10-item Beliefs about Positive Parenting Scale created by the local evaluators. All items on the scale on a four-point Likert scale ranging from "Not at all important" to "Very important". Mean outcome measure scores were derived by adding all completed items (at least 8 of 10 to meet the 80% analytic sample threshold) together then dividing by the number of items completed. Responses were coded such that higher scores are considered to indicate better parenting skills. Cronbach's α: .931	Immediately following the completion of the intervention

Outcome measure	Description of the outcome measure	Timing of measure
Outcome Measure 5: Engagement with Child	The outcome measure is the 11-item Engagement with Child Scale created by the local evaluators. All items on the scale are on a six-point scale ranging from “Never” to “More than 10 times”. Responses were coded to capture the increasing frequency of behavior measured by the scale (i.e., Never = 0, Once = 1, Twice = 2, Two to five times = 4, Six to ten times = 8, and Ten or more times = 16). Mean outcome measure scores were derived by adding all completed items (at least 9 of 11 to meet the 80% analytic sample threshold) together then dividing by the number of items completed. Responses were coded such that higher scores are considered to indicate higher levels of engagement. Cronbach’s α: .959	Immediately following the completion of the intervention
Outcome Measure 6: Conflict with Co-Parent	The outcome measure is the 7-item Conflict with Custodial Parent Scale created by the local evaluators. All items on the scale are on a six-point Likert scale ranging from “Never” to “More than 10 times”. Responses were coded to capture the increasing frequency of behavior measured by the scale (i.e., Never = 0, Once = 1, Twice = 2, Two to five times = 4, Six to ten times = 8, and Ten or more times = 16). Mean outcome measure scores were derived by adding all completed items (at least 6 of 7 to meet the 80% analytic sample threshold) together then dividing by the number of items completed. Responses were coded such that higher scores indicate higher levels of conflict (less desirable outcome). Cronbach’s α: .877	Immediately following the completion of the intervention

Note: Source of all measures was the UWM PFF Survey. Reliability results are based on the first analytic sample.

C. Baseline equivalence and sample characteristics

The equivalence of the study groups at baseline was examined for several demographic characteristics collected through nFORM (e.g., race/ethnicity, income level, age, education, employment status, and marital status) and on baseline outcome measures (i.e., the subscales of the UWM PFF Pre-Survey). The original demographic questions from the nFORM survey were recorded in several categories for each question (for example, there were six response options for income level). An examination of the response distributions showed that the responses could be reasonably summarized in two categories for each variable. As such, the categorical demographic variables were re-coded into dichotomous variables (for example, the six response options for income level were summarized by using "under $500 per month" and "$500 or more per month" as response categories).

Chi-squared tests and t-tests were used to test for significant group differences on the categorical demographic variables and the continuous survey subscale scores, respectively. Statistical significance tests used the $\alpha = .05$ level (two-tailed) to determine significance. Corresponding effect size measures (Hedge's g and Cox's index for continuous and dichotomous variables, respectively) were used to evaluate any differences between the treatment and control groups at baseline using the appropriate analytic sample (e.g., the first analytic sample if the outcome analyzed was immediately post-study). Because this study is treated as a high-attrition RCT, in accordance with WWC guidelines effect sizes of .05 or less would satisfy baseline equivalence, effect sizes above .05 but less than or equal to .25 would necessitate inclusion of the variables (for adjustment) in the analyses, and effect sizes greater than .25 would not satisfy equivalence. Table IV.4 summarizes the results of the significance tests and effect sizes (used to determine baseline equivalence) for the first analytic sample.

Using the first analytic sample, baseline equivalence was not met for any of the demographic variables (i.e., effect sizes were between .05 and .25). Additionally, effect sizes for four of the baseline subscale measures were also between .05 and 0.25. Thus, all of these variables were included as covariates in the primary analyses, as will be described in Section V.B.

Table IV.4. Summary statistics of key baseline measures and baseline equivalence across study groups, for individuals completing the UWM Post-Survey

Baseline measure	Intervention mean (standard deviation)	Comparison mean (standard deviation)	Intervention versus comparison mean difference (p-value of difference)	Effect size
Race/ethnicity (%)				
Black	92.1 (0.27)	92.8 (0.26)	0.7 (.819)	.061
Other	7.9	7.2		
Income level (%)				
Less than $500/month	46.0 (0.50)	38.1 (0.49)	7.9 (.193)	.196
$500/month or more	54.0	61.9		
Age (%)				
44 years and younger	73.0 (0.44)	74.8 (0.43)	1.8 (.738)	.056
45 years and older	27.0	25.2		
Education (%)				
High school diploma or less	66.7 (0.47)	64.0 (0.48)	2.7 (.652)	.072
Some college or more	33.3	36.0		
Employment (%)				
Employed (including occasional work)	57.9 (0.49)	61.9 (0.49)	4.0 (.514)	.101
Not currently employed	42.1	38.1		
Marital Status (%)				
Married (currently or previously)	34.9 (0.48)	38.1 (0.49)	3.2 (.588)	.083
Never married	65.1	61.9		
1: Co-Parenting	3.499 (1.11)	3.556 (1.06)	0.057 (.671)	.052
2: Communication with Co-Parent	3.346 (1.35)	3.391 (1.31)	0.045 (.780)	.034
3: Conflict Resolution Skills	3.537 (1.17)	3.595 (1.12)	0.058 (.680)	.050
4: Positive Parenting Skills	3.708 (0.39)	3.664 (0.44)	0.044 (.384)	.106
5: Engagement with Child	10.143 (5.47)	10.188 (5.58)	0.045 (.948)	.008
6: Conflict with Co-Parent	0.627 (1.43)	0.759 (1.95)	0.132 (.533)	.076
Sample size	126	139		

Notes: Results are based on the first analytic sample.

Demographic variables were dichotomized from original multi-category variables.

Effect sizes are calculated using Cox's Index for dichotomous demographic variables and Hedge's g for continuous subscale variables.

Standard deviations are included in parentheses in the "Intervention mean" and "Comparison mean" columns, and p-values are included in parentheses in the "Intervention versus comparison mean difference" column.

V. Findings and estimation approach

A. Implementation evaluation

1. Key findings

In evaluating the implementation evaluation research questions during and at the completion of the study, one question measured fidelity, the second question measured staffing, and the third question measured dosage. The intervention was implemented with a high level of fidelity for both treatment and control groups, with facilitators covering the content of the sessions in the intended order and delivering the sessions for the expected length of time (question 1). The intended staff delivered services to fathers and received training (question 2). Finally, the treatment and control groups did not exhibit high program completion rates (question 3).

2. Analysis details

AMTC compiled data from nFORM on the number of sessions and attendance; training logs and organization implementation plans; and attendance quality assurance logs. AMTC used this information to address the three research questions. Below is the analysis and results for each research question.

Implementation Research Question 1: Were all intended intervention components offered and for the expected duration?

Measure: Percentage of cohorts that delivered all intended sessions within the workshop

The implementation evaluation consisted of 25 cohorts. Of the 25 treatment cohorts, 21 (84%) implemented the 19 sessions as planned, and of the 25 counterfactual cohorts 23 (92%) implemented the 4 sessions as planned. Primary reasons for these differences included three cohorts that had sessions cancelled due to inclement weather and one cohort that experienced building/maintenance problems that led to a cancellation of sessions. Although MFH needed to cancel or reposition seven sessions from four cohorts due to weather and facility maintenance work, MFH rescheduled sessions and held make-up sessions on Fridays to ensure that all participants were able to make up the material that was missed during any canceled sessions.

Measure: Percentage of sessions delivered in order of intended implementation plans

Of the 25 treatment cohorts, four contained sessions (7 in total) facilitated out of sequence or eliminated; thus, out of the 475 sessions scheduled for delivery throughout the study, only 1.5% (7) of the sessions were cancelled or facilitated out of order. The other 21 treatment cohorts did not have any out-of-sequence lessons. Of the 25 counterfactual cohorts, none had lessons facilitated out of sequence because missed sessions were rescheduled prior to the next scheduled session.

Implementation Research Question 2: Did the intended staff deliver services to fathers and receive training?

Measure: Same staff delivering content to fathers across sessions

Program facilitation was consistent because the same staff/educator facilitated the sessions throughout the program. The only deviation occurred when an educator was sick or on vacation (which amounted to 12 sessions out of 575 sessions); however, another educator filled in during these sessions to ensure that no sessions would be facilitated out of sequence. There was also no turnover of the educators throughout the impact evaluation.

Measure: Percentage of staff trained

All (100%) of the Pathways staff at MFH were trained on all methodologies to implement the program with fidelity. The list of all training provided is detailed in Section II.A.

Implementation Research Question 3: What was the completion rate for fathers?

Measures: Number (or percentage) of sessions attended (including % that did not attend sessions at all)

The no-show rate (i.e., percentage of participants who did not attend any sessions) was 21% in the treatment group and 17% in the counterfactual group. In the treatment group, 52% of enrolled fathers attended at least 75% (at least 15 sessions) of the 19 sessions offered, and 13% attended 100% of the 19 sessions offered. In the counterfactual group, 72% of enrolled fathers attended at least 75% (at least 3 sessions) of the 4 sessions offered, and 43% attended 100% of the sessions offered.

3. Summary: Program implementation received

The impact evaluation implementation was facilitated with fidelity for both treatment and counterfactual groups. Sessions and lessons were delivered as intended and only deviated from the original plan when a session was cancelled for weather-related reasons. When a session was cancelled, make-up sessions were created to accommodate the missing days of implementation prior to or after the cancelled session. Only one deviation to the initial plan was made but it did not alter the fidelity of the model; namely, due to the need to serve more fathers and not have long wait periods between enrollment periods, UWM, MCCSS, AMTC and MFH decided to enroll participants every four weeks instead of (the initially proposed) every eight weeks. This meant that some fathers in the treatment group received the fatherhood lessons (Fatherhood Development and 24/7 DAD A.M.) prior to relationship lessons (Within My Reach) while some fathers received relationship lessons first, followed by fatherhood lessons. There was a natural break between fatherhood and relationship lessons, and this allowed MFH to introduce a new cohort of fathers during this break every four weeks. AMTC reviewed this sequencing with the curriculum authors and it was agreed that fidelity to the model was not compromised.

4. Limitations

There were a few limitations to the implementation evaluation. These limitations affected the third evaluation question (i.e., program completion rate). The two main limitations were that there was not enough space or staff time to conduct as many make-up sessions as were needed. MFH had one large conference room that could be split into two different spaces; however, this did not provide enough classrooms to hold multiple sessions at the same time. Discussed below in the Lessons Learned section, it is mentioned that MFH held sessions during the day, at night and make-up sessions on Fridays. However, even with that flexibility, fathers still needed more options to attend make-up sessions. Also, due to the heavy emphasis on recruiting, the MFH staff did not have enough time to hold as many make-up sessions as were needed for more fathers to complete the program.

5. Lessons learned

Lesson Learned 1: Completion

When developing a schedule for a program that is seven to eight weeks in length, there is a need for multiple options (in terms of days and/or times) for fathers to attend a certain session. MFH conducted the same session in the morning and again at night, with makeup sessions on Fridays, and even with this flexibility it was hard for fathers in the treatment group to attend the number of sessions needed to complete the program.

Lesson Learned 2: Cohort scheduling

Although the attrition rate from recruitment to enrollment was not a measure of the implementation study, MFH, UWM and AMTC noticed the time between cohorts should be minimized so that wait time to enroll into the program (and potentially passing up on enrolling) is reduced. The first two cohorts were enrolled eight weeks apart and it became apparent very quickly that there would not be a sufficient number of fathers enrolling into the study. This was due to a number of fathers showing interest in the program during recruitment, but then losing interest because they could not wait eight weeks. After completing the first two cohorts, changing enrollment to every four weeks and hiring a recruitment specialist greatly helped increase enrollment. As discussed previously, with this change some fathers started their session series with fatherhood lessons while the other half started with healthy relationship lessons.

Lesson Learned 3: Childcare

Although childcare was not designed to be a measure of completion in the implementation evaluation, MFH provided childcare during the sessions to support fathers whose only time to attend the program was when they had their child/ren. Offering childcare allowed fathers to have a comfortable and convenient alternative to dropping their children off at a formal childcare facility. While we did not measure whether this could be linked to increased attendance, MFH and AMTC believe it helped eliminate a potential barrier for attendance.

Table V.1. Completion rates by session (frequency; percent)

Group	Enrolled	No Sessions Attended	Attended First Session in cohort	Attended at least one session	75%*	75%**	100%*	100%**
Treatment	243	51; 21%	168; 69%	192; 66%	126; 66%	126; 52%	31; 16%	31; 13%
Counterfactual	239	41; 17%	177; 74%	198; 83%	172; 88%	172; 72%	103; 52%	103; 43%
Total	482	92; 19%	345; 72%	390; 81%	298; 76%	298; 62%	134; 34%	134; 28%

Notes: *Out of those that attended at least ONE session

**Out of those that were ENROLLED

75% Completion = 15 out of 19 sessions for the Treatment group and 3 out of 4 sessions for the Counterfactual group.

Fathers can be counted in any one or all of the categories/columns.

B. Primary impact evaluation

1. Key findings

In evaluating the outcome measures immediately at study completion (for the primary research questions), no statistically significant group differences were found. Although participants in the treatment group did show some gains on several outcomes (after controlling for a number of covariates) following the completion of the training program, the effect sizes were small and no statistically significant differences were found in comparison with the control group. The results are described below.

2. Analysis details

Multiple linear regression modeling was used to answer all of the primary research questions. The statistical software packages SAS 9.4 and SPSS 25 were used to conduct the analyses, and the statistical significance level was set at $\alpha = .05$ using two-tailed tests. No correction methods (e.g., Bonferroni, Tukey) were used to adjust this α-level. The analysis did not account for any clustering because enrollment and random assignment were conducted at the individual level.

Each outcome measure (Post-Survey subscale score) was modeled using (a) the UWM PFF Pre-Survey score on the same measure and (b) intervention group membership (a 0/1 binary variable, where 0 = Control, 1 = Treatment) as predictors. Additional covariates in the models included the demographic variables as well as the UWM PFF Pre-Survey subscale scores that did not satisfy baseline equivalence (see Table V.2). The value of the regression coefficient for group membership in this model indicates the (covariate-adjusted) difference between the group means on the outcome measure. A statistically significant effect of the group membership regression coefficient (at $\alpha = .05$, two-tailed) indicates a statistically significant difference between the groups, after adjusting for the covariates. All participants who met criteria for inclusion in the analytic sample were included in the analysis, regardless of whether or not they completed the intervention (in concordance with an intent-to-treat approach).

Table V.2. Covariates included in impact analyses

Covariate	Description of the covariate
Race/ethnicity	0 = Other, 1 = Black
Income	0 = Less than $500 per month, 1 = $500 or more per month
Age	0 = 44 years of age and younger, 1 = 45 years of age and older
Employment	0 = Not employed, 1 = Employed
Education	0 = High school diploma or less, 1 = Some college or more
Marital status	0 = Never married, 1 = Married
Healthy Co-Parenting Pre-Score	Healthy Co-Parenting mean subscale score from the UWM PFF Pre-Survey
Conflict Resolution Skills Pre-Score	Conflict Resolution Skills mean subscale score from the UWM PFF Pre-Survey
Positive Parenting Skills Pre-Score	Positive Parenting Skills mean subscale score from the UWM PFF Pre-Survey
Conflict with Co-Parent Pre-Score	Conflict with Co-Parent mean subscale score from the UWM PFF Pre-Survey

Note: Demographic variables were dichotomized by recoding the original multi-category variables.

3. Results of primary analyses

Multiple regression models were used to test group (treatment/control) differences in each of the UWM PFF Post-Survey subscale scores after controlling for the corresponding UWM PFF Pre-Survey subscale score as well as other covariates (see Table V.2). The relevant (group effect) results are summarized in Table V.3, and the full regression results (where the focus is on the "Group" effect) are provided in Table D.1 (Appendix D).

The last two columns of Table V.3 are used to compare the groups on each outcome of interest. The mean difference column represents the difference between the group means in terms of the subscale scores, adjusted for the covariates; positive values for these differences indicate that the treatment group scored higher than the control group on average (in terms of adjusted mean on the outcome). The p-value for the statistical test of this difference is provided in parentheses (where p-values below 0.05 are considered statistically significant). The effect size column reports the difference between the group means on a standardized scale, representing this difference in terms of standard deviations. Using the Healthy Co-Parenting measure as an example, the difference in adjusted means is about 0.06; thus, on average, the treatment group scored 0.06 points higher on the Healthy Co-Parenting scale (representing slightly more positive co-parenting attitudes) than the control group. The p-value is 0.527, indicating that this difference is not statistically significant. The effect size for the Healthy Co-Parenting scale is about 0.085, indicating that the difference between the treatment and control group means on this outcome represents less than one tenth of a standard deviation. Thus, these results point to a very small difference between the groups on this scale. The other results can be similarly interpreted. We note that the treatment group was expected to score higher than the control group on the first five outcomes listed in Table V.3 (Healthy Co-Parenting, Communication with Co-Parent, Conflict Resolution Skills, Beliefs about Healthy Parenting, and Engagement with Child subscales), and to score lower than the control group on the Conflict with Co-Parent subscale.

The results (see Table V.3) indicate that none of the group differences reached statistical significance (all p-values are well above 0.05). For the Healthy Co-Parenting, Communication with Co-Parent, and Conflict Resolution Skills outcomes, results are in the expected direction in that the treatment group scored higher than the control group. Similarly, for the Conflict with Co-Parent outcome the results are also in the expected direction in that the treatment group scored lower than the control group. For the remaining two outcomes, Beliefs about Positive Parenting and Engagement with Child, results are not in the expected direction as the treatment group scored lower than the control group. None of these group differences were statistically significant and their effect sizes were small. The largest effect size, of 0.264, was found for the Conflict Resolution Skills outcome; the other effect sizes are considered quite small. Overall, after controlling for the covariates, the group differences on these outcomes at program completion were negligible.

Table V.3. Post-intervention estimated effects using data from UWM Post-Survey to address the primary research questions

Outcome measure	Intervention mean or % (standard deviation)	Comparison mean or % (standard deviation)	Intervention compared to comparison mean difference (p-value of difference)	Effect size
Healthy Co-Parenting (Outcome 1)	3.684 (0.702)	3.624 (0.691)	0.060 (.527)	0.085
Communication with Co-Parent (Outcome 2)	3.448 (0.759)	3.365 (0.772)	0.083 (.463)	0.108
Conflict Resolution Skills (Outcome 3)	3.541 (0.649)	3.364 (0.690)	0.177 (.096)	0.264
Beliefs about Positive Parenting (Outcome 4)	3.572 (0.273)	3.603 (0.306)	-0.031 (.596)	0.105
Engagement with Child (Outcome 5)	9.296 (2.959)	9.739 (2.945)	-0.443 (.426)	0.150
Conflict with Co-Parent (Outcome 6)*	0.502 (0.458)	0.603 (0.571)	-0.100 (.502)	0.192
Sample Size	126	139		

Source: UWM PFF Post-Surveys administered immediately following the completion of the intervention.

Notes: Effect sizes were calculated using Hedge's g formula found in Appendix D. Reported means were estimated using the regression models.
Results are based on the first analytic sample.
See Table IV.2 for a more detailed description of each measure.

*For Conflict with Co-Parent (Outcome 6), lower scores indicate better outcomes (less conflict). For all other measures, higher scores indicate better outcomes.

C. Sensitivity analyses

In this section we discuss sensitivity analyses that were conducted to evaluate the robustness of the results from the planned analyses (reported in the previous section). Specifically, the sensitivity analyses tested if and how the group effects might change with the addition of a group by Pre-Survey score interaction effect for each outcome.

1. Key findings

Overall, the sensitivity analyses did not result in any changes (compared to the findings of the initial primary analyses) in terms of the statistical significance of the group effect. That is, the conclusions from the primary analyses, which indicated negligible group differences at the post-study time point, did not change.

2. Results of sensitivity analyses

The sensitivity analyses were conducted to test an added interaction between the effect of group and the Pre-Survey subscale score in each of the original models (i.e., after controlling for the same covariates as in the primary analyses). These interaction effects indicate whether the group effect on a particular subscale varies with the Pre-Survey subscale score. If the interaction effect is not significant, then the initial group effect results are robust in the sense that they are consistent (i.e., remain the same) regardless of one's Pre-Survey score. If the interaction effect is statistically significant, this will refine the original results by providing information on the Pre-Survey scores for which the group effects might be smaller or larger. The results of the interaction models are summarized in the last column of Table E.1 (Appendix E) under Primary Research Questions. Note that, because in this model the group effect can potentially vary depending on the value of the Pre-Survey score, the Pre-Survey score was mean-centered so the group effect in the last column of Table E.1 refers to the group difference specifically at the mean Pre-Survey score.

Significant interaction effects were found for the Healthy Co-Parenting (Outcome 1), Communication with Co-Parent (Outcome 2), and Conflict with Co-Parent (Outcome 6) subscales at the post-study time point. Interaction plots are provided in Figure E.1 to help with visualizing these results. The interaction effect was manifested similarly for both the Healthy Co-Parenting and Communication with Co-Parent outcomes, and it is represented by the plot in the left panel of Figure E.1. Namely, the group effect (where the treatment group scored higher than the control group) decreased significantly as the pre-study score increased. That is, the advantage observed for the treatment group on the post-study scores was generally larger for those with lower pre-study scores (and vice versa). The interaction effect for the Conflict with Co-Parent subscale is represented by the plot in the right panel of Figure E.1. It shows that for most individuals there was little if any effect of group, or a slight advantage (i.e., less conflict) for the treatment group. For the few individuals who reported high pre-study conflict, the treatment group reported higher levels of conflict than those in the control group, and this effect increased significantly as the pre-study score increased. In other words, more post-study conflict was

observed for the treatment group relative to the control group for individuals with larger conflict scores pre-study (and vice versa).

It is noted that when an interaction effect is not statistically significant (and not of primary interest), it can be removed from the model, resulting in the model we analyzed originally. Thus, for outcomes without a significant interaction, the results (regarding group effect) did not change. For the outcomes that showed a significant interaction, the original group effects were not significant; thus, while the interaction models add information on how the group effect varies depending on the pre-study score, they do not change the main findings as previously reported.

Finally, the group effects (reported in Table E.1 as the mean differences) were almost identical for the two models. That is, for each outcome, the group effect at the mean Pre-Survey score (in the interaction model) is nearly identical to the group effect observed in the primary analysis (model with no interaction). Thus, the main results did not change by allowing for an interaction effect.

D. Additional analyses

1. Key findings

Evaluating the outcome measures six months after study completion, the group differences were in the expected direction on all outcomes. The group difference was also in the expected direction for the Economic Stability measure (evaluated immediately at the completion of the program). These group differences were statistically significant for all outcomes except for the Conflict with Co-Parent outcome and the Economic Stability outcome. The results are summarized in Table F.3 (Appendix F) and described below.

2. Analysis details

The additional analyses focused on examining group effects on all outcomes (previously discussed for the primary analyses) at the six-month follow-up time point. A final additional research question included the economic stability outcome, which was measured immediately after completion of the intervention. The data collection processes, survey instruments, sample formation, research design, and analytic methods for the additional analyses were the same as previously detailed for the primary analyses, but used the second analytic sample (described in Section IV.A) for outcomes evaluated at the follow-up time point.

The second analytic sample included 301 participants who had complete data on all relevant variables at the baseline and Follow-Up Survey time points (see Section IV.A). The reliability of the UWM PFF Pre-Survey subscales using the second analytic sample can be found in Table F.1 (Appendix F). The second analytic sample met the WWC guidelines for a low-attrition RCT because there were both less overall attrition and less differential attrition in this sample (see Table IV.1). Thus, it was not required that the analyses utilizing this sample involve covariates that violate baseline equivalence. Nonetheless, for the sake of consistency with the primary

analyses and improved precision, the decision was made to adjust for covariates that did not satisfy baseline equivalence.

Baseline equivalence results using the second analytic sample are summarized in Table F.2 (Appendix F). Baseline equivalence was not met (i.e., effect sizes were above .05, but less than .25) for income level, education level, and employment status. Additionally, effect sizes for the baseline measures on all of the UWM PFF Pre-Survey subscales were between .05 and .25. All of these variables (that did not satisfy equivalence) were thus included in the subsequent analyses as covariates.

3. Results of additional analyses

Multiple regression models were used to test group (treatment/control) differences in each of the UWM PFF Follow-Up Survey subscale scores after controlling for income level, education level, employment status, and the baseline measures on all of the UWM PFF subscales. The relevant (group effect) results are summarized in Table F.3, and the full regression results (where the focus is on the "Group" effect) are provided in Table F.4.

The last two columns of Table F.3 are used to compare the groups on each outcome of interest, (as described in Section V.B). The results indicate that the group differences were statistically significant (the p-values are below 0.05) and in the expected direction (i.e., the treatment group scored higher than the control group) for the first five Follow-Up outcomes; namely, Healthy Co-Parenting, Communication with Co-Parent, Conflict Resolution Skills, Beliefs about Positive Parenting, and Engagement with Child. In addition, the effect sizes for all of these differences were large (ranging from about 1.2 to 2.0). As these analyses were exploratory and the large effect sizes are rather unusual, these results should be interpreted with caution (see Discussion Section).

For the last two outcomes in Table F.3, Conflict with Co-Parent and Economic Stability, the results were not statistically significant but were in the expected direction in that the treatment group scored lower than the control group. The effect sizes for these two outcomes were 0.2 and 0.1, which are considered relatively small.

4. Additional sensitivity analyses

Two sensitivity analyses were conducted to check whether the results of the additional analyses would be affected by changing certain features of the analytic approach. The first sensitivity analysis was the same as that conducted with the primary analyses; namely, it tested how the group effects might change with the addition of a group by Pre-Survey interaction effect. The second sensitivity analysis adjusted the scaling approach used to measure two of the outcomes to examine whether this may affect the results. These particular subscales were scaled such that scores increase exponentially with response categories, so this sensitivity analysis was conducted to determine whether the results would be robust if the scoring scales were changed for these two measures.

The results of the first sensitivity analysis, with an added interaction between the effect of group and the pre-study score (after controlling for the same covariates as in the benchmark additional analyses), are summarized under "Additional Research Questions" in the last column of Table E.1 (Appendix E). The group effect in this last column of Table E.1 refers to the group difference specifically at the mean pre-study score (because it was mean-centered).

A significant interaction effect was found only for the Conflict with Co-Parent outcome at the Follow-Up time point. The nature of this interaction was very similar to that presented in the right panel of Figure E.1 and can be interpreted as it was for the primary sensitivity analyses. That is, for most individuals there was little if any effect of group, or a slight advantage (i.e., less conflict) for the treatment group. More conflict was observed for the treatment group relative to the control group for the few individuals with larger pre-study conflict scores. For all other outcomes, the interaction effect was not statistically significant so the results were not changed; the interaction can be removed from the model and the originally reported results would be obtained.

The second sensitivity analysis examined the scaling of the Engagement with Child and Conflict with Co-Parent outcomes. It was hypothesized that the results observed may have been influenced by the coding scheme used for these scales, which were scored to represent the non-linear nature of the response categories. Specifically, items on these scales asked "how often" participants exhibited certain behaviors (such as helping their child with homework or yelling at their co-parent) in the previous month. Because the frequency options were not linear, and reflect exponentially increasing frequencies given the fixed time period (i.e., the previous month), they were originally coded to reflect the exponential nature of the frequency options; namely, Never = 0, Once = 1, Twice = 2, Two to five times = 4, Six to ten times = 8, and Ten or more times = 16. The sensitivity analysis was designed to examine whether any skewness in the data towards the high end of the scale, combined with this particular scoring approach, may have resulted in the unusually large effects. Thus, the last category (i.e., "Ten or more times") was recoded and scored as a value of 10 instead of 16.

The results of the second sensitivity analysis are summarized in Table E.2. Interestingly, after recoding the scales, the effect sizes increased for both of these outcomes (from $g = 2.001$ to $g = 2.299$ for the Beliefs about Positive Parenting outcome and from $g = 0.228$ to $g = 0.413$ for the Conflict with Co-Parent outcome). This appears to have occurred because the skewness on these scales affected both the means and the standard deviations. That is, while the means decreased after re-scaling, so did the standard deviations, resulting in higher effect sizes. In conclusion, the re-scaling of these measures did not affect the original conclusions obtained for the follow-up time point. As mentioned previously, these unusually large effects should be interpreted with caution, and we will elaborate on potential reasons for this in the discussion section.

VI. Discussion

The Pathways for Fathers and Families (PFF) impact evaluation study did not find significantly better outcomes for fathers immediately upon the completion of the PFF program. Compared to fathers who received a subset of the program, fathers who engaged in an intensive, eight-week training program did not report significantly more positive parenting beliefs, significantly less conflict with their co-parent, significantly more contact with their child, or significantly improved skills in either conflict resolution or co-parenting. Nonetheless, many of the results were in the expected direction, suggesting that the program could be beneficial.

Significant effects of the program were found when examining the same outcomes six months after its completion. Specifically, compared to fathers who received a subset of the program, fathers who received the full treatment program reported more positive beliefs about healthy co-parenting, better communication skills with their co-parent, better conflict resolution skills, more positive beliefs about healthy parenting, and more engagement with their child when measured six months after the completion of the PFF program.

These significant effects must be interpreted with caution. First, the effect sizes observed were very large and especially so given the rather modest effects typically seen in similar studies. In examining potential sources for these unusual results, the evaluators conducted a rigorous audit of the data collected to ensure correct data entry, used a conservative approach for estimating the models, and conducted several sensitivity analyses. Even so, the data indicated rather large effects, though it is noted that findings for the follow-up time point are out-of-the-norm relative to prior research.

Second, we note that the effect sizes may not necessarily represent an increase in skills, beliefs, or behaviors in treatment group fathers, but instead they might reflect a decrease in skills, beliefs, or behaviors among control group fathers. Between the Post-Survey and Follow-Up Survey time points, while treatment group scores on all outcomes remained relatively stable, control group scores on most outcomes decreased (the group means are reported in Tables V.3 and F.3 for the post-study and follow-up time points, respectively). In addition, seemingly small differences between mean scores can result in very large effects if the corresponding standard deviations are relatively small. In this study, the large effect sizes seen in the additional analyses appear to be largely due to a substantial reduction in the standard deviations. As an example, at the Post-Survey time point (see Table V.3) the standard deviations for the Beliefs about Positive Parenting outcome were 0.273 and 0.306 for the treatment and control groups, respectively. At the six-month follow-up time point (see Table F.3), those same standard deviations (although based on a different analytic sample) decreased to 0.141 and 0.150 for the treatment and control groups, respectively. This reduction in variability between the post-study and follow-up time points translated to larger effect sizes at the follow-up relative to the post-study time point (even if the differences in means remained relatively similar). It is also possible that the obtained variability is (at least in part) due to a somewhat small sample, which has the potential to introduce a larger amount of sampling error. In summary, based on previous research, it is rather

unlikely that effect sizes of this magnitude would be seen if the current study was replicated with different samples.

It may indeed be possible that the additional training received by fathers in the treatment group acted as some sort of buffer, allowing them to retain relatively better outcomes scores while those in the control group naturally regressed as time passed. It is also possible that skills and behaviors (as self-reported) become more consistent over time, thus reducing variability among individuals. However, as this is speculative, the evaluators urge caution in the interpretation and do not recommend drawing high-stake inferences from these results.

Of interest to future researchers may be the significant effects of various demographic variables in predicting outcome scores. Examining the models for the primary and additional analyses, the evaluators noted two variables that were significantly associated with the outcomes (even after controlling for all other covariates): age and education level. The demographic variable of age, while interesting, is likely of less interest to future research as it is a static property of any father receiving training; in other words, although younger fathers were predicted to have better communication skills (at the post-study time point), researchers would likely not want to limit programming to only those under a certain age. Of more interest were the results found for education level. At the six-month follow-up time point, participants who had at least some technical school or collegiate education were predicted to have better conflict resolution skills and more engagement with their child. These results, while extremely preliminary and not the purpose of the study, may suggest that further research into the effect of education on non-educational outcomes (such as engagement with child) may be warranted.

The implementation evaluation revealed a number of lessons learned for future programs. First, although this is a common problem in real-world settings, recruitment of eligible participants was a challenge. Two large changes to study procedures, namely rolling monthly enrollment and the hiring of a recruitment specialist, appeared to have substantially improved participant recruitment. Future evaluations should consider utilizing a recruitment specialist as soon as possible and an implementation design that allows for more frequent enrollment sessions.

Additionally, the implementation evaluation revealed that a few cohorts did not receive the full intervention (i.e., did not receive every intervention session that was part of the curriculum) or did not receive the full intervention in the order intended (i.e., received every intervention session but out of the prescribed order). While some deviation from exact intervention procedures can be expected in real-world settings, these changes may have introduced unintended confounds into the study that cannot be fully evaluated. To avoid such ambiguity, future interventions should adhere to the planned implementation as closely as possible.

Further, the implementation evaluation discovered slightly less than optimal intervention attendance for fathers in the treatment group. A majority of fathers did not attend every training session scheduled, and nearly half of fathers did not attend a sufficient number (75%) of sessions to be considered as having completed the full curriculum. The substantial time commitment required from fathers in the treatment group (i.e., 42 hours of classes over eight weeks) is the most likely reason that so few fathers attended all intervention sessions. While perfect attendance

is extremely unlikely for any intervention, if the program could be delivered in a more efficient way (e.g., delivering all content on a few weekends) with the same level of integrity, this may increase session attendance and program completion.

The study utilized a randomized control trial (RCT) study design. Attendance and other real-world implementation challenges were not considered in the analyses, allowing for increased external validity of the results. In other words, the impact evaluation tested for effects of treatment without accounting for challenges with attendance. Because of this, the results of this study are more likely to generalize to an implementation of the PFF intervention in other, similar, populations.

There are some limitations to the study related to the scales used to measure the outcomes. First, all measures were self-reported and, while self-reports have their advantages, there are also well-known limitations inherent to self-reports. For instance, the Conflict with Co-Parent subscale appeared to have substantial social desirability issues. Many of the items on this subscale asked fathers to report how often they swore at, hit, or otherwise hurt (emotionally, verbally, or physically) their co-parent. Perhaps unsurprisingly, fathers appeared reticent to report any of these behaviors. For example, only 27 of the 482 fathers in the study (less than 6%) reported ever threatening to hit their co-parent prior to intervention (i.e., at the Pre-Survey time point). With these extremely low reported rates at the Pre-Survey time point, there is little room for improvement to be identified on the Post- or Follow-Up Surveys, even if the PFF intervention was effective at reducing conflict with co-parent. Future researchers may consider using scales that ask questions about verbal, emotional, or physical abuse in indirect ways, finding non self-report measures to use, or validating the measures with some other evaluation methods (e.g., in-depth interviews).

Another limitation associated with social desirability bias was the use of a phone-based survey system. The evaluators added the option for fathers to complete the UWM PFF Post- and Follow-Up surveys over the phone approximately halfway through the study out of concern with low-response rates and in recognition of the fact that some fathers may not have been comfortable with using a link to complete a survey online. While only evaluators trained in administering phone surveys conducted these, any phone-based survey can introduce social desirability bias as participants must verbalize their response to another person. Although social desirability bias is certainly a concern, participants were randomly assigned to groups so there is very little reason to suspect that fathers in one group of participants (i.e., fathers in the treatment group or fathers in the control group) systematically chose the phone survey option at a greater or lesser rate than the other. Additionally, based on available information, we have no reason to suspect that there were differences between the groups in terms of survey modes in this study, so any variability introduced by social desirability bias should have affected both treatment and control group participants equally. In future studies it may be of interest to consider survey mode in greater detail and possibly include it in the data analyses.

Some of the subscales might have also limited the results due to possible issues with individual survey items. Specifically, some of the items may have not been applicable based on certain

demographic or environmental characteristics of the participants enrolled in the study. For instance, one item on the Engagement with Child subscale asked fathers how many times in the prior month they had taken their child to a park. If, for instance, a father responding to the survey had an older child who no longer required supervision when going to a park, or even if the survey was completed in the winter (the study was completed in Milwaukee, WI), they may have answered this item differently than a father with a younger child or one who would be answering this item in the summer. In our study, fathers were asked to respond regarding their youngest child, and the distribution of this variable indicates that most fathers had very young children (i.e., at baseline the majority of fathers reported their youngest child's age as under 3 years of age, and only 4 fathers reported their youngest child's age as 18 or over). While it may be useful to incorporate age as a covariate in future investigations, we recommend that researchers carefully consider the items included in subscales to ensure that they are not conditional on demographic or environmental circumstances. Despite some conceptual challenges with individual items, overall subscale reliability indices were strong.

Two of the subscales, Engagement with Child and Conflict with Co-Parent, used frequency response categories (i.e., number of times in the last month) and the scaling was non-linear to best reflect the increasing frequency of engagement. In future research, we suggest that researchers consider other potential ways of assessing or scaling these outcomes, or consider analyses that treat them as ordinal. For all scales, it would also be advisable, when possible, to measure outcomes in other ways that may help to confirm, verify, or complement self-reported measures (e.g., objective observations of behaviors, verification of behaviors from others, in-depth interviews with a random subset of participants). While there is no reason to suspect that these issues affected one group (treatment or control) more than they affected the other given the RCT study design, considering such potential issues (or collecting detailed data to address them) could be useful when designing future studies.

Finally, despite robust orientation and informed consent procedures, the researchers did note that some participants expressed general distrust of the study for multiple reasons. The researchers were told by participants on multiple occasions that they believed their UWM PFF Pre-Survey responses dictated their random group assignment. Despite multiple attempts to explain that group assignment was truly random and unconnected with any surveys, participants continued to occasionally express this belief throughout the study. Additionally, during orientation sessions some participants expressed concerns that data or participant contact information was going to be used against them by Milwaukee County Child Support Services. The researchers made explicit written and verbal statements to prospective participants assuring them that this would not be the case; however, it was apparent that this was not universally accepted. Future studies may consider spending additional time in developing study procedures with an emphasis on gaining the trust of participants. In particular, researchers could consider in-depth explanations about the relationship of governmental agencies to the study itself, or having frank and genuine discussions with prospective participants about the injustices committed by researchers on marginalized communities in the past and the specific procedures that current researchers have put in place to ensure that these injustices are not repeated.

In conclusion, the results found in this study were mixed. The effect of treatment was not significant for any of the outcomes immediately following the completion of the program. Significant results were found at the six-month follow-up time point, but these need to be regarded with caution as discussed above. Preliminary findings indicating that education level may have an effect on various fatherhood skills such as child engagement and conflict resolution provide a basis for further research. Finally, many of the results were in the expected direction, indicating that the intervention shows some promise and further research is warranted. It may be that the PFF program helps non-custodial fathers to interact more positively with both their children and their co-parents; if so, the impacts on fathers' and families' mental health, emotional well-being, and life satisfaction could be very beneficial.

VII. References

Avellar, S., Covington, R., Moore, Q., Patnaik, A., & Wu, A. Effects of four responsible fatherhood programs for low-income fathers: Evidence from the parents and children together evaluation (No. d1b1f6c40fd24df5b03102d8e15a3064). Mathematica Policy Research.

Campbell, C. A., Howard, D., Rayford, B. S., & Gordon, D. M. (2015). Fathers matter: Involving and engaging fathers in the child welfare system process. *Children and Youth Services Review*, 53, 84-91.

Del Boca, D., & Flinn, C. J. (1995). Rationalizing child-support decisions. *The American Economic Review*, 1241-1262.

Dyer, J., Fagan, J., Kaufman, R., Pearson, J., & Cabrera, N. J. (2015). Fatherhood research and practice network coparenting relationship scale. *Philadelphia, PA: Fatherhood Research and Practice Network*.

Ferguson, S., & Morley, P. (2011). Improving engagement in the role of father for homeless, noncustodial fathers: A program evaluation. *Journal of Poverty*, 15, 206-225.

Finzi-Dottan, R., & Cohen, O. (2019). Involvement and acceptance of custodial fathers: The role of narcissism and caregiving. *Psychology of Men & Masculinities*, 20, 1-11.

Gordon, D. M., Hunter, B., Woods, L. N., Tinney, B., Bostic, B., Malone, S., ... & Smith, A. (2012). Increasing outreach, connection, and services to low-income non-custodial fathers: How did we get here and what do we know. *Fathering*, 10, 101.

Hamer, J. F. (1998). What African-American noncustodial fathers say inhibits and enhances their involvement with children. *Western Journal of Black Studies*, 22, 117.

Mathematica. (n.d.). *Parents and children together (PACT)*. https://www.mathematica.org/our-publications-and-findings/projects/parents-and-children-together

McKenry, P. C., Price, S. J., Fine, M. A., & Serovich, J. (1992). Predictors of single, noncustodial fathers' physical involvement with their children. *The Journal of Genetic Psychology*, 153, 305-319.

Meyer, D. R., Kim, Y., & Cancian, M. (2019). Satisfaction with child support agency services and its relationship to child support payments. Child Support Noncustodial Parent Employment Demonstration (grant number 90FD0184). Washington, DC: Office of Child Support Enforcement within the Administration for Children and Families, U.S. Department of Health and Human Services.

Self-Brown, S., Osborne, M. C., Lai, B. S., Brown, N. D. V., Glasheen, T. L., & Adams, M. C. (2017). Initial findings from a feasibility trial examining the SafeCare Dad to Kids Program with marginalized fathers. *Journal of family violence*, 32, 751-766.

Sorensen, E. (1997). A national profile of nonresident fathers and their ability to pay child support. *Journal of Marriage and the Family*, 785-797.

VIII. Appendices

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

Table A.1. Logic model (or theory of change) for the program

Goals	Activities	Outputs/Implem. Obj's	Short-term Outcomes	Long-Term Outcomes
1. Strengthen positive father-child engagement 2. Develop healthy relationships & marriage 3. = Improve fathers' economic stability and mobility **Target Populations** • 750 low-income non-custodial fathers ages 18 and older with children ages 0 to 21 **Input** • **MFH:** Experienced Fatherhood agency • **Evidence-based curricula:** parenting & healthy relationship • **Workforce Dev. Coordinator:** Employ Milwaukee • **Legal Serv. Partner:** CL • **Trauma Services:** Alma • **Local Evaluator:** • AMTC & UWM • **Budget:** $1.55 million (evaluation) 5 years	**Curricula** • Fatherhood Development • Within My Reach • Children First • Child Support 101 • Family Law 101 **Employment** • Employ Milwaukee: Workforce Development/ Preparation **Case Management** • Indiv. Service Plan & Coaching • Comp. Support Services Network **Specialty Services** • Alt Dispute Resolution • Trauma Services	**1:** 750 enrolled, randomly assigned as treatment (T) or control (C) **2:** 750 T&C fathers get RF needs assessment **3:** 375 T fathers get Resp. Parenting curriculum **4:** 40 T fathers get Alternative Dispute Resolution **5**: 375 T fathers get Healthy Relationship curriculum **6:** 750 T&C fathers get Career Pathway Consult. **7:** 750 T&C fathers get workforce services **8**: 750 T & C fathers offered Child Support 101 **9:** 50% of Ts attending CS 101 get CS services **10:** 375 T fathers get Family Law 101 **11:** 40 T fathers get legal support services. **12:** 100% in need of support services receive them	**1. Improve rel-ship skills** • > knowledge re: communication • > knowledge re: conflict resolution • < conflict w/co-parent **2. Improve parenting, co-parenting skills** • > knowledge re: pos. parenting • > positive/nurturing parenting **3. Increased father/ child engagement** • > freq of time w/ child • > involv. of father in specific activities w/child • > rel quality w/child **4. Increase in fathers' financial responsibility** • > frequency of making child support payments • > amount contributed to child support **5. Increase econ. stability** • > employment status • > income or access to financial resources	**1.** Improved marital/ couple satisfaction **2.** Improved co-parenting satisfaction **3.** Improved adult and child well-being **4.** Increased economic stability and mobility; **5.** Reduced poverty **6.** Reduced recidivism (for crim justice pop) **Assumptions** 1. Lack of skills underlies relationship conflict 2. Lack of skills underlies ineffective parenting 3. Fathers want to be involved with their children 4. Parental conflict undermines effective parenting 5. Lack of knowledge contributes to poor child support compliance 6. Matching of worker job training to employer needs will increase employment for *fathers*

B. DATA AND STUDY SAMPLE

Table B.1. Example of Program Implementation Plan for Group A (intervention).

Partner: MFHI
Site: MFHI
Educator: _____
Session Series Name: MFH-MFH-YR3-MMDDYYYY
Start Date: Month Day, Year
End Date: Month Day, Year
Days of the Week: M,T,W
Start and End Time: 10:00am-12:00pm & 5:30pm-7:30pm

Key: FD = Fatherhood Development Responsible Parenting Curriculum; 24/7 = 24/7 Dad A.M. Responsible Parenting Curriculum; ES = Economic Stability; SS = Support Services; CS 101 = Child Support 101; WMR = Within My Reach Healthy Marriage curriculum

Date	Topic(s) Covered	Lesson/ Module	Length (mins.)
Week One (Month-Date Range)			**390 min**
Monday	(SS) Pre-Survey	N/A	60 min
	Case Management/Community Resources	N/A	90 min
Tuesday	(FD) What is a Man?	Session III pg.17	40 min
	(FD) Your Goals and Values	Session II pg. 5	40 min
	(FD) Values Voting	Session II pg. 9	40 min
Wednesday	(24/7) Entire Discipline Section	Pg. 148	120 min
Week Two (Month-Date Range)			**360 min**
Monday	(FD) Understanding Stress	Session VIII pg. 5	35 min
	(FD) How to Cope with Stress/Stress Strategies	Session VIII pg.12	45 min
	(FD) What's Communication	Session VI pg. 5	20 min
	(FD) One/Two Way Communication	Session VI pg. 9	20 min
Tuesday	(ES) About Credit/Credit Cards-Albert H.	Financial	60 min
	(ES) Cars & Loans-Albert H.	Financial	60 min
Wednesday	(24/7) Session 10 Entire Module Working with Mom & Co-Parenting	pg. 198	60 min
	(24/7) Dads and work	pg. 218	60 min
Week Three (Month-Date Range)			**360 min**
Monday	(FD) Alcohol and Drugs Facts and Feelings	FD Session XXII pg.5	30 minb
	(FD) Recent Decisions	FD Session VII pg.5	60 min
	(FD) Tests for Life	FD Session XXI pg.19	30 min
Tuesday	(ES) Consumer awareness/Saving and investing-Albert H.	Financial	60 min
	(ES) In Trouble-Albert H.	Financial	60 min
Wednesday	(24/7) Activity 5.2 Ways to Communicate "Fight or Flight"	Pg. 112	30 min
	(24/7) Activity 3.2 Showing and Handling Feelings	Pg. 78	30 min

Date	Topic(s) Covered	Lesson/ Module	Length (mins.)
	(FD) Maintaining Your Cool	Session XVIII pg. 21	60 min
Week Four (Month-Date Range)			**240 min**
Monday	(FD) Being a "Real Man" Hazardous to Your Health	FD Session XXII pg.5	60 min
	(FD) Children's Growth	24/7 Session 8pg. 168	60 min
Tuesday	(FD) What's Your Influence (Myth or Fact)	Session XIII pg. 5	15 min
	(FD) Creative Caring	Session XV pg. 18	25 min
	(FD) Building Self Worth (Trust Walk)	Session XV pg. 5	40 min
	(FD) Father Knows Best Post-Survey for A2	Session XIII pg. 15	40 min
Wednesday	OFF		
Week Five (Month-Date Range)			**360 min**
Monday	Off for current cohort, orientation and Pre-Survey A4		60 min
	(WMR) The State of Relationships Today		60 min
Tuesday	(WMR)Healthy Relationships: What They Are and What They Aren't	Unit 1	30 min
	(WMR) Sliding vs. Deciding	Unite 2	30 min
	(WMR) Smart Love	Unit 3	30 min
Wednesday	Employ Milwaukee	Unit 4	30 min
	(ES) The Art of Budgeting/Living on Your Own	Employ Milw.Financial	60 min
Week Six (Month-Date Range)			**360 min**
Monday	(WMR) Knowing Yourself First	Unit 5	60 min
	(WMR) Making Your own Decisions	Unit 6	60 min
Tuesday	(WMR) Dangerous Patterns in Relationships	Unit 7	60 min
	(WMR) Where Conflict Begins	Unit 8	60 min
Wednesday	(CS 101) Child Support	Child Support	60 min
	Centro Legal	Centro Legal	60 min

Table B.2. Example of Program Implementation Plan for Group B (Control).

Partner: MFHI
Site: MFHI
Educator: _____
Session Series Name: MFH-MFH-YR3-03082018
Start Date: Month Day, Year
End Date: Month Day, Year
Days of the Week: Thursday
Start and End Time: 1:00pm – 3:00pm & 5:30pm-7:30pm

Key: ES = Economic Stability; SS = Support Services; CS 101= Child Support 101

Date	Topic(s) Covered	Lesson/ Module	Length (mins.)
Week One (Month-Date Range)			**150 min**
Monday	(SS) Orientation, Pre-Survey	N/A	60 min
	Case Management/Community Resources	N/A	90 min
Thursday	(ES) The Art of Budgeting/Living on your own–Albert H.	Financial	60 min
Week Two (Month-Date Range)			**120 min**
Thursday	(ES) About Credit/Credit Cards	Financial	60 min
	(ES) Cars & Loans	Financial	60 min
Week Three (Month-Date Range)			**120 min**
Thursday	Employ Milwaukee	Employ	60 min
	(CS101) Child Support 101	CSS	60 min
Week Four (Month-Date Range)			**120 min**
Thursday	(ES) Consumer Awareness/Saving and investing	Financial	60 min
	(ES) In Trouble	Financial	60 min
Week Five (Month-Date Range)			**60 min**
Thursday	Exit survey		

Table B.3. Data for addressing implementation evaluation research questions

Implementation element	Research question	Data source	Timing/frequency of data collection	Party responsible for data collection
Fidelity	Were all intended intervention components offered and for the expected duration?	Workshop sessions entered into nFORM and AMTC findings based on program implementation plan	All sessions delivered	Intervention staff and AMTC and Associates
Staffing	Who delivered services to fathers?	Hiring records and training logs and program implementation plan	Start of each workshop	Management intervention staff and AMTC and Associates
Dosage	What was the completion rate for fathers?	Workshop attendance in nFORM as well as sign-in sheets and attendance quality assurance logs	After each workshop was completed, attendance quality assurance logs reviewed quarterly	Intervention staff and AMTC and Associates

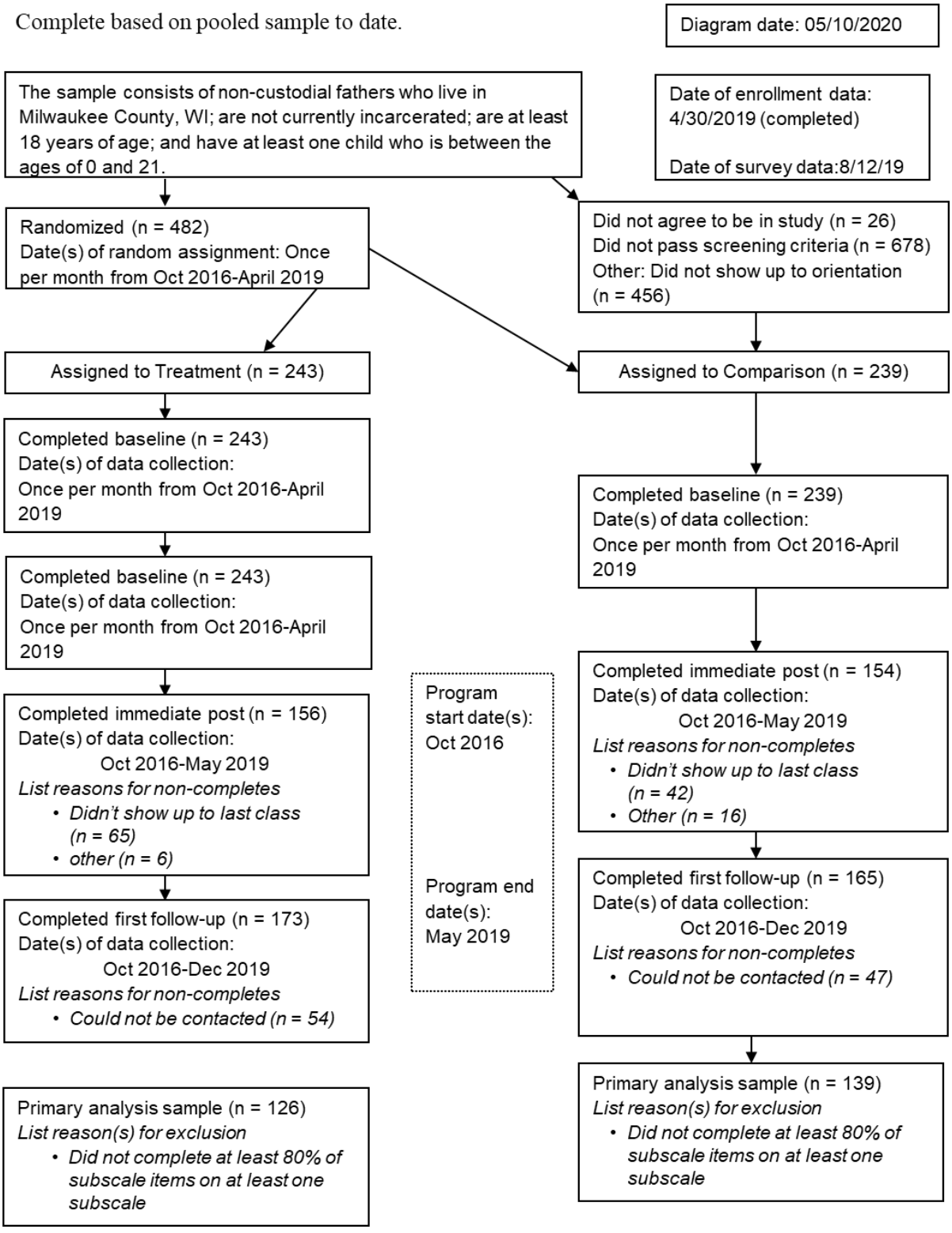
Table B.4. Session Schedule by Educator

Educator	Monday	Tuesday	Wednesday	Thursday	Friday
Educator 1 (Teaches Treatment PM session)	X	X	X	–	–
Educator 2 (Teaches Treatment AM session)	X	X	X	–	As needed for makeups
Educator 3 (Teaches Counterfactual PM session)	–	–	–	X	As needed for makeups

Table B.5. Key features of the data collection

Intervention/ counterfactual	Data source	Timing of data collection	Mode of data collection	Party responsible for data collection	Start and end date of data collection
Intervention	Intervention group study participants	Enrollment (baseline)	In-person online surveys (nFORM & Qualtrics)	Program staff, UWM & AMTC	October 2016-December 2019
Intervention	Intervention group study participants	End of intervention (2 months after enrollment)	In-person, phone, or emailed online surveys (nFORM & Qualtrics)	Program staff, UWM & AMTC	October 2016-December 2019
Intervention	Intervention group study participants	Follow-up (6 months after intervention ends)	In-person, phone, or emailed online surveys (Qualtrics)	Program staff, UWM & AMTC	October 2016-December 2019
Counterfactual	Comparison group study participants	Enrollment (baseline)	In-person online surveys (nFORM & Qualtrics)	Program staff, UWM & AMTC	October 2016-December 2019
Counterfactual	Comparison group study participants	End of intervention (1 month after enrollment)	In-person, phone, or emailed online surveys (nFORM & Qualtrics)	Program staff, UWM & AMTC	October 2016-December 2019
Counterfactual	Comparison group study participants	Follow-up (6 months after intervention ends)	In-person, phone, or emailed online surveys (Qualtrics)	Program staff, UWM & AMTC	October 2016-December 2019

Table B.6. CONSORT diagram for individual clients, for studies in which consent occurred before assignment



C. Data preparation

Data were merged from multiple sources using the statistical software SAS 9.4. Data sources include web-based Qualtrics surveys and nFORM surveys. After merging all participant data by ID, the data were examined using descriptive statistics (e.g., variable distributions, frequencies, ranges) for unusual observations that may signal data entry mistakes. Raw data were exported from the Qualtrics online survey software into an Excel document. Any individuals who were given incorrect IDs (i.e., their ID number did not match the format used by the study and was the result of human error in entry) for either their UWM PFF Post-Survey or UWM PFF Follow-Up Survey were identified. Those individuals' IDs were fixed by the researchers when possible (for example, when the resulting incorrect ID was only missing a single digit and could be easily matched to a known ID). After these corrections, 18 surveys could not be matched with the participant who completed them as the ID numbers entered were too dissimilar from anything recorded. These surveys were removed from the larger data set.

Reliability was calculated for all subscales of the researcher-created survey. Item-level reliability analyses did not suggest that any individual survey items be removed. Reliability of all subscales was over 0.6. Missing data (i.e., outcome measures that were missing or insufficiently complete, using the threshold of completing less than 80% of items) was handled by creating individual analytic samples for the primary and secondary research questions.

Participant surveys were examined for consistency based on ID to ensure that Pre-, Post-, and Follow-Up surveys reflect the same participant. A number of individuals contributed multiple UWM PFF surveys at the post-study and follow-up time points, and it was determined that this represented individuals who completed a survey on each of their children, rather than just their youngest child as was intended, due to human error. Post- and Follow-Up surveys that were completed on a child that was not the participant's youngest (as identified in the UWM PFF Pre-Survey) were deleted.

D. Impact estimation

Model specifications for the analyses can be seen below.

1. Baseline equivalence

Hedge's g:

$$g = \frac{\omega(y_i - y_c)}{\sqrt{\frac{(n_i - 1)s_i^2 + (n_c - 1)s_c^2}{n_i + n_c - 2}}}$$

where y_i is the adjusted (or unadjusted) mean for the intervention group, y_c is the adjusted (or unadjusted) mean for the comparison group, n_i is the sample size of the intervention group, n_c is the sample size of the comparison group, s_i is the unadjusted standard deviation for the intervention group, and s_c is the unadjusted standard deviation for the comparison group. Omega (ω) is the small sample size correction:

ω = [1 – 3/(4N – 9)],

with N being the total sample size.

Cox's Index:

$$d_{Cox} = \frac{\omega\left[\ln\left(\frac{p_i}{1-p_i}\right) - \ln\left(\frac{p_c}{1-p_c}\right)\right]}{1.65}$$

where p_i is the prevalence rate of an outcome observed in the intervention group, p_c is the prevalence rate of an outcome observed in the comparison group, and Omega (ω) is the small sample size correction.

2. Program impacts

Primary research questions multiple linear regression model:

$$\hat{y}_j = b_0 + b_1 \times group + b_2 \times Presurvey_j + b_3 \times x_3 + \ldots + b_k \times x_k$$

where j = the outcome subscale (j = 1, 2, …, 6), $\hat{y}_j$ is the (model predicted or adjusted) Post-Survey subscale score, *group* (1 or 0) indicates group membership, $Presurvey_j$ is the subscale score (for the outcome variable) at baseline, and x_3 to x_k are the remaining baseline covariates.

Secondary research questions multiple linear regression model

$$\hat{y}_j = b_0 + b_1 \times group + b_2 \times Presurvey_j + b_3 \times x_3 + \ldots + b_k \times x_k$$

where j = the outcome subscale (j = 1, 2, …, 7), $\hat{y}_j$ is the (model predicted or adjusted) Follow-Up (outcomes 1-6) or Post-Survey (outcome 7) subscale score, *group* (1 or 0) indicates group membership, $Presurvey_j$ is the subscale score (for the outcome variable) at baseline, and x_3 to x_k are the remaining baseline covariates.

Table D.1. Regression results using data from UWM Post-Survey to address the primary research questions

Outcome measure	Predictor	B	95% CI	β	SE	t	p
Healthy Co-Parenting	Intercept	1.101	[0.120, 2.082]				
	Group	0.060	[-0.126, 0.245]	.029	0.094	0.63	.527
	Race	0.385	[0.024, 0.747]	.100	0.183	2.10	0.37
	Age	0.078	[-0.299, 0.142]	-.034	0.112	0.70	.486
	Employment	0.031	[-0.197, 0.260]	.015	0.116	0.27	.786
	Income	-0.064	[-0.291, 0.163]	-.031	0.115	0.55	.581
	Education	0.024	[-0.177, 0.224]	.011	0.102	0.23	.817
	Marital status	-0.095	[-0.300, 0.110]	-.045	0.104	0.91	.363
	Healthy Co-Parenting Pre-Survey score	0.586	[0.489, 0.682]	.622	0.049	11.92	<.001
	Conflict Resolution Skills Pre-Survey score	0.106	[0.007, 0.205]	.120	0.050	2.12	.035
	Positive Parenting Beliefs Pre-Survey score	-0.038	[-0.293, 0.217]	-.015	0.129	0.29	.769
	Conflict with Co-Parent Pre-Survey score	-0.019	[-0.076, 0.037]	-.033	0.029	0.68	.496
Co-Parent Communication	Intercept	0.680	[-0.510, 1.870]				
	Group	0.083	[-0.139, 0.304]	.035	0.112	0.74	.463
	Race	0.312	[-0.119, 0.744]	.071	0.219	1.43	.155
	Age	-0.316	[-0.579, -0.051]	-.119	0.134	2.36	.019
	Employment	0.036	[-0.236, 0.309]	.015	0.138	0.26	.794
	Income	-0.104	[-0.375, 0.167]	-.044	0.138	0.75	0.452
	Education	0.210	[-0.030, 0.449]	.086	0.122	1.73	.086
	Marital status	-0.152	[-0.398, 0.094]	-.063	0.125	1.21	.226
	Healthy Co-Parenting Pre-Survey score	0.118	[-0.035, 0.271]	.109	0.077	1.52	.129
	Communication with Co-Parent Pre-Survey score	0.294	[0.133, 0.545]	.333	0.082	3.60	<.001
	Conflict Resolution Skills Pre-Survey score	0.240	[0.091, 0.389]	.235	0.076	3.18	<.005
	Positive Parenting Beliefs Pre-Survey score	0.088	[-0.217, 0.392]	.031	0.155	0.57	.570
	Conflict with Co-Parent Pre-Survey score	-0.005	[-0.072, 0.062]	-.007	0.034	0.14	.891

Outcome measure	Predictor	B	95% CI	β	SE	t	p
Conflict Resolution Skills	Intercept	0.512	[-0.592, 1.616]				
	Group	0.177	[-0.031, 0.386]	.083	0.106	1.67	.096
	Race	0.147	[-0.260, 0.553]	.036	0.206	0.71	.478
	Age	-0.106	[-0.354, 0.143]	-.043	0.126	0.84	.403
	Employment	0.105	[-0.152, 0.362]	.048	0.130	0.80	.423
	Income	-0.046	[-0.302, 0.209]	-.021	0.130	0.36	.721
	Education	0.312	[0.086, 0.537]	.138	0.115	2.72	.007
	Marital Status	-0.159	[-0.390, 0.209]	-.072	0.117	1.36	.176
	Healthy Co-Parenting Pre-Survey score	0.089	[-0.020, 0.198]	.090	0.055	1.61	.108
	Conflict Resolution Skills Pre-Survey score	0.459	[0.348, 0.570]	.490	0.056	8.13	<.001
	Positive Parenting Beliefs Pre-Survey score	0.240	[-0.047, 0.527]	.092	0.146	1.65	.100
	Conflict with Co-Parent pre-survey score	-0.017	[-0.065, 0.062]	-.003	0.032	0.05	.958
Positive Parenting Beliefs	Intercept	1.313	[0.710, 1.912]				
	Group	-0.031	[-0.145, 0.083]	-.028	0.058	0.53	.596
	Race	0.049	[-0.173, 0.271]	.024	0.113	0.44	.664
	Age	-0.044	[-0.180, 0.092]	-.036	0.069	0.64	.524
	Employment	-0.026	[-0.167, 0.114]	-.024	0.072	0.37	.711
	Income	0.029	[-0.111, 0.168]	.026	0.071	0.40	.688
	Education	0.051	[-0.072, 0.174]	.045	0.063	0.82	.415
	Marital status	-0.038	[-0.164, 0.088]	-.034	0.064	0.59	.557
	Healthy Co-Parenting Pre-Survey score	-0.008	[-0.068, 0.051]	-.017	0.030	0.28	.782
	Conflict Resolution Skills	0.093	[0.032, 0.154]	.197	0.031	3.02	.003
	Positive Parenting Beliefs Pre-Survey score	0.522	[0.366, 0.679]	.398	0.079	6.57	<.001
	Conflict with Co-Parent Pre-Survey score	-0.008	[-0.043, 0.026]	-.026	0.018	0.46	.644
Engagement with Child	Intercept	2.864	[-3.048, 8.775]				
	Group	-0.443	[-1.537, 0.651]	-.042	0.556	0.80	.426
	Race	1.640	[-0.489, 3.769]	.082	1.081	1.52	.131
	Age	0.285	[-1.028, 1.598]	.024	0.667	0.43	.670
	Employment	-0.632	[-1.980, 0.716]	-.059	0.684	0.92	.357
	Income	0.169	[-1.171, 1.508]	.016	0.604	0.38	.706

Outcome measure	Predictor	B	95% CI	β	SE	t	p
	Education	0.228	[-0.961, 1.417]	.021	0.604	0.38	.706
	Marital status	-0.091	[-1.301, 1.120]	-.008	0.615	0.15	.883
	Healthy Co-Parenting Pre-Survey score	0.486	[-0.156, 1.128]	.100	0.326	1.149	.137
	Conflict Resolution Skills Pre-Survey score	0.731	[0.147, 1.315]	.159	0.297	2.47	.014
	Positive Parenting Pre-Survey score	-1.005	[-2.567, 0.558]	-.078	0.793	1.27	.207
	Engagement with Child Pre-Survey score	0.443	[0.321, 1.315]	.463	0.062	7.17	<.001
	Conflict with Co-Parent Pre-Survey score	-0.069	[-0.404, 0.266]	-.023	0.170	0.41	.685
Conflict with Co-Parent	Intercept	1.703	[0.148, 3.259]				
	Group	-0.100	[-0.395, 0.194]	-.039	0.149	0.67	.502
	Race	0.165	[-0.408, 0.738]	.034	0.291	0.57	.571
	Age	-0.231	[-0.581, 0.119]	-.079	1.778	1.30	.195
	Employment	-0.328	[-0.690, 0.035]	-.125	0.184	1.78	.076
	Income	0.153	[-0.207, 0.513]	.059	0.183	0.84	.403
	Education	-0.180	[-0.498, 0.138]	-.067	0.162	1.11	.266
	Marital status	-0.068	[-0.394, 0.257]	-.026	0.165	0.41	.679
	Healthy Co-Parenting Pre-Survey score	0.068	[-0.086, 0.221]	.057	0.078	0.87	.384
	Conflict Resolution Skills Pre-Survey score	-0.184	[-0.341, -0.027]	-.163	0.080	2.31	.022
	Positive Parenting Pre-Survey score	-0.226	[-0.630, 0.178]	-.072	0.205	1.10	.271
	Conflict with Co-Parent Pre-Survey score	0.218	[0.129, 0.307]	.291	0.045	4.82	<.001

Notes: Categorical variables were coded as follows: For Group, Control group = 0, Treatment group = 1; for Race, Other = 0, Black = 1; for Age, 44 years and younger = 0, 45 years and older = 1; for Employment, Not Employed = 0, Employed = 1; for Income, Less than $500 per month = 0, $500 a month or more = 1; for Marital Status, Never Married = 0, Married = 1.
Higher scores represent better outcomes, except for Outcome 6 where lower scores represent better outcomes. Results are based on the first analytic sample

E. Sensitivity analyses and alternative model specifications

Table E.1. Differences in means between intervention and comparison groups estimated using alternative methods

Outcome	Benchmark approach	Group x pre-score interaction [interaction effect]
Primary Research Questions		
1. Healthy Co-Parenting	0.060 (.527)	0.063 [-0.18] (.042)
2. Communication with Co-Parent	0.083 (.463)	0.089 [-0.18] (.040)
3. Conflict Resolution Skills	0.177 (.096)	0.178 [-0.10] (.269)
4. Beliefs about Positive Parenting	-0.031 (.596)	-0.031 [-0.01] (.945)
5. Engagement with Child	-0.443 (.426)	-0.442 [-.003] (.972)
6. Conflict with Co-Parent	-0.100 (.502)	-0.104 [0.28] (.002)
Additional Research Questions		
1. Healthy Co-Parenting	0.476 (<.001)	0.477 [-0.11] (.271)
2. Communication with Co-Parent	0.360 (.001)	0.361 [-0.06] (.485)
3. Conflict Resolution Skills	0.456 (<.001)	0.455 [-0.08] (.394)
4. Beliefs about Positive Parenting	0.181 (<.005)	0.178 [0.27] (.052)
5. Engagement with Child	2.302 (<.001)	2.300 [-0.09] (.435)
6. Conflict with Co-Parent	-0.057 (.547)	-0.058 [0.13] (.019)
7. Economic Stability	-0.029 (.598)	-0.030 [0.08] (.513)

Source: UWM PFF Post- and Follow-Up surveys, administered immediately following and six months after the completion of intervention

Notes: All values refer to the model-adjusted group differences, and values in brackets (last column) refer to the interaction effect.

The p-values (in parentheses) refer to the test of the Group effect, except in the last column, where they refer to a test of the interaction effect.

Figure E.1. Interaction plots: Post-study group effect on the Healthy Co-Parenting (left) and Conflict with Co-Parent (right) scores as a function of the corresponding pre-study scores.

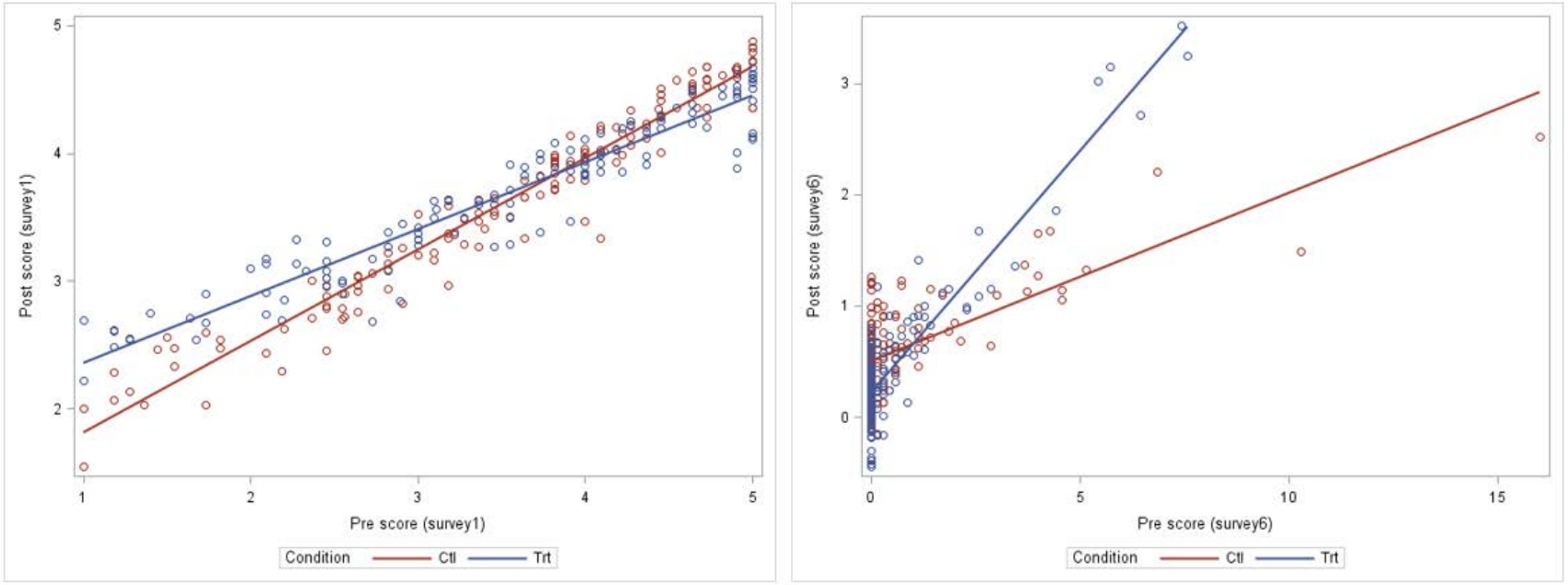


Table E.2. Differences in means and effect sizes between intervention and comparison groups estimated using alternative coding structure for two scales (outcomes 5 and 6) at the follow-up time point

	Benchmark approach				Alternative coding structure			
Outcome measure	Intervention mean (SD)	Comparison mean (SD)	Mean difference (p-value)	Effect size	Intervention mean (SD)	Comparison mean (SD)	Mean difference (p-value)	Effect size
Engagement with Child	9.219 (1.084)	6.917 (1.209)	2.302 (<.001)	2.001	6.842 (0.628)	5.322 (0.691)	1.520 (<.001)	2.299
Conflict with Co-Parent	0.254 (0.230)	0.311 (0.268)	-0.057 (.547)	0.228	0.217 (0.169)	0.291 (0.185)	-0.074 (.328)	0.413
Sample Size	152	149						

Note: Alternative coding structure refers to the Likert response values. Original coding structure for both outcomes was: Never = 0, Once = 1, Twice = 2, Two to five times = 4, Six to ten times = 8, and Ten or more times = 16. Alternative coding structure for both outcomes was: Never = 0, Once = 1, Twice = 2, Two to five times = 4, Six to ten times = 8, and Ten or more times = 10.

F. Additional analyses

This section includes information on the additional impact analyses.

Table F.1. Outcome measures used for the additional impact analyses research questions

Outcome measure	Description of outcome measure	Timing of measure
Outcome Measure 1: Healthy Co-Parenting	The outcome measure is the 11-item Co-parenting Relationship Scale adapted from Dyer, Fagan, Kaufman, Pearson, & Cabrera (2015). All items are on a five-point Likert scale ranging from “Strongly disagree” to “Strongly agree”. We reverse-coded items 1, 2, 3, 9, 10, and 11. Mean outcome measure scores were derived by adding all completed items (at least 9 of 11 to meet the 80% analytic sample threshold) together then dividing by the number of items completed. Higher scores are considered to indicate better co-parenting skills. Cronbach’s α: .966	Six months after the completion of the intervention
Outcome Measure 2: Communication with Co-Parent	The outcome measure is the 11-item Communication with Co-parent Scale created by the local evaluators. All items are on a five-point Likert scale ranging from “Strongly disagree” to “Strongly agree”. Mean outcome measure scores were derived by adding all completed items (at least 9 of 11 to meet the 80% analytic sample threshold) together then dividing by the number of items completed. Responses were coded such that higher scores are considered to indicate better communication skills Cronbach’s α: .971	Six months after the completion of the intervention
Outcome Measure 3: Conflict Resolution Skills	The outcome measure is the 8-item Resolving Conflict with Co-parent Scale created by the local evaluators. All items are on a five-point Likert scale ranging from “Not at all confident” to “Completely confident”. Mean outcome measure scores were derived by adding all completed items (at least 7 of 8 to meet the 80% analytic sample threshold) together then dividing by the number of items completed. Responses were coded such that higher scores are considered to indicate better conflict resolution skills. Cronbach’s α: .969	Six months after the completion of the intervention
Outcome Measure 4: Positive Parenting Skills	The outcome measure is the 10-item Beliefs about Positive Parenting Scale created by the local evaluators. All items are on a four-point Likert scale ranging from “Not at all important” to “Very important”. Mean outcome measure scores were derived by adding all completed items (at least 8 of 10 to meet the 80% analytic sample threshold) together then dividing by the number of items completed. Responses were coded such that higher scores are considered to indicate more positive parenting skills. Cronbach’s α: .931	Six months after the completion of the intervention

Outcome measure	Description of outcome measure	Timing of measure
Outcome Measure 5: Engagement with Child	The outcome measure is the 11-item Engagement with Child Scale created by the local evaluators. All items are on a six-point scale ranging from “Never” to “More than 10 times”. Responses were coded exponentially to capture the increasing frequency of behavior measured by the scale (i.e., Never = 0, Once = 1, Twice = 2, Two to five times = 4, Six to ten times = 8, and Ten or more times = 16). Mean outcome measure scores were derived by adding all completed items (at least 9 of 11 to meet the 80% analytic sample threshold) together then dividing by the number of items completed. Responses were coded such that higher scores are considered to indicate higher levels of engagement. Cronbach’s α: .978	Six months after the completion of the intervention
Outcome Measure 6: Conflict with Co-Parent	The outcome measure is the 7-item Conflict with Custodial Parent Scale created by the local evaluators. All items are on a six-point Likert scale ranging from “Never” to “More than 10 times”. Responses were coded exponentially to capture the increasing frequency of behavior measured by the scale (i.e., Never = 0, Once = 1, Twice = 2, Two to five times = 4, Six to ten times = 8, and Ten or more times = 16). Mean outcome measure scores were derived by adding all completed items (at least 6 of 7 to meet the 80% analytic sample threshold) together then dividing by the number of items completed. Responses were coded such that higher scores are considered to indicate higher levels of conflict (less desirable outcome). Cronbach’s α: .779	Six months after the completion of the intervention
Outcome Measure 7: Economic Stability	The outcome measure is the 10-item Attitudes Towards Economic Stability subscale taken directly from the nFORM Survey. All items are on a four-point Likert scale ranging from “Strongly agree” to “Strongly disagree”, with a “N/A” option. Mean outcome measure scores were derived by adding all completed items (at least 8 of 10 to meet the 80% analytic sample threshold) together then dividing by the number of items completed. Responses were coded such that higher scores are considered to indicate worse economic skills (less desirable outcome). Cronbach’s α: .918	Immediately following the completion of the intervention

Table F.2. Summary statistics of key baseline measures and baseline equivalence across study groups, for individuals completing the UWM Follow-Up Survey

Baseline measure	Intervention mean (standard deviation)	Comparison mean (standard deviation)	Intervention versus comparison mean difference (p-value of difference)	Effect size
Race/ethnicity (%)				
Black	91.4 (0.28)	91.3 (0.28)	0.1 (.958)	.008
Other	8.6	8.7		
Income level (%)				
Less than $500/month	45.4 (0.50)	40.9 (0.49)	4.5 (.435)	.111
$500/month or more	54.6	59.1		
Age (%)				
44 years and younger	74.3 (0.44)	73.2 (0.44)	1.1 (.815)	.034
45 years and older	25.7	26.8		
Education (%)				
High school diploma or less	69.1 (0.46)	62.4 (0.48)	6.7 (.223)	.180
Some college or more	30.9	37.6		
Employment (%)				
Employed (including occasional work)	55.9 (0.50)	61.1 (0.49)	5.2 (.364)	.130
Not currently employed	44.1	38.9		
Marital Status (%)				
Married (currently or previously)	38.2 (0.49)	38.3 (0.49)	0.1 (.986)	.003
Never married	61.8	61.7		
1: Co-Parenting	3.615 (1.10)	3.460 (1.07)	0.155 (.216)	.142
2: Communication with Co-Parent	3.497 (1.31)	3.337 (1.34)	0.160 (.295)	.121
3: Conflict Resolution Skills	3.583 (1.19)	3.502 (1.18)	0.081 (.553)	.068
4: Positive Parenting Skills	3.705 (0.38)	3.682 (0.42)	0.023 (.612)	.056
5: Engagement with Child	10.703 (5.26)	9.872 (5.65)	0.831 (.188)	.152
6: Conflict with Co-Parent	0.665 (1.46)	0.799 (2.23)	0.134 (.539)	.071
7: Economic Stability	1.652 (0.49)	1.677 (0.44)	0.025 (.741)	.054

Baseline measure	Intervention mean (standard deviation)	Comparison mean (standard deviation)	Intervention versus comparison mean difference (p-value of difference)	Effect size
Sample size	152	149		

Notes: p-values are included in parentheses. Results are for the second analytic sample. Demographic variables were dichotomized from original categorical variables. Effect sizes are calculated using Cox's Index for dichotomous demographic variables and Hedge's g for continuous subscale variables.

Table F.3. Post-intervention estimated effects using data from UWM Follow-Up and Post-Survey to address the secondary research questions

Outcome measure	Intervention mean or % (standard deviation)	Comparison mean or % (standard deviation)	Intervention compared to comparison mean difference (p-value of difference)	Effect size
Healthy Co-Parenting (Outcome 1)	3.935 (0.277)	3.459 (0.295)	0.476 (<.001)	1.660
Communication with Co-Parent (Outcome 2)	3.713 (0.287)	3.354 (0.318)	0.360 (<.001)	1.180
Conflict Resolution Skills (Outcome 3)	3.759 (0.294)	3.304 (0.330)	0.456 (<.001)	1.455
Beliefs about Positive Parenting (Outcome 4)	3.647 (0.141)	3.466 (0.150)	0.181 (<.005)	1.245
Engagement with Child (Outcome 5)	9.219 (1.084)	6.917 (1.209)	2.302 (<.001)	2.001
Conflict with Co-Parent (Outcome 6)*	0.254 (0.230)	0.311 (0.268)	-0.057 (.547)	0.228
Economic Stability (Outcome 7)*	1.625 (0.301)	1.655 (0.286)	-0.029 (.598)	0.100
Sample Size	152	149		

Source: UWM PFF Six-Month Follow-Up Surveys administered six months following the completion of the intervention. UWM PFF Post-Survey data used only to evaluate Outcome 7.

Notes: Effect sizes are calculated using the Hedge's g formula found in Appendix D.

Results for outcomes 1-6 are based on the second analytic sample; results for outcome 7 are based on the first analytic sample (i.e., intervention sample size = 126, comparison sample size = 139) because it involves the Post-Survey time point.

See Table F.1 for a more detailed description of each measure.

*For Conflict with Co-Parent (Outcome 6) and Economic Stability (Outcome 7), lower scores indicate better outcomes. For all other outcomes higher scores indicate better outcomes.

Table F.4. Regression results using data from UWM Follow-Up and Post-Survey to address the secondary research questions

Outcome measure	Predictor	B	95% CI	β	SE	t	p
Healthy Co-Parenting	Intercept	2.087	[0.874, 3.300]				
	Group	0.476	[0.256, 0.693]	.234	0.112	4.25	<.001
	Employment	-0.031	[-0.300, 0.238]	-.015	0.137	0.22	.822
	Education	0.055	[-0.188, 0.297]	.025	0.123	0.44	.658
	Income	0.266	[-0.004, 0.536]	.130	0.137	1.94	.053
	Healthy Co-Parenting pre-score	0.194	[0.039, 0.349]	.208	0.079	2.47	.014
	Communication with Co-Parent pre-score	-0.039	[-0.196, 0.118]	-.050	0.080	0.49	.626
	Conflict Resolution Skills pre-score	0.023	[-0.120, 0.166]	.027	0.073	0.32	.749
	Positive Parenting pre-score	0.254	[-0.060, 0.569]	.100	0.160	1.59	.113
	Engagement with Child pre-score	0.004	[-0.020, 0.028]	.022	0.012	0.33	.742
	Conflict with Co-Parent pre-score	0.012	[-0.047, 0.072]	.023	0.030	0.41	.685
	Economic Stability pre-score	0.042	[-0.202, 0.286]	.019	0.124	0.34	.734
Co-Parent Communication	Intercept	2.163	[0.998, 3.328]				
	Group	0.360	[0.148, 0.571]	.185	0.107	3.35	<.001
	Employment	-0.037	[-0.295, 0.221]	-.019	0.131	0.28	.777
	Education	0.185	[-0.048, 0.418]	.090	0.118	1.56	.120
	Income	0.111	[-0.148, 0.370]	.057	0.132	0.84	.400
	Healthy Co-Parenting pre-score	0.080	[-0.069, 0.229]	.090	0.076	1.06	.289
	Communication with Co-Parent pre-score	0.162	[0.011, 0.312]	.220	0.077	2.11	.035
	Conflict Resolution Skills pre-score	-0.016	[-0.153, 0.121]	-.019	0.070	0.23	.819
	Positive Parenting pre-score	0.179	[-0.123, 0.481]	.074	0.153	1.17	.244
	Engagement with Child pre-score	-0.008	[-0.031, 0.015]	-.046	0.012	0.71	.481
	Conflict with Co-Parent pre-score	0.032	[-0.026, 0.089]	.061	0.029	1.09	.277
	Economic Stability pre-score	0.040	[-0.194, 0.274]	.019	0.119	0.34	.736
Conflict Resolution Skills	Intercept	1.665	[0.518, 2.812]				
	Group	0.456	[0.247, 0.664]	.235	0.106	4.31	<.001
	Employment	0.010	[-0.244, 0.265]	.005	0.129	0.08	.936
	Education	0.272	[0.043, 0.502]	.133	0.117	2.34	.020
	Income	0.118	[-0.138, 0.373]	.060	0.130	0.91	.365

Outcome measure	Predictor	B	95% CI	β	SE	t	p
	Healthy Co-Parenting pre-score	0.108	[-0.039, 0.254]	.121	0.074	1.45	.149
	Communication with Co-Parent pre-score	0.090	[-0.059, 0.238]	.122	0.075	1.19	.235
	Conflict Resolution Skills pre-score	-0.001	[-0.136, 0.135]	-.001	0.069	0.01	.995
	Positive Parenting pre-score	0.357	[0.060, 0.654]	.148	0.151	2.36	.019
	Engagement with Child pre-score	-0.026	[-0.488, -0.003]	-.147	0.012	2.27	.024
	Conflict with Co-Parent pre-score	0.044	[-0.012, 0.101]	.086	0.029	1.55	.123
	Economic Stability pre-score	0.097	[-0.134, 0.327]	.046	0.117	0.82	.410
Positive Parenting Beliefs	Intercept	2.495	[1.899, 3.091]				
	Group	0.181	[0.073, 0.290]	.184	0.055	3.30	<.005
	Employment	-0.006	[-0.138, 0.126]	-.006	0.067	0.09	.929
	Education	0.105	[-0.015, 0.224]	.101	0.061	1.73	.085
	Income	0.017	[-0.116, 0.149]	.017	0.067	0.25	.803
	Healthy Co-Parenting pre-score	0.024	[-0.053, 0.100]	.052	0.039	0.61	.544
	Communication with Co-Parent pre-score	0.038	[-0.039, 0.115]	.101	0.039	0.97	.334
	Conflict Resolution Skills pre-score	-0.075	[-0.146, -0.005]	-.180	0.036	2.11	.035
	Positive Parenting pre-score	0.334	[0.179, 0.488]	.271	0.079	4.25	<.001
	Engagement with Child pre-score	-0.008	[-0.020, 0.004]	-.089	0.006	1.34	.181
	Conflict with Co-Parent pre-score	0.005	[-0.024, 0.034]	.019	0.015	0.31	.735
	Economic Stability pre-score	0.007	[-0.113, 0.127]	.007	0.061	0.11	.910
Engagement with Child	Intercept	3.650	[-2.870, 10.171]				
	Group	2.302	[1.118, 3.486]	.216	0.602	3.83	<.001
	Employment	-0.583	[-2.029, 0.862]	-.054	0.735	0.79	.427
	Education	1.379	[0.074, 2.684]	.123	0.663	2.08	.038
	Income	0.382	[-1.070, 1.833]	.036	0.737	0.52	.605
	Healthy Co-Parenting pre-score	0.245	[-0.588, 1.078]	.050	0.423	0.58	.563
	Communication with Co-Parent pre-score	0.425	[-0.419, 1.268]	.105	0.428	0.99	.322
	Conflict Resolution Skills pre-score	-0.791	[-1.559, -0.023]	-.175	0.390	2.03	.044
	Positive Parenting pre-score	1.247	[-0.444, 2.938]	.094	0.859	1.45	.148
	Engagement with Child pre-score	0.049	[-0.080, 0.178]	.051	0.066	0.75	.453
	Conflict with Co-Parent pre-score	-0.089	[-0.409, 0.231]	-.032	0.163	0.55	.583

Outcome measure	Predictor	B	95% CI	β	SE	t	p
	Economic Stability pre-score	0.397	[-0.914, 1.708]	.035	0.666	0.60	.552
Conflict with Co-Parent	Intercept	1.690	[0.665, 2.714]				
	Group	-0.057	[-0.243, 0.129]	-.034	0.095	0.60	.547
	Employment	0.071	[-0.156, 0.298]	.042	0.115	0.62	.538
	Education	0.132	[-0.073, 0.337]	.075	0.104	1.27	.206
	Income	-0.100	[-0.324, 0.132]	-.057	0.116	0.83	.409
	Healthy Co-Parenting pre-score	-0.006	[-0.137, 0.125]	-.007	0.067	0.08	.933
	Communication with Co-Parent pre-score	0.063	[-0.070, 0.195]	.100	0.067	0.94	.350
	Conflict Resolution Skills pre-score	-0.022	[-0.142, 0.099]	-.030	0.061	0.35	.726
	Positive Parenting pre-score	-0.442	[-0.708, -0.176]	-.212	0.135	3.27	.001
	Engagement with Child pre-score	0.015	[-0.005, 0.036]	.101	0.010	1.50	.135
	Conflict with Co-Parent pre-score	0.086	[0.036, 0.137]	.194	0.256	3.38	<.001
	Economic Stability pre-score	-0.106	[-0.312, 0.100]	-.059	0.105	1.02	.311
Economic Stability	Intercept	0.989	[0.397, 1.582]				
	Group	-0.029	[-0.138, 0.080]		0.055	0.53	.598
	Employment	-0.094	[-0.234, 0.047]		0.071	1.31	.190
	Education	0.004	[-0.117, 0.125]		0.031	0.06	.952
	Income	0.107	[-0.035, 0.249]		0.072	1.48	.140
	Healthy Co-Parenting pre-score	0.043	[-0.035, 0.120]		0.039	1.09	.278
	Communication with Co-Parent pre-score	-0.011	[-0.094, 0.073]		0.042	0.25	.805
	Conflict Resolution Skills pre-score	-0.006	[-0.086, 0.074]		0.041	0.15	.879
	Positive Parenting pre-score	-0.071	[-0.223, 0.082]		0.077	0.91	.362
	Engagement with Child pre-score	-0.013	[-0.025, -0.001]		0.006	2.14	.033
	Conflict with Co-Parent pre-score	0.034	[0.003, 0.066]		0.016	2.13	.034
	Economic Stability pre-score	0.523	[0.407, 0.640]		0.059	8.83	<.001

Notes: Categorical variables were coded as follows: For Group, Control group = 0, Treatment group = 1; for Race, Other = 0, Black = 1; for Age, 44 years and younger = 0, 45 years and older = 1; for Employment, Not Employed = 0, Employed = 1; for Income, Less than $500 per month = 0, $500 a month or more = 1; for Marital Status, Never Married = 0, Married = 1. Higher scores represent better outcomes, except for Outcomes 6 and 7, where lower scores represent better outcomes. Results for outcomes 1-6 are based on the second analytic sample; results for outcome 7 are based on the first analytic sample because it involves the Post-Survey time point.