Impact Evaluation of the Brothers United Fatherhood Program in Toledo, Ohio
Final Impact Evaluation Report for Pathway, Inc.

November 9, 2020

Prepared by

Educational Evaluators, Inc.,

Steph Fluegeman, MPH, Aida Hernandez, MA,

Joseph Donnelly, PhD, and Michael Young, PhD

Recommended Citation:

Fluegeman, S. Hernandez, A. Young, M. (2020). *Impact Evaluation of the Brothers United Fatherhood Program in Toledo, Ohio*. Educational Evaluators, Inc.

Acknowledgements:

The Brothers United Fatherhood Program was a collaborative project developed and administered with the assistance of the program director, facilitators and staff of Pathway, Inc. and evaluated by Educational Evaluators, Inc.

Disclosure:

Steph Fluegeman, Aida Hernandez, Michael Young, and Joseph Donnelly declare that they have no conflict of interest.

This publication was prepared under Grant Number 90FK0087 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 Brothers United Fatherhood Program in Toledo, Ohio"

Objective. The purpose of the study was to examine the impact of the Brothers United Fatherhood Program (BUFP) on participants' scores on four fatherhood-related outcome measures. The 58-hour program includes case management and four curricula that addressed the major topics guiding this research: fatherhood development, relationship enhancement, financial literacy/work.

Study design. The study used a randomized control trial (RCT) design. Participants were young fathers, mostly African American, ages 20-24 living in low-income, high crime areas, of Toledo, Ohio. All participants completed a baseline survey and then were assigned to either the intervention or the control group, in a 2:1 ratio. The intervention participants received the BUFP. Control participants received a resource list only. Twelve months after enrollment, participants completed a follow-up survey. Matched pretest and follow-up surveys were obtained from 151 intervention participants and 71 control participants. Data were analyzed using one-way analysis of co-variance, controlling for pretest scores, employment status, arrest record, and relationship status.

Results. Intervention participants received an average of 43 hours out of 58 total hours and attended an average of 14 out of 24 sessions. The results indicated the BUFP helped participants gain a greater understanding of financial planning ($p < .01$). The effect size, or the magnitude of the relationship between study group and measure of financial planning, indicates a medium effect size (g=0.74) (Cohen, 1977). Intervention participants did not demonstrate progress to greater economic stability ($p = .29$), nor did they improve on conflict resolution with their partner (p =.58) or increase their understanding of healthy marriage ($p = .79$) compared to control participants.

Conclusion. Based on the results of this study it appears that the BUFP can help young men improve their understanding in financial planning and this could be valuable to them in becoming more responsible fathers.

Contents

Tables

Figures

Impact Evaluation of Brothers United Fatherhood Program in Toledo, Ohio

I. INTRODUCTION

A. Introduction and study overview

This study was the first evaluation of a program developed by Pathway, Inc.: Brothers United Fatherhood Program (BUFP). Pathway, Inc. has many years of experience in providing local programs to support disadvantaged populations in Toledo, Ohio. The purpose of the BUFP was to serve young, low-income fathers in areas where high crime, poverty and lack of services exist for fathers. The BUFP integrated curricula and activities based on areas of priority identified through a community needs assessment. The program designed long-term goals in the logic model addressing the following priority areas: Improve family functioning, improve adult and child well-being, increase economic mobility, decrease poverty, and reduce recidivism. The study was developed to assess the effectiveness of the BUFP specifically with 20- to 24-year-old fathers. Using a randomized control trial (RCT), the study examined whether the intervention based on fatherhood development, relationship enhancement, financial management, and job-readiness skills produced significant improvements for participants in four outcome measures when compared to control participants over the long-term (12-month follow-up).

B. Primary research question(s)

This study had four primary research questions to understand the impact of the BUFP on intervention participants' attitudes and behaviors related to healthy marriage and their finances: Do program participants, when compared to control participants, at one year after enrollment have more positive self-reported attitudes and behaviors with respect to: (1) Learning new opportunities for economic mobility and the demonstration of job readiness skills; (2) Improved communication and conflict resolution toward partner, including resolving conflicts leading to domestic violence; (3) Understanding of healthy marriage and its value to fathers; (4) Demonstrated understanding of financial planning. The study is registered with ClinicalTrials.gov (Identifier: NCT03021226).

II. INTERVENTION AND CONTROL CONDITIONS

This section provides a brief description of the intervention being evaluated and the services that are intended as the control to the intervention, including any "business-as-usual" resources available.

A. Description of program as intended

The BUFP was a comprehensive multi-component program that offered Responsible Fatherhood education, healthy relationship education, individualized financial planning, job and career advancement training, and related activities, including subsidized employment to increase economic stability and self-sufficiency. The 58-hour intervention utilized four "core" curricula that address the three major topic areas guiding the research study: fatherhood development, relationship enhancement, and financial literacy and employment. The four curricula were selected for implementation because they addressed the priority areas identified by the program team and were found to be culturally sensitive and appropriate for the participants recruited and served (young low-income fathers).

The program team delivered a curriculum-based intervention over a six-week period with sessions occurring four times a week for two to three hours per session. Participants received the intervention in small group sessions at designated times and locations in the community of Toledo. For each program cohort, there were sessions offered in the morning and in the afternoon. If a participant missed a session, they had the opportunity to make up the session through a one-on-one session with either their coach or a trained facilitator. In addition to being offered these curriculum-based sessions, participants in the intervention group were assigned to coaches who provided individual meetings with participants to identify potential needs and determine eligibility for linkages to community partners and resources in their communities. These resources included, but were not limited to, mental health services, housing services, job placement, and child support.

In addition to the curricula and case management components, the BUFP facilitated activities for program graduates including monthly BU Nation meetings, a BU basketball team, and community events hosted with partners. All intervention components were logged and tracked in nFORM.

Table II.1 provides a description of the program by component, including the curriculum and content covered, the dosage and schedule of the program, how the program was delivered, and the target population. See Appendix A for the program logic model.

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

Component	Curriculum and content	Dosage and schedule	Delivery	Target Population
Intervention				
Fatherhood Development	*Fatherhood Development*: Parenting skills, team-parenting and relationships, domestic violence, health	24 hours total: Comprising 16 90-minute sessions, occurring four times a week for four weeks of the program	Group lessons provided at a designated location in the community by two trained facilitators in every session	Low-income fathers 20-24 years of age
Relationship Enhancement	*Basic Training for Singles*: Marriage education, conflict management; communication	8 hours total: Comprising eight one-hour sessions, occurring four times a week for one week of the program	Group lessons provided at a designated location in the community by two trained facilitators in every session	Low-income fathers 20-24 years of age
Financial Literacy/ Employment	*Workplace Survival Skills*: Respecting yourself and others, planning and reaching goals, communicating with others, advancing work and school, understanding your legal rights and responsibilities, getting ready for employment, keeping a job, managing money, taking care of your health, and celebrating success	20 hours: Comprising 10 two-hour sessions, occurring four times a week for one week of the program	Group lessons provided at a designated location in the community by two trained facilitators in every session	Low-income fathers 20-24 years of age
Financial Literacy/ Employment	*Money Habitudes 2*: Financial literacy, money management and personal finance, asset building	3 hours total: Comprising three one-hour sessions three times a week for one week of the program	Group lessons provided at a designated location in the community by two trained facilitators in every session	Low-income fathers 20-24 years of age
Fatherhood Development	BU Nation comprised of program graduates	2 hours total: Meets one time per month for family-centered activities and support group	Meets one time per month at a location in the community (church, park, school) to engage fathers in family-centered activities (Brunch, Father's Day activities)	Low-income fathers 20-24 years of age

Component	Curriculum and content	Dosage and schedule	Delivery	Target Population
Case Management	Case management: mental health services, housing services, job placement, and child support	At least 1 hour total: Meets individually with Coaches; variability in length of coaching	One-on-one; Coaches assigned to provide individual meetings with participants to identify potential needs and determine eligibility for linkages to community partners and resources in their communities	Low-income fathers 20-24 years of age
Control				
Resource List	Did not receive the components of the intervention; received Resource List with organizations to contact for housing assistance, legal aid, basic needs like food and clothes, mental health, substance abuse help, education assistance, and general health needs	Data collection specialist provided participants a resource list	Approximately 1-hour interaction during study enrollment	Low-income fathers 20-24 years of age fathers

The BUFP team included men and women of different ages and backgrounds who were familiar with community approaches to programming and had connections to Toledo. There were 13 positions created to address all programmatic activities and operational duties. Four of the positions included coaches who were assigned intervention participants for case management. Each of the four coaches served an average of 54 sample participants during the tenure of the study. Coaches, a program director, a program assistant, and a program aide received training in program facilitation. One data collection specialist, one intake coordinator and four facilitators were also part of the BUFP team. All the BUFP Facilitators that delivered curriculum components received training in curriculum delivery. In addition to curriculum-training, all the BUFP staff received additional trainings in child maltreatment and domestic violence awareness. All the BUFP staff completed training in ethics and human subjects research provided through Protecting Human Research Participants (PHRP) endorsed by the National Institutes of Health (NIH). Table II.2 provides details specific to the educational and training requirements of staff that deliver curriculum components and resources.

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		
Fatherhood Development Relationship skills workshops	Facilitators held at least a bachelor's degree and received four days of initial training.	Facilitators received a half-day of semi-annual refresher training in the intervention's curricula from program staff.
Relationship Enhancement Relationship skills workshops	Facilitators held at least a bachelor's degree and received two days of initial training.	Facilitators received a half-day of semi-annual refresher training in the intervention's curricula from program staff.
Financial Literacy/Work: Economic stability workshops	Facilitators held at least a bachelor's degree and received two days of initial training.	Facilitators received a half-day of semi-annual refresher training in the intervention's curricula from program staff.
Counterfactual		
Employment and mental health services and family resources provided through partner agencies	Data collection specialist held at least a bachelor's degree and received training and information about resources available that was organized into resource lists.	Not applicable

B. Description of counterfactual condition as intended

Participants in the control group were not provided with any programmatic content but were provided with a resource list of other resources in the community after assignment to the control group (see Table II.1). The resource list included organizations to contact for housing assistance, legal aid, basic needs like food and clothes, mental health, substance abuse help, education assistance, and general health needs. Participants were asked about their participation in services during enrollment, and participation data for government programs such as WIC, TANF, SSI, SNAP, etc. is stored in nFORM. After the completion of the evaluation activities, including completion of the 12-month follow-up survey, control participants could participate in the full program provided to intervention participants.

Control participants received monthly calls and messages from the data collection specialist to ensure that the BUFP staff had the most up-to-date contact information (Table II.2 provides details specific to the educational and training requirements of staff responsible for contacting control participants). Control participants received periodic incentives for completing a posttest 6 weeks after baseline/pretest, a 12-month follow-up survey, and an incentive 6 months after enrollment to reinforce their participation in the study.

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

The local evaluation examined data on implementation fidelity, the average dosage of programming, the quality of staff-partner interactions, participants' engagement, and the context of implementation. These data provide information on how the program was implemented and received by the participants. These questions were selected to direct the process evaluation with

the intent to integrate findings together with those of the impact evaluation. The research questions examined if differences in planned implementation versus actual implementation occurred that may have affected the impact of the program (see Table II.3 for research questions of the implementation study).

Table II.3. Research questions of the implementation study, by implementation element

Implementation element	Research question
Intervention Group Questions	
Fidelity	• Were all intended intervention sessions offered and for the expected duration?
Dosage	• On average, how many hours of the intervention did the intervention group participate in? • What was the average number of sessions the intervention group participated in?
Quality	• What was the quality of staff–participant interactions?
Engagement	• How engaged were intervention group members in the intervention?
Context	• What external events occurred that affected implementation?

III. STUDY DESIGN

This section provides a brief description of the study design and the process for creating intervention and control groups.

A. Sample formation and research design

The target population of the study was low-income fathers and fathers at-risk for recidivism ages 20 to 24 that reside primarily in high-crime areas in the Toledo area within Lucas County, Ohio. Participants also included expectant fathers. Participants were married, single, or in a relationship. Any eligible father interested in the BUFP could enroll in the study.

The process for random assignment was that all participants took a Qualtrics pre-survey and then at the end of the survey, Qualtrics randomly assigned participants to the intervention or control groups through an electronic random number generator. Randomization happened on a rolling basis as participants enrolled in the intervention or control groups for the cohorts of programming. Participants could join ongoing cohorts up to three weeks after the start date of the cohort. After the three-week cut off, participants were put on hold and placed into the following cohort start date. Participants were assigned on a 2:1 ratio of intervention to control. In other words, for every two participants in the 20- to 24-year-old intervention cohort, one participant in this age group was assigned to control, which was a 0.67 probability that an individual was assigned to the intervention group. Program staff found out participants' assigned status through Qualtrics and informed participants of their assignment immediately after they completed the survey. The evaluation team informed the program staff, including case managers, data collection specialist, and intake coordinator, that group selection could not be changed due to this being a research study. Program staff informed participants that group selection could not be changed.

Participants were required to provide informed consent before participation in the program and prior to the collection of any data.

B. Data collection

1. Implementation analysis

This section lays out how measures were constructed to address the implementation study research questions.

Fidelity. To measure the number of sessions delivered and average session duration, the study examined data logged by the program staff in nFORM, OFA's management information system for logging the dates of each of the program sessions and session completion. Data was exported from nFORM and sent to the evaluation team a minimum of one time per month.

Dosage. To determine the average number of sessions and the number of intervention hours received, the study reviewed information gathered via nFORM. The program staff utilized the nFORM management information system for logging attendance at scheduled small group

sessions. Attendance was recorded at the start of each workshop session and amended if participants arrived at the session later or completed a make-up session in another group or one-on-one with the coaches. Data was exported from nFORM and sent to the evaluation team a minimum of one time per month.

Quality. The study reviewed participants' responses gathered from the Internal Feedback Form that consisted of six Likert scale questions and four open-ended questions designed to acquire specific information regarding their experience in participating in curriculum sessions. The form was provided to participants at the close of each curriculum in the program. Questions asked about their thoughts on interacting with the program facilitators, the applicability of the skills learned, if they enjoyed participating, and how participants would describe the curriculum to a potential participant. The form was also voluntary for participants. If participants did not attend the last session, coaches reached out to participants to remind them to complete missing forms and survey items. Due to the voluntary nature of the form and timing of collection at the last session, the form had a low response rate of 27%. Participants that were engaged in the sessions and attended regularly were more likely to be the participants represented in the Internal Feedback Forms data because they were in attendance during the collection and willing to complete another data collection survey.

To track differences that may have impacted program delivery or the implementation elements, the program and evaluation teams scheduled meetings two times a month to discuss what modifications were necessary.

Engagement. To measure engagement, the study reviewed attendance data collected in nFORM to determine the percentage of participants that completed the program. Engagement was measured by the number of participants completing the program. Data was exported from nFORM and sent to the evaluation team a minimum of one time per month.

Context. The evaluation team reviewed monthly evaluation meeting notes to identify external events that may have impacted attendance such as policy changes or community events. All members of the evaluation team attended the meetings and select members of the BUFP team attended, including the program director, program assistant, program aide and the data collection specialist. These monthly meetings were used as time to discuss the RCT enrollment, RCT retention numbers, preliminary findings, and external factors that may have impacted that month's attendance or survey collection.

See Table B.1 in Appendix B for a description of the research questions and measures developed for each implementation element, and Table B.2 for data used to address implementation research questions.

2. Impact analysis

This section provides a brief description of the quantitative data collection process during the research study. Quantitative data for the impact analysis was collected electronically using tablets or computers with wireless internet connection. The study utilized surveys developed by

OFA and the local evaluation team. The surveys developed by OFA were the nFORM Entrance and Exit Surveys. These surveys served as a baseline/pretest and posttest that participants completed at the start and end of their participation in the program. The survey developed by the study team was referred to as the Local Evaluation Survey. This survey was administered using Qualtrics a total of three times, as a pretest/baseline survey, posttest survey, and a 12-month follow-up survey. The follow-up version of the Local Evaluation Survey included a selection of questions from nFORM needed to construct outcome measure subscales to match baseline subscales. All survey instruments utilized in the study were self-reported surveys. To gather demographic information, the study reviewed responses to the Applicant Characteristics Survey, a one-time survey developed by OFA and administered electronically via nFORM during participant enrollment.

Data were collected on-site by the data collection specialist and the intake coordinator. If participants had moved out of the area or were otherwise unreachable for in-person data collection, the survey was administered over the phone with participants. All these methods were used to achieve a high response rate (as will be presented in section IV.). Protocols for collecting data remained the same, regardless of the method used to collect the data. See Table B.3 in Appendix B for key features of the impact analysis data collection.

Participants who provided written consent to participate in the program completed the baseline, or pretest, survey prior to program implementation. Participants were surveyed again at the close of the program and 12 months after enrollment. The evaluation team trained program staff and the data collection specialist in evaluation procedures with specific attention to survey administration and data collection protocols. The data collection specialist or intake coordinator inputted the following for each participant's survey: the participant number (nFORM ID) and test time (baseline/pretest, posttest, or 12-month follow-up). The nFORM ID number allowed the study to match participants' initial survey responses with responses across administrations.

To maximize the validity of these self-report data, program staff emphasized that (a) all individual responses are confidential and will not be revealed to family members, or law enforcement officials, (b) participants were to keep their eyes on their own survey and not to discuss any of the questions with other participants, and (c) participants could skip any question they did not wish to answer. They were, however, encouraged to answer all questions, with emphasis placed on the importance of honest answers. To minimize problems with survey completion and submission, the data collection specialist demonstrated how to answer questions, move to the next page of the survey, and how to submit the completed survey on the electronic devices.

IV. ANALYSIS METHODS

A. Analytic sample

The analytic sample for the longer-term effects of the program include those participants who have completed a baseline and follow-up survey. Within these groupings, participants must have complete data for the outcome measures involved in a specific analysis. Some participants with missing baseline data for the outcome measures involved in a specific analysis were included, as per the description concerning the handling of missing data (See Appendix C). The primary analysis sample included 151 intervention participants and 71 control participants.

Although program staff clearly communicated to control participants that they would not receive any programmatic components, some circumstances led to control group participants attending program sessions, such as the father needed services immediately due to court orders or emotional distress and so the program staff placed them in the sessions regardless of the control assignment. The control participants who received intervention services were tracked, but their results were analyzed with the control group to follow the intent-to-treat approach. In total, 7% (7/103) of fathers randomly assigned to the control group attended sessions alongside the treatment participants. About 6% (4/71) of control group follow-up surveys were from fathers who had attended at least one session.

The majority of participants who were assigned to the control group did not receive any of the intervention components offered by the program while they were on the wait list to receive the intervention (after completing the 12-month follow-up survey). It is unlikely that control participants received services similar to the intervention through other organizations due to the scope of the intervention – most programs in the community for fathers do not involve such a time-intensive intervention.

Control participants interacted with the data collection specialist, who was trained to follow data collection procedures. The data collection specialist held a master's degree and completed training in ethics and human subjects research provided through Protecting Human Research Participants (PHRP) endorsed by NIH. If a participant was exposed to a component of the intervention, either through a curriculum session or a BU-sponsored community event, the BUFP staff alerted the evaluation team to track the participant. The BUFP staff took steps to ensure control participants did not receive services. Staff carried around clipboards with names of control participants and cross-referenced the list before allowing participation. If control participants were identified, staff would remind the participant they are not allowed to take part until after they completed their 12-month follow-up survey. There were 7 control participants that received programming. These control participants were fathers in need of services immediately and program staff placed them in the sessions despite the control assignment; however, the analysis upheld random assignment and control participants who received any programming hours were in the control group for the analysis.

Sample Attrition

The local evaluators tracked the numbers of intervention participants and control participants that enrolled and completed each of the three surveys. The overall attrition rate (29.7%) and differential attrition rate (30.1% intervention, 31.1% control) were low according to the What Works Clearinghouse guidance, indicating that this is a well-executed RCT.

Table IV.1. describes the individual sample sizes by intervention status and includes the response rates for each study group at each data collection time point. Both study groups had a 100% total response rate at baseline/pretest, a 72% total response rate at the posttest survey and a 70% total response rate at the 12-month follow-up survey. See Appendix B for the CONSORT diagram with details on the final analytic sample.

Table IV.1. Individual sample sizes by intervention status

Number of individuals	Intervention sample size	Control sample size	Total sample size	Total response rate	Intervention response rate	Control response rate
Assigned to condition	216	103	319	n.a.	n.a.	n.a.
Contributed a baseline/pretest survey	216	103	319	100%	100%	100%
Contributed to posttest survey (6-weeks after baseline survey)	157	73	229	72%	72%	71%
Contributed to 12-month follow-up survey	151	71	222	70%	70%	69%

n.a. = not applicable.

B. Outcome measures

The evaluation study examined four outcomes of interest. For the first outcome, economic mobility and job readiness, existing items from the nFORM survey were selected. The measure included items that grouped together to express knowledge around good job habits and job skills like applying for jobs and keeping jobs. The second outcome, improved communication and conflict resolution toward partner, was comprised of survey questions from the Local Evaluation (LE) survey. The four items came from a Healthy Marriage Program survey published by Mathematica. The third outcome, understanding of healthy marriage and its value to fathers, was compiled using Confirmatory Factor Analysis to achieve high internal reliability. Items grouping around healthy marriage behaviors and attitudes were examined for closeness of fit. A combination of six items from nFORM and the LE survey comprises the outcome measure. The items from the LE Survey come from a Healthy Marriage Program survey published by Mathematica. Measuring an understanding of financial planning (the fourth outcome) starts with scales items around budgeting, attitudes about money, and the relationship between working and being financially well. All four items come from the LE Survey. For each outcome, the item mean score was divided by the number of category responses so the outcome measures all range from 0 to 1 and averaged for a composite mean score. See Table IV.2 for a description of each outcome measure, source of the measure, and questions that compile each scale.

The Cronbach's alpha score for each outcome measure at baseline is reported to assess internal consistency. Confirmatory Factor Analysis (CFA) indicate a good fit for each pre-survey measure and follow-up measure. See Table D.1 in Appendix D for results from CFA for each factor.

Baseline/pretest survey instrument

The internal consistency was measured on the pre-survey instrument. The economic mobility and job readiness skills subscale consisted of 6 items (α= .72), the communication and conflict resolution subscale consisted of 4 items (α= .88), the healthy relationship and marriage subscale consisted of 6 items (α= .72), and the financial planning subscale consisted of 4 items (α= .62). Although the financial planning subscale had lower alpha levels at pretest, the scale performed satisfactory at follow-up and the CFA results indicate the outcome is adequately defined (see Table D.1 in Appendix D).

Follow-up survey instrument

When assessing internal consistency on the follow-up survey, similar scores for Cronbach's alpha were found. The economic mobility and job readiness skills subscale consisted of 6 items (α= .94), the communication and conflict resolution subscale consisted of 4 items (α= .81), the healthy relationship and marriage subscale consisted of 6 items (α= .70), and the financial planning subscale consisted of 4 items (α= .71).

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

Outcome measure	Description of the outcome measure	Source	Timing of measure
1. Learning new opportunities for economic mobility and the demonstration of job readiness skills	The outcome measure is derived from six items on both the nFORM survey and local evaluation survey; compared pretest to follow-up. **1.** I have good job skills. **2.** I know how to apply for a job. **3.** I feel confident in my ability to conduct an effective job search for a job I want. **4.** I feel confident in my interviewing skills. **5.** I am usually on time for work. **6.** If I'm not going to go to work, I let my supervisor know ahead of time. Scale: Strongly Agree to Strongly Disagree, 4-point scale. Cronbach's alpha pre-survey measure: 0.72 Cronbach's alpha follow-up survey measure: 0.94	nFORM entry survey; local evaluation entry and follow-up survey	At follow-up (1 year after enrollment)
2. Improved communication and conflict resolution toward partner, including those leading to domestic violence	The outcome measure is derived from four items on the local evaluation survey; compared pretest to follow-up. **1.** My partner/spouse and I were good at working out our differences. **2.** I felt respected even when my partner/spouse and I disagreed. **3.** When my partner/spouse and I had a serious disagreement, we worked on it together to find a resolution. **4.** When my partner/spouse and I had a serious disagreement, we discussed our disagreement respectfully. Scale: Strongly disagree to Strongly agree, 4-point scale. Cronbach's alpha pre-survey measure: 0.88 Cronbach's alpha follow-up survey measure: 0.81	Local evaluation entry and follow-up survey	At follow-up (1 year after enrollment)

Outcome measure	Description of the outcome measure	Source	Timing of measure
3. Understanding of healthy marriage and its value to fathers	The outcome measure is derived from six items on the local evaluation survey; compared pretest to follow-up. **1.** I know the basics of what makes a marriage succeed or fail. **2.** I view our co-parenting or marriage/ relationship as lifelong. Scale: Strongly disagree to Strongly agree, 4-point scale. **1.** My partner/spouse seemed to view my words or actions more negatively than I meant them to be. **2.** Our arguments became very heated. **3.** Small issues suddenly became big arguments. **4.** My partner/spouse or I stayed mad at one another after an argument. Scale: Often, Sometimes, Hardly, Never, 4-point scale. Cronbach's alpha pre-survey measure: 0.72 Cronbach's alpha follow-up survey measure: 0.70	nFORM entry survey; local evaluation entry and follow-up survey	At follow-up (1 year after enrollment)
4. Demonstrated understanding of financial planning	The outcome measure is derived from four items on the local evaluation survey; compared pretest to follow-up. **1.** Budgeting is the best way to make sure I know where my money is going every month. **2.** I set financial goals for my personal needs and my children. **3.** Part of keeping a job is working hard, even when I would rather be somewhere else. **4.** I know how I can use my current job to advance to the job I really want. Scale: Strongly disagree to Strongly agree, 4-point scale. Cronbach's alpha pre-survey measure: 0.62 Cronbach's alpha follow-up survey measure: 0.71	nFORM entry survey; local evaluation entry and follow-up survey	At follow-up (1 year after enrollment)

C. Baseline equivalence and sample characteristics

Baseline equivalence was assessed comparing the intervention and control groups on demographic measures and scores for the outcome measures. Chi-square analysis was used to determine whether the specific demographic variables are independent of intervention condition (intervention or control).

There were no statistically significant demographic differences by group (intervention and control) at baseline among employment status, arrest record, education level, and relationship status. The percentage of participants that were employed did not differ by group, X^2 (1, N = 222) = 0.00, p = .99. The percentage of participants with an arrest record did not differ by group, X^2 (1, N = 222) = 0.01, p = .92. The percentage of participants who did not have a high school degree or GED at baseline did not differ by group, X^2 (1, N = 222) = 0.33, p = .56. The percentage of participants who were single at baseline did not differ by group, X^2 (1, N = 222) = 1.87, p = .17. See Table IV.3 below for a full description of demographic comparisons by group.

Analysis of variance was used to determine whether there are baseline differences between the intervention group and the control group for each of the outcome variables. The evaluation examined effect sizes using Hedge's g for the continuous outcome variables. Additionally, unadjusted means and standard deviations for the baseline measures for the intervention and control group for each outcome measure as well as p-values are provided.

Overall, there were no statistically significant differences detected between means at baseline for any of the outcome measures: progress toward greater economic stability, (F(1, 295) = 0.107, p = 0.74); improved communication and conflict resolution towards partner, (F(1, 311) = 0.168, p = 0.68); improved healthy relationships and marriage skills (F(1, 311) = 0.173, p = 0.68); improved understanding of financial planning (F(1, 312) = 0.081, p = 0.78).

Using the WWC standard for baseline equivalence, the effect sizes indicate that the intervention and control groups are similar at baseline across all demographic measures of interest and most baseline scores for the outcome measures (see Table IV.3 for baseline equivalence information). There were pre-intervention differences for the outcome measures of improved communication and conflict resolution towards partner, and improved healthy relationship and marriage skills, so the benchmark analysis approach uses ANCOVA to adjust for these differences at baseline.

Table IV.3. Summary statistics of key baseline measures and baseline equivalence across study groups, for individuals completing the 12-month follow-up

Baseline measure	Intervention mean (standard deviation)	Control mean (standard deviation)	Intervention versus control mean difference (p-value of difference)	Effect size
Not employed	57.6%	57.7%	-0.1% (.99)	n.a.
Has an arrest record	66.9%	67.6%	-0.7% (.92)	n.a.
No high school diploma/GED	41.2%	36.9%	4.3% (.56)	n.a.
Single	48.8%	37.9%	10.9% (.17)	n.a.
Progress toward greater economic stability, including skill attainment and employment	0.83 (0.14)	0.83 (0.13)	0.00 (.97)	0.00
Improved communication & conflict resolution towards partner	0.65 (0.22)	0.62 (0.25)	0.03 (.49)	0.12
Improved healthy relationship and marriage skills	0.60 (0.19)	0.67 (0.20)	-0.07 (.06)	-0.36
Improved understanding of financial planning	0.76 (0.16)	0.76 (0.18)	0.00 (.89)	0.00
Sample size	151	71	n.a.	n.a.

n.a. = not applicable.

V. FINDINGS AND ESTIMATION APPROACH

A. Implementation evaluation

1. Key findings

The implementation research questions focused on the fidelity of the BUFP implementation, the dosage of the BUFP the program team provided, the quality of implementation, participant engagement, and the context for implementation. The BUFP was implemented with fidelity and the program achieved its goals of delivering all the intended intervention components of Fatherhood Development, Basic Training for Singles, Workplace Survival Skills, and Money Habitudes. Participants completed an average of 43 hours of the 58 hours of total programming. The program was implemented as planned and participants who completed feedback forms reported that they perceived the BUFP to be helpful, the facilitators to be friendly and supportive, and that they enjoyed participating in the group. Fifty-five percent of participants completed 75% or more of the program; this level of program completion was considered as engaged in the program. Barriers such as transient housing and job placements, prevented participants from attending every session; however, there were no other contextual issues to note that negatively affected the program's capacity to deliver the program. Details related to the implementation evaluation results are presented below.

2. Fidelity and dosage results

Over the course of the evaluation study, the BUFP was implemented with fidelity and trained facilitators implemented all programmatic components. Sessions lasted an average of 175 minutes as planned. Upon examining the program dosage, the study found intervention participants received an average of 43 hours out of 58 total hours and attended an average of 14 out of 28 sessions. Engagement was measured by examining what percentage of participants completed the program. Of all the intervention participants, twenty-eight percent attended 58 hours of programming. Fifty-five percent of participants completed 75% or more of the program. Eleven percent of intervention participants did not attend any sessions and had no recorded hours. On average, intervention participants attended 14 sessions. No participants were flagged as receiving similar services outside of the programming.

3. Participants' perceptions of the quality of the BUFP sessions

At the close of the curriculum sessions, facilitators provided participants with an Internal Feedback Form to collect feedback regarding their impression of the quality of the program implementation. Response rate from the Internal Feedback Form was low and the most engaged participants contributed responses; therefore, these results do not represent the full sample. The overall average score when including the responses collected from 20 to 24-year-old respondents to the Internal Feedback Forms is 27.8/30, or 93%, indicating positive feedback to the presentation of the workshop classes and activities. Although it is important to note that feedback to the program was positive, one limitation to this measure was that not every participant completed the Internal Feedback Form if they did not attend the last curriculum session. Participants were also informed that completing the form was voluntary and this may have

resulted in the response rate of 27%. For those that did respond to the form, the following questions measured the participants' experience of the quality of program and demonstrated high feedback scores of 5 (Strongly Agree) and 4 (Agree) across each of the four curricula: Question 1, "I enjoyed participating in this group"; Question 2,"I learned new skills and strengths in this group"; and Question 6, "The facilitators were friendly and supportive." To gain a sense from participants regarding their experiences with how the curriculum components were delivered, on the Internal Feedback Form, participants were asked, "What activity from this group affected you the most in a positive way?" Overall, nearly one-fifth of respondents were most impacted by relationship improvement skill activities, communication activities, and money management exercises. Respondents also frequently reported that "everything" or "all" skills they learned were perceived as impactful.

When asked, "How would you describe this workshop to a father who is interested in coming to group," one-third of the respondents stated that they would describe the program as helpful to fathers. Respondents also frequently reported they would describe the program as great and helps to become a better parent, indicating that the relationship skills components of the program that were implemented were perceived to be the most impactful when self-reported by the respondents.

Although some information can be gleaned from the Internal Feedback Forms and the limited sample of responses, it is unknown how the majority of participants perceived the quality of the program because of the lack of data.

4. Contextual factors that influenced attendance

Contextual factors that influenced program attendance appeared to be related to challenges and barriers outside of the scope of the program and not the result of policy changes or significantly impactful external events. Based on a review of the responses from program staff in monthly call notes, factors that contributed to program attrition rates included a transient population and unstable housing that resulted in homelessness and moving in and outside the city of Toledo. The high-crime area in which participants were recruited to the program presented as a barrier when there was an increase in community conflict or violence, with some participants avoiding interactions with other participants. Program staff also attributed attrition to higher rates of joblessness, and some participants left the program if it conflicted with opportunities to financially support themselves. Although the program postponed sessions due to inclement weather, this did not have a significant impact on how the program was implemented and all programmatic activities resumed shortly afterwards. In cases of missed sessions due to weather conditions, sessions were rescheduled, or participants could attend make up sessions.

B. Primary impact evaluation

1. Key findings

The primary research questions of the evaluation focused on the impact of the BUFP on intervention participants' job readiness, communication and conflict resolution with their partner, understanding of the value of healthy marriage and relationships, and financial planning when compared to control participants. The BUFP had a significant impact on intervention participants' understanding of financial planning compared to control participants ($p < .01$); however, the BUFP did not have a statistically significant impact on intervention participants' behavior with respect to understanding of the value healthy marriage and relationships ($p = .79$), communication and conflict resolution with their partner ($p = .58$), or with progress toward greater economic stability, including skill attainment and employment ($p = .29$).

A One-way ANCOVA was conducted to determine a statistically significant difference between study groups on progress toward the measure of greater economic stability, controlling for relationship status, employment status, arrest record, and pretest score. There is not a significant effect from study group on this measure after controlling for relationship status, employment status, arrest record, and pretest score, $F(1, 162) = 1.14$, $p = .287$.

To determine a statistically significant difference between study groups on improved conflict resolution toward partner, a one-way ANCOVA was conducted, controlling for relationship status, employment status, arrest record, and pretest score. There is not a significant effect from study group on this measure after controlling for relationship status, employment status, arrest record, and pretest score, $F(1, 170) = 0.31$, $p = .581$.

To determine a statistically significant difference between study groups on improved healthy relationship and marriage skills, a one-way ANCOVA was conducted, controlling for relationship status, employment status, arrest record, and pretest score. There is not a significant effect from study group on this measure after controlling for relationship status, employment status, arrest record, and pretest score, $F(1, 166) = 0.07$, $p = .790$.

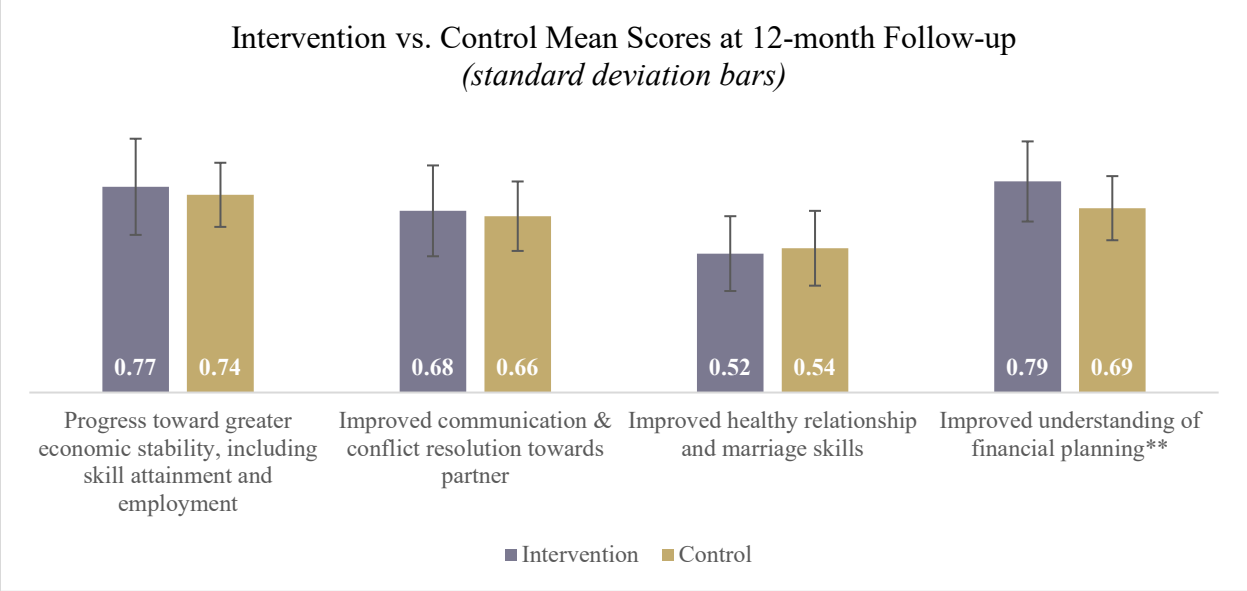
Study groups differed significantly at follow-up for the measure of improved understanding of financial planning, $F(1, 172) = 18.91$, $p < .01$. A One-way ANCOVA was conducted to determine a statistically significant difference between study groups on improved healthy relationship and marriage skills. After controlling for relationship status, employment status, arrest record, and pretest score, the covariate of pretest score was significantly related to the outcome scores and held up to additional testing, $F(1,175) = 9.60$, $p < .01$. After testing for the assumption of homogeneity of regression slope and the interaction between the study group with pretest scores, is found to be insignificant and therefore the assumption is tenable.

See Figure 1 for a comparison of mean scores at follow-up, and Table V.2 for more detailed analysis findings.

Table V.1. Covariates included in impact analyses

Covariate	Description of the covariate
Relationship status	Relationship status (0=no partner, 1=has a partner) as of the baseline data collection
Employment status	Employment status (0=unemployed, 1=employed) as of the baselines data collection
Arrest record	Police record (0=no police record, 1=has a police record) as of the baseline data collection
Pretest mean score	Mean score (0-1) at pretest for the outcome measure of interest at follow-up

Figure 1. Intervention versus Control Mean Scores at 12-month Follow-up (ANCOVA test)



Notes: p-values are included as punctuation for each outcome label. **/*/+ Differences are statistically significant at the .01/.05/.10 levels, respectively.

Table V.2. Post-intervention estimated effects using data from 12-month follow-up to address the primary research questions, controlling for covariates

Outcome measure	Intervention mean or % (standard deviation)	Control mean or % (standard deviation)	Intervention compared to control mean difference (p-value of difference)	Effect size
Progress toward a greater economic stability, including skill attainment and employment	0.77 (0.18)	0.74 (0.12)	0.03 (.29)	0.20
Improved communication & conflict resolution towards partner	0.68 (0.17)	0.66 (0.13)	0.02 (.58)	0.16
Improved healthy relationship and marriage skills	0.52 (0.14)	0.54 (0.14)	-0.02 (.79)	-0.14
Improved understanding of financial planning	0.79 (0.15)	0.69 (0.12)	0.10 (<.01**)	0.74
Sample Size	151	71	n.a.	n.a.

Source: 12-month follow-up Surveys, administered 12 months after the start of the program.

Notes: p-values are included in parentheses. Effect sizes are calculated by dividing the differences in means by the standard deviation of the control group. **/*/+ Differences are statistically significant at the .01/.05/.10 levels, respectively.

n.a. = not applicable

C. Sensitivity analyses

1. Key findings

The sensitivity analyses compared results from the benchmark approach results to analysis with no covariate adjustment. Like the benchmark approach results, the analysis with no covariate adjustment did not produce statistically significant results for the outcomes of progress toward a greater economic stability, improved communication and conflict resolution towards partner, and improved healthy relationship and marriage skills.

Statistically significant results were seen in both methods of analysis for the outcome of improved understanding of financial planning. The benchmark approach using a one-way ANCOVA showed a significant effect from study group on this measure after controlling for relationship status, employment status, arrest record, and pretest score, $F(1, 172) = 18.91$, $p < 0.01$ (see Table V.2.). With no covariate adjustment, there was still a statistically significant difference between group means at follow-up as determined by one-way ANOVA ($F(1, 218) = 22.78$, $p < .01$.) The results between the two methods of analysis are very similar and do not challenge the primary findings. See Table E.1 in Appendix E for the differences in means between intervention and control groups estimated using this alternative method.

VI. DISCUSSION

The purpose of the study was to examine the impact of BUFP on participants' scores for four fatherhood-related outcomes. Results showed that for one of the outcomes, understanding of financial planning, intervention participants scored significantly higher than control participants ($p<0.01$). For the other three outcomes, progress toward a greater economic stability, understanding of healthy marriage and relationship skills, and improved communication/conflict resolution toward partner, results showed no statistically significant differences between intervention participants and control participants.

Why was there a lack of impact on these three outcomes? One reason might be the level of participation. For example, on average, intervention participants attended about 50% of the intervention sessions. In fact, only 28% of the participants received the full 58-hour intervention. It should also be noted that 11% of the intervention participants, for whom data were collected, did not attend any of the intervention sessions, i.e. these 11% received no more of the programming than those in the control group. It should be clear that even a highly effective intervention, cannot produce change among those who do not participate. Moving forward, it will be important for BUFP program staff to seriously consider how the program might increase the degree of participant engagement in the intervention.

The following discussion serves to identify programs that have examined outcomes like the outcomes used in the BUFP evaluation. We compare results and note the contribution the current evaluation makes to the literature.

Healthy Marriage and Relationship Skills

To our knowledge, there are no results in the literature from fatherhood programs concerning the effects of programs on healthy marriage and relationship skills. There are, however, several studies that have examined the effects of relationship enhancement programs as a part of a healthy marriage program. In 2005 Dion highlighted healthy marriage programs and indicated several studies found the programs to have positive effects on relationship satisfaction and communication among romantically involved couples. Hawkins and colleagues (2008) conducted a meta-analysis of such studies and noted similar effects. In both the 2005 publication and the 2008 publication, researchers noted that the studies did not include non-white, disadvantaged populations and were concerned as to whether the results of the studies could be generalized to these other population groups.

Other researchers (Johnson, 2012; Jackson et al, 2016; Karney, Bradbury, & Lavner, 2018) have noted that results of federal healthy marriage initiatives targeting low-income couples and couples of color, have been disappointing because they had no effect or have had negligible benefits. Amato (2014) however, suggested that contrary to the notion that disadvantaged couples do not benefit from relationship education, these couples may be the main beneficiaries of these services, provided they are able to keep their unions intact. More recently Hawkins (2019) provided a positive review of federal evaluations of healthy marriage programs, while

also noting that most scholars disagreed with him. This included a report from the National Academy of Science. He indicated that in his view the reason for others taking a negative view concerning the results of healthy marriage programming was their conclusions were based on "an early and limited view of evaluation work." Hawkins went on to say he found the positive results from a number of programs to be encouraging. For example, he stated that participants liked the programs and thought they were of benefit. Programs also seemed to be more or less effective depending on the type of outcome under consideration. For example, some evaluations have shown small effects in helping distressed, low-income married couple increase their commitment and stay married. There is no evidence, however, indicating unmarried couple participating in these programs are more likely to marry. There is evidence showing program participation can help couples reduce destructive conflict and experience less domestic abuse (both physical and emotional). Additionally, there is growing evidence that programs can improve couples' positive communication skills understanding, warmth, support, and co-parenting, as well as other effects.

Hawkins noted this was a relatively new area of federal funding and, in comparison to other social policy initiatives, that had received greater funding, done less early evaluation work, and shown less evidence of effectiveness, healthy marriage programming was promising and merited continued support. He indicated that the healthy marriage and relationship initiative should fund research giving more attention to long-term results, strive for programming that will produce greater impact, and to move beyond program success to population impact.

Hawkins did not identify the factors that seemed to be responsible for program success, in those programs that have achieved positive results. Also note the results Hawkins discussed were not from fatherhood projects, but men, including fathers, were part of the interventions.

Like the results in the literature, the current evaluation did not find the BUFP to have an impact on healthy marriage and relationship skills. The most highly engaged participants in the Brothers United program indicated that they liked the program and thought it had helped them. These participants indicated they did not want the program to end and wanted to participate in further programming. The question then is, why the program did not achieve positive results in this important area. The most straightforward answer is, we don't know. While the participants indicated they liked the program, their positive view may have had to do more with the fellowship and identification with "BU Nation" than any skills actually learned. Were the curricula that were used, and there were several, appropriate for the participants' cultural background and current situation? Program staff believed the curricula chosen were relevant to the target population, but were they? In the future it might be helpful to conduct focus groups with members of the target population to discuss the curricula and content before implementing them. A pilot study designed to examine participants' level of engagement might be helpful in identifying and addressing factors influencing the degree to which participants actually participate in the intervention sessions. This might allow staff to make adjustments in the curricula, or in the level or type of participant support, to increase participant engagement. Was there a theoretical basis for expecting that implementing these programs would produce positive changes? Was the program implemented with fidelity? For participants who have not seen

models of healthy marriage relationships in their own upbringing, developing such skills themselves may not be an easy task. Future programming should take these factors into account and be confident that the program they propose to implement with their group of participants has real potential to produce the desired changes.

Understanding Financial Planning

Landers (2017) found participants enrolled in fatherhood programs offered in a non-rural context showed no significant change in general financial literacy from pre- to post-test. Fathers who participated in programs provided in rural areas, however, did show significant positive changes in financial literacy. There was also a significant time-by-group interaction effect, showing that there was greater improvement from pretest to posttest for the rural participants than for the non-rural participants. There were no control participants, that is, fathers who completed the pretest and posttest, but did not participate in a fatherhood program.

Unlike the Landers (2017) study, the current evaluation's results show that BUFP made positive changes in intervention fathers' financial literacy compared to control fathers, at the 12-month follow-up. In addition, the current evaluation uses an RCT design and had low attrition, which gives us confidence that the program, in fact, did produce this effect. The intervention, design, and additionally, the 12-month follow-up, instead of a pretest-posttest, no control group, no follow-up design, points to the greater rigor used in this study and enhances the confidence we have in the findings. While the current evaluation did not include participants in rural settings, it may be useful to implement the BUFP with participants in rural settings to see whether BUFP can impact the understanding of financial planning among fathers in other settings.

Progress to Economic Stability

Holmes et al. (2018) conducted a meta-analysis of 30 studies (published and unpublished) concerned with evaluations of fatherhood programs targeting unmarried, never married, and low-income fathers. Although the researchers found some positive effects in other areas, relative to economic stability, the programs did not improve employment for fathers, or their economic well-being. Additionally, the programs did not increase fathers' child support payments. That is their findings were similar to this study, in that that the BUFP did not influence participants' progress toward economic stability. An important exception, though not a part of the RCT, was that BUFP participants were more likely to have made at least one child support payment. Terry and Azmitia (2018) reported that participants in the Father Involvement Support Program demonstrated changes from pretest to posttest indicating they were having fewer challenges in paying their bills. The study design, however, included no long-term follow-up and no control group. There was no change in fathers' employment status.

The results of our study show that an increase in understanding of financial planning does not necessarily translate into progress in economic stability within 12 months. Additionally, when more than 30 studies indicate various fatherhood interventions do not show positive results in employment or in economic welling-being, one can only conclude that the programs that have been provided do not produce the desired results or that more time might be needed to detect

such results. What can produce positive results? Perhaps this needs to be something more than what programs have been doing. Helping participants acquire the education and training necessary to secure a job that pays well, and the time needed to support them in becoming well-established in a career, may simply be beyond the scope of most projects. For a project to actually make a difference in this area, there may need to be a serious investment in the lives of the participants. Projects may need to work with participants to help them first develop an understanding of what might be necessary, on their part, to make a real difference. Second, for those participants who want and potentially can do that which is necessary, projects may need to make a stronger effort to provide the resources and work with individual participants to help them secure the education, training, and skills necessary to secure and maintain jobs that both pay well and provide opportunities for career advancement.

Conflict Resolution with Partner

Avellar and her colleagues (2018) found that, in the four responsible fatherhood programs they examined, the programs did not produce changes in using constructive-conflict behaviors or in avoiding destructive conflict behaviors. Terry and Azmitia (2018) noted fatherhood program participants reported no change from pretest to posttest in "conflict satisfaction" scores.

These results are much like this study's finding that fathers in the BUFP intervention did not score differently from control fathers relative to conflict resolution with their partner. Learning a different pattern of behavior to resolve conflict is not an easy task, and again may be beyond the scope of the programs that have been previously provided. If programs are to make a difference in this important area they may need to do something different than has been done in previous programs. This may include providing more intensive training and relationship counseling with individual participants and couples counseling with participants and the mothers of their children.

Limitations

The study population was comprised of young fathers (ages 20-24), almost all African Americans, living in low-income, high-crime, urban areas. Thus, the results seen in this study may not generalize to other groups. Additionally, data were from self-report questionnaires, with the potential problems that accompany this method of data collection. With self-report data there are always some concerns, and thus these concerns are also applicable, but not unique, to the present study. For example, the improvements shown may be due to participants responding based on perceived social desirability, rather than actual improvements. Additionally, we do not have data from other sources, such as family members and/or third-party observations, to corroborate what participants reported. Nevertheless, despite these limitations, these are important findings from an experimental design study that make a real contribution to the literature, and to professionals working in fatherhood programs.

Conclusions

Overall, the BUFP produced positive results for one out of the four outcome variables: understanding financial planning, indicating that the program is working to support financial

planning among the sample of young African American fathers. It also suggests, that given the positive results BUFP achieved with this sample of fathers, it may be that programs like BUFP can potentially produce these changes with other groups in other parts of the country.

The findings for the three other outcome measures—progress toward economic stability, understanding of healthy marriage and relationship skills, and conflict resolution with partner—were consistent with previous work that found no positive results. Thus, it appears that such fatherhood programs, structured as they currently are, may not work in these areas as long-term solutions to change. To begin to make change in these important areas, researchers may need to examine other possible factors. We have made suggestions as to factors that may be preventing positive results, and suggestions as to programming changes to consider that may yield positive results. Future research/evaluation efforts should take these into account.

VII. REFERENCES

Avellar S, Covington R, Moore Q, Patnaik A, Wu A. Parents and children together: effects of four responsible fatherhood programs for low-income fathers. Report by Mathematica to the Office of Planning, Research, and Evaluation Administration for Children and Families, June, 2018, OPRE Report 2018-50. Available at: https://www.acf.hhs.gov/sites/default/files/opre/pact_rf_impacts_to_opre_508.pdf. Accessed May 19, 2020.

Cohen, J. (1977). Statistical power analysis for the behavioral sciences. Routledge.

Dion, M. R. (2005). Healthy Marriage Programs: Learning What Works. *The Future of Children*, 15(2), 139–156.

Hawkins, A. J., Blanchard, V. L., Baldwin, S. A., & Fawcett, E. B. (2008). Does marriage and relationship education work? A meta-analytic study. *Journal of Consulting and Clinical Psychology*, 76(5), 723–734.

Holmes EK, Egginton BM, Hawkins AJ, Robbins NL, Shafer K. Do responsible fatherhood programs work? A comprehensive met-analytic study. Family Relations, 2018;

Jackson, G. L., Trail, T. E., Kennedy, D. P., Williamson, H. C., Bradbury, T. N., & Karney, B. R. (2016). The salience and severity of relationship problems among low-income couples. *Journal of Family Psychology*, 30(1), 2–11.

Johnson, M. D. (2012). Healthy marriage initiatives: On the need for empiricism in policy implementation. *American Psychologist*, 67(4), 296-308.

Karberg E, Aldoney D, Cabrera N. Fatherhood in America: the context, practice and gaps in responsible fatherhood programs. In Mazza C, Perry AR. Fatherhood in America: social work perspectives on a changing society, Charles C. Thomas Publisher, January, 2017.

Karney, B. R., Bradbury, T. N., & Lavner, J. A. (2018). Supporting healthy relationships in low income couples: Lessons learned and policy implications. *Policy Insights from the Behavioral and Brain Sciences*, 5(1), 33-39.

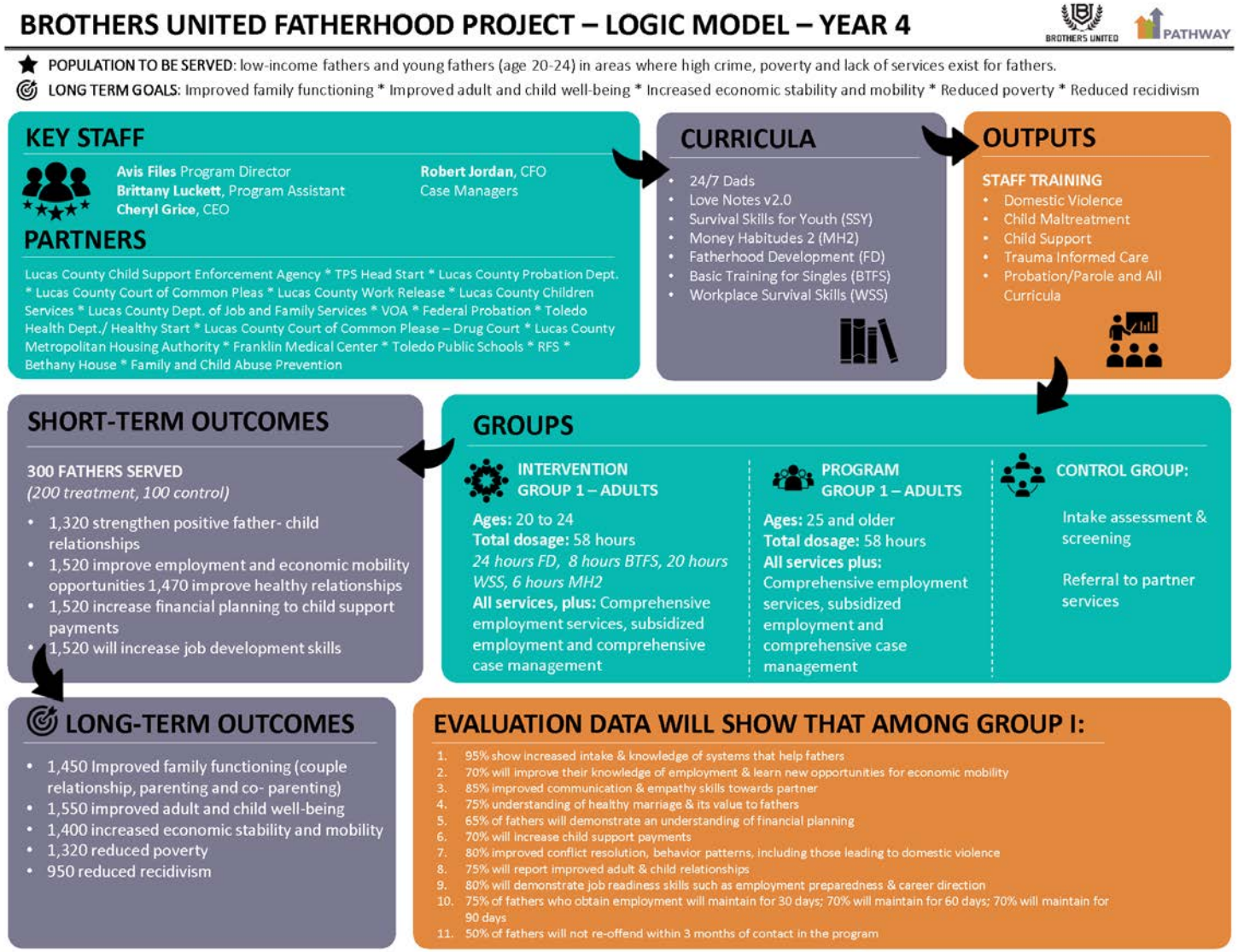
Landers, AG. The Examination of Outcomes of Fatherhood Program Participants in Rural and Non-Rural Communities, Masters Thesis, Auburn University, Educational Foundations, Leadership and Technology, May, 2017. Available at: https://etd.auburn.edu/bitstream/handle/10415/5691/Thesis%20AL%204.17.17%20for%20AUETD.pdf?sequence=2&isAllowed=y Accessed May 19, 2020.

Terry J, Azmitia N. Father involvement support program. Graduate project, California State University – Northridge, Social Work, May, 2018. Available at: http://scholarworks.csun.edu/bitstream/handle/10211.3/203819/Terry-Jasmine-thesis-2018.pdf?sequence=1 Accessed May 19, 2020

Wood RG, Moore Q, Clarkwest A, Killewa A. The long-term effects of building strong families: a program for unmarried parents. Journal of Marriage and the Family, 2014 (March)

VIII. APPENDICES

A. Logic Model (or Theory of Change) for the Program



B. Data and Study Sample

Table B.1. Description of implementation measures

Implementation element	Research question	Measures
Fidelity	Were all intended intervention components offered and for the expected duration?	• Total number of sessions delivered • Average session duration, calculated as the average of the recorded session lengths in nFORM attendance data (in minutes)
Dosage	On average, how many hours of the intervention did the intervention group participate in?	• Average number of hours clients attended • Percentage of the sample attending the required or recommended proportion of hours • Percentage of the sample with no hours recorded (did not attend)
Dosage	What was the average number of sessions the intervention group participated in?	• Average number of sessions clients attended • Percentage of the sample attending the required or recommended number of sessions • Percentage of the sample with no session attendance recorded
Quality	What was the quality of staff–participant interactions?	• Percentage of scores of agreement that the quality was high using the 6-item Internal Feedback Form measuring participants' experience of each curriculum at the close of each cohort • Verbatim responses to the open-ended question on the Internal Feedback Format the close of each cohort;
Engagement	How engaged were intervention group members in the intervention?	• Percentage of participants who complete the program
Context	What external events occurred that affected implementation?	• Responses provided by participants from monthly call notes about external events that may have impacted attendance – responses will be listed, and frequently reported responses will be summarized • Number of sites/schools that were closed as a result of weather events or policy changes (unrelated to the HM/RF programming), if any

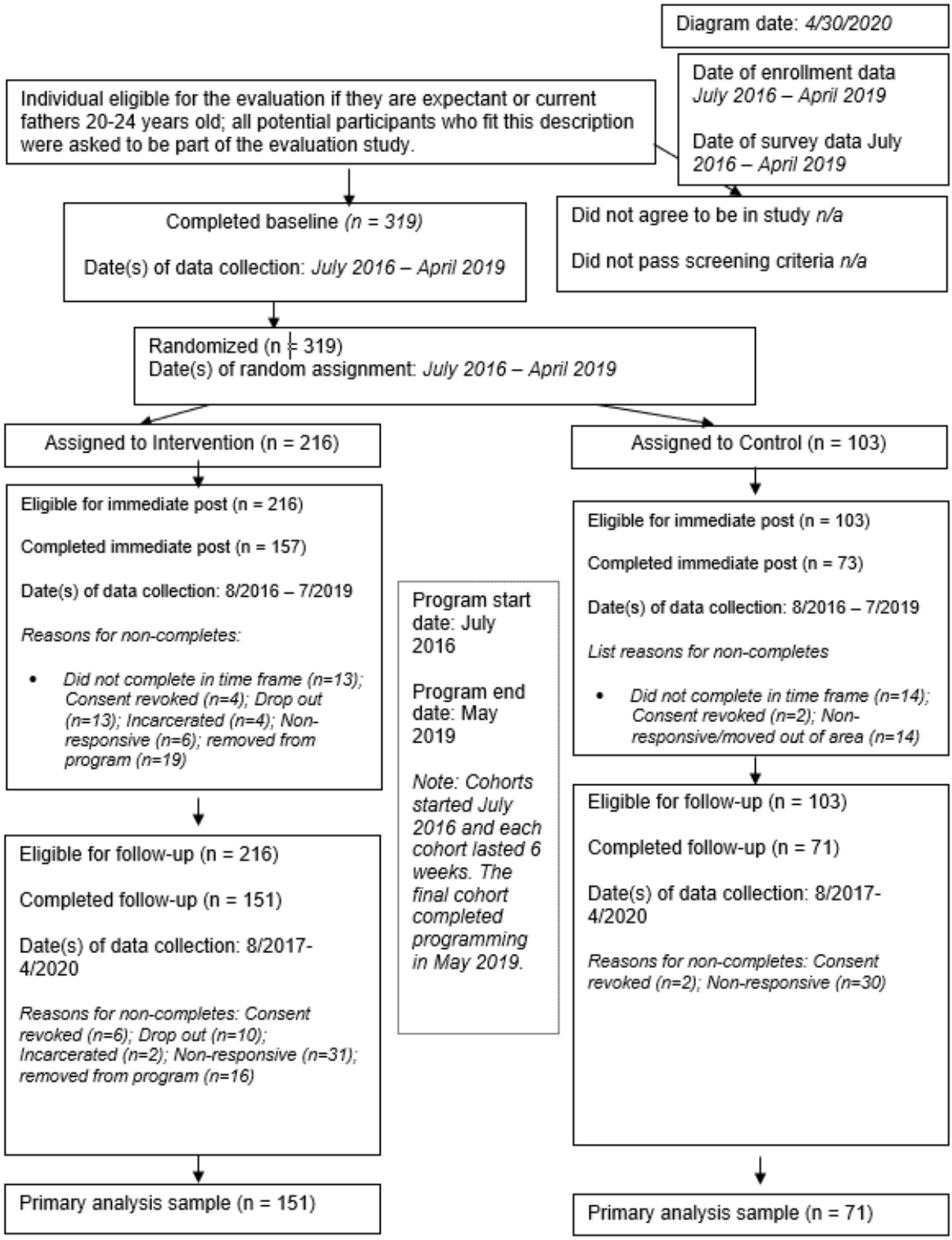
Table B.2. Data used to address implementation research questions

Implementation element	Research question	Data source	Timing/frequency of data collection	Party responsible for data collection
Fidelity	Were all intended intervention sessions offered and for the expected duration?	Attendance data in nFORM	At each of the 24 sessions per cohort	Intervention staff
Fidelity	Were all topics of the BU program covered in the intervention sessions?	Attendance data in nFORM	At each of the 24 sessions per cohort	Intervention staff
Dosage	What sessions did the intervention group receive?	Attendance data in nFORM	At each of the 24 sessions per cohort	Intervention staff
Dosage	How many hours did the intervention group participate in the intervention on average?	Attendance data in nFORM	At each of the 24 sessions per cohort	Intervention staff
Dosage	How many sessions did the intervention group participate in the intervention on average?	Attendance data in nFORM	At each of the 24 sessions per cohort	Intervention staff
Quality	What was the quality of staff–participant interactions?	The BUFP Internal Feedback Forms	At the end of each curriculum for each cohort in Year 5	Intervention Staff
Engagement	How engaged were clients in the intervention?	nFORM attendance data	At each of the 24 sessions per cohort	Intervention Staff
Context	What external events affected implementation?	Program documentation from monthly evaluation call with program staff	At the end of each quarter from 1/2017 to 4/2020	Evaluation team

Table B.3. Key features of the impact analysis data collection

	Data source	Timing of data collection	Mode of data collection	Party responsible for data collection	Start and end date of data collection
Intervention	Intervention group study participants	Enrollment (baseline) End of intervention (6 weeks after enrollment)	In-person online survey (nFORM)	Program staff	July 2016 through April 2019
		Enrollment (baseline) End of intervention (6 weeks after enrollment) Follow-up (1 year after enrollment)	In-person online survey (Qualtrics local evaluation)	Program staff	July 2016 through April 2020
Control	Control group study participants	Enrollment (baseline) End of intervention (6 weeks after enrollment)	In-person online survey (nFORM)	Program staff	July 2016 through April 2019
		Enrollment (baseline) End of intervention (6 weeks after enrollment) Follow-up (1 year after enrollment)	In-person online survey (Qualtrics local evaluation)	Program staff	July 2016 through April 2020

Figure B.1. CONSORT diagram for individual clients



C. Data Preparation

Data were merged from the nFORM and the local evaluation survey. If data were missing for single, stand-alone items crucial to the analysis, adjustments were made to try to recoup those data by cross-checking information in the local evaluation survey and the nFORM entry survey. For example, some missing data for employment status in the local evaluation survey were recouped using employment questions from the nFORM survey. In an analysis that involved a single variable for which data are missing and no similar data can be examined across instruments—for example, a participant leaves the item concerning arrest record blank—list-wise deletion of missing data removed this participant's data for any analysis that involves the variable arrest record.

Missing data concerning items that comprise multi-item scales for baseline and outcome data used an averaging method to impute. Scores for multi-item scales are reported as the average item score. For example, individual items are scored from 0 to 3. For a five-item scale, summing the five items gives a total score somewhere between 0 (if the participant scored each item 0) to 15 (if the participant scored each item 3). Dividing the total score by 5 gives an average item score, which is the score used. Average item scores were between 0 and 1. When data were missing for any item in a multi-item scale, then list-wise deletion of missing data would result in no score shown for any participant for which there were missing data for one or more of the items comprising that scale. To reduce this data loss, the following adjustment was made for the two scales that are comprised of more than three items (Progress toward greater economic stability – 6 items, and Improved communication and conflict resolution toward partner – 4 items). If a participant failed to respond to no more than 20% of the items comprising a scale (for the two scales mentioned, this was no more than 1 item), then the average of the remaining items that comprise the scale was used to determine an average item score. If a participant failed to respond to more than 20% of the items comprising a scale (for the two scales mentioned, this was more than 1 item missing), then no score for that scale was reported for that participant. No other imputations were performed.

Analytic Approach

The analyses of the impact of the intervention were conducted using an intent-to-treat approach. All study participants who were randomly assigned to the study groups (intervention and control) who have completed surveys were included in the analysis, regardless of the amount of intervention services they received or their cross-over status.

The model specification was analysis of co-variance. This approach addressed whether there was a difference at follow-up between the intervention group and the control group for each of the outcome variables. The covariates of employment, relationship status, arrest record at baseline, and pretest score for the outcome variable under consideration were included in the analysis.

The local evaluation study used JASP 0.13.1 for Confirmatory Factory Analysis and analysis of scale reliability. IBM SPSS Statistics 26 to analyze results. Findings were considered statistically significant based on $p < .05$, two-tailed test. Additionally, the study used Hedge's g to measure

effect size. Together, these two procedures allowed the study to first determine whether the interventions produced a difference in the outcome variables, when compared to the control condition, and then establish the size of such differences.

The study assessed, using covariates, whether there is a difference at follow-up between the intervention group and the control group. Sex as a covariate was not included because the study only included males/fathers. The study did not include race/ethnicity as a covariate because most of the sample were Black or African-American. Covariates used in the follow-up analysis included:

- Employment status – Employment status (unemployed, employed) at baseline data collection
- Relationship status – Single vs in a relationship at baseline data collection
- Arrest record – Arrest record at baseline data collection
- Pretest mean score – Mean score (0-1) at pretest for the outcome measure of interest

D. Impact Estimation

The text description of formulas for each estimation approach are detailed below. The description includes the outcome measured, the dependent variable, the covariates, and the independent variable.

A. Outcome 1 Follow-up Score (Progress toward greater economic stability, including skill attainment and employment) ANCOVA

- Outcome 1 Follow-Up Score (dependent variable)
- Outcome 1 Baseline Score + Employment Status + Relationship Status + Arrest Record (covariates)
- Group – Intervention vs. Control (independent variable)

B. Outcome 2 Follow-up Score (Improved communication & conflict resolution towards partner) ANCOVA

- Outcome 2 Follow-Up Score (dependent variable)
- Outcome 2 Baseline Score + Employment Status + Relationship Status + Arrest Record (covariates)
- Group – Intervention vs. Control (independent variable)

C. Outcome 3 Follow-up Score (Improved healthy relationship and marriage skills) ANCOVA

- Outcome 3 Follow-Up Score (dependent variable)
- Outcome 3 Baseline Score + Employment Status + Relationship Status + Arrest Record (covariates)
- Group – Intervention vs. Control (independent variable)

D. Outcome 4 Follow-up Score (Improved understanding of financial planning) ANCOVA

- Outcome 4 Follow-Up Score (dependent variable)
- Outcome 4 Baseline Score + Employment Status + Relationship Status + Arrest Record (covariates)
- Group – Intervention vs. Control (independent variable)

Table D.1. Confirmatory Factor Analysis of Baseline Measures

Model	X^2	df	TLI	CFI	RMSEA	SRMR
Progress toward greater economic stability, including skill attainment and employment	25.27	9	0.87	0.92	0.10	0.06
Improved communication & conflict resolution towards partner	0.08	2	1.01	1.00	0.00	0.00
Improved healthy relationship and marriage skills	14.31	9	0.96	0.98	0.07	0.06
Improved understanding of financial planning	16.61	2	0.63	0.88	0.18	0.06

Notes: X^2 = a chi-square statistic; df = degrees of freedom; TLI = Tucker–Lewis index; CFI = comparative fit index; RMSEA = root mean square error of approximation; SRMR = standardized root mean squared residual fit index.

Table D.2. Confirmatory Factor Analysis of Follow-Up Measures

Model	X^2	df	TLI	CFI	RMSEA	SRMR
Progress toward greater economic stability, including skill attainment and employment	93.77	9	0.87	0.92	0.21	0.04
Improved communication & conflict resolution towards partner	2.75	2	0.99	1.00	0.04	0.02
Improved healthy relationship and marriage skills	14.49	9	0.97	0.98	0.05	0.04
Improved understanding of financial planning	2.56	2	0.99	1.00	0.04	0.02

Notes: X^2 = a chi-square statistic; df = degrees of freedom; TLI = Tucker–Lewis index; CFI = comparative fit index; RMSEA = root mean square error of approximation; SRMR = standardized root mean squared residual fit index.

E. SENSITIVITY ANALYSES AND ALTERNATIVE MODEL SPECIFICATIONS

The text description of each estimation approach for the sensitivity analyses are detailed below. The description includes the outcome measured, the measurement, and the statistical test.

a. Outcome 1 Follow-up Score (mean score, ANOVA)
b. Outcome 2 Follow-up Score (mean score, ANOVA)
c. Outcome 3 Follow-up Score (mean score, ANOVA)
d. Outcome 4 Follow-up Score (mean score, ANOVA)

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

Outcome	Benchmark approach Intervention compared to control mean difference (p-value of difference)	No covariate adjustment Intervention compared to control mean difference (p-value of difference)
Progress toward a greater economic stability, including skill attainment and employment	0.03 (.29)	0.02 (.61)
Improved communication & conflict resolution towards partner	0.02 (.58)	0.01 (.50)
Improved healthy relationship and marriage skills	-0.02 (.79)	0.01 (.70)
Improved understanding of financial planning	0.10 (<.01**)	0.10 (<0.01**)

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

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