Impact Evaluation of Phone Services Added to the R3 Academy in California

Final Impact Evaluation Report for Healthy Relationships California

Prepared by

Kevin Shafer, Ph.D.,

Brigham Young University

Recommended Citation:

Shafer, Kevin. 2020. “Impact Evaluation of Phone Services Added to the R3 Academy in California.” Washington, D.C.: Department of Health & Human Services.

Acknowledgements:

Special thanks to the leadership and staff of Healthy Relationships California without whom this report would not have been possible, especially HRC Research Manager Shelece Luke, HRC President Patty Howell and Vice President K. Jason Krafsky, and HRC’s skilled teams of instructors, phone coaches, case managers, and registration staff.

Disclosure:

The report author has no conflicts of interest to disclose.

This publication was prepared under Grant Number 90FK0108 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 Phone Services Added to the R3 Academy in California"

Objective. This study considers the effects of the R3 Academy and added phone services in the Responsible Fatherhood program provided by Healthy Relationships California across multiple sites in the state of California.

Study design. All fathers participated in the R3 Academy, which offered content on parenting, coparenting, and economic provision. Fathers were randomly assigned to receive added phone services. Fathers received: (1) case management, which included case management around basic needs and referrals to community programs, (2) phone coaching, which included weekly calls with a trained phone coach reviewing the course content provided in class sessions and applying material to their individual life situations, or (3) a weekly class reminder call only. Primary research questions examined the impact on father involvement, coparenting quality, and economic outcome scale scores between pre- and post-test. Secondary research questions focused on describing changes in father involvement and coparenting from posttest to the one-year follow-up survey for fathers who remained in the study sample at posttest.

Results. Using an intent-to-treat framework, results obtained from ordinary least squares regression indicated that all groups experienced an increase across all outcomes. Compared to the control group, individuals who received phone coaching showed larger increases in one component of father involvement (caregiving) and coparenting quality between pre- and post-test. Case management participants reported larger increases in economic provision scores between pre-and post-test than those in the comparison group. Results were robust across various modeling strategies. Individuals who received phone coaching had higher caregiving and coparenting scores at one-year follow-up, as well.

Conclusion. Different add-on strategies may be particularly relevant given program goals. Phone coaching appears to provide additional benefits for relational outcomes, while case management is beneficial for economic outcomes. Participant engagement may be a particularly relevant concern for programs interested in additional services outside the classroom.

Contents

Tables

Figures

Impact Evaluation of Phone Services Added to the R3 Academy in California

I. Introduction

A. Introduction and study overview

Fathers play a vital role in their children's lives. When children have involved fathers, they show a range of positive outcomes including better social skills (Lieberman et al., 1999), develop greater empathy (Koestner et al., 1990), develop greater self-esteem (Lam et al., 2012), and achieve higher grades (The National Center for Education Statistics, 1997). A study analyzing more than 100 studies on parent-child relationships concluded that father love is at least as important as mother love (if not more so) in predicting a number of different child outcomes, including psychological adjustment, conduct problems, cognitive and academic performance, mental illness, and substance abuse (Rohner & Veneziano, 2001).

Responsible Fatherhood education teaches dads the essential skills they need to parent their children. Beyond parenting, many fathers also need assistance in maintaining co-parenting relationships with their children's mother, with establishing healthy romantic relationships, and obtaining employment and the resources they need to provide for their families. The R3 Academy program is a 24-hour Responsible Fatherhood workshop series designed to meet these needs. It includes the *Raising World Class Kids*™ course, which teaches parenting skills as well as co-parenting relationship skills, and the *Jobs and Money (JAM) Session* course, which teaches money management and job search skills. The R3 Academy program was offered across the state of California, specifically in Los Angeles, Sacramento, Alameda, and San Diego counties.

Compared to the many studies available on Healthy Marriages programs, little research has been done on the effect of Responsible Fatherhood programs to date (Holmes et al., 2020). Existing studies suggest Responsible Fatherhood programs can benefit participants. A recent meta-analysis on fatherhood programs for unmarried, nonresident, low-income fathers found small but statistically significant effects on father involvement, parenting, and coparenting, although no significant impacts were found for father employment, economic well-being, or payment of child support (Holmes et al., 2020). Avellar et al. (2018) examined the effects of four responsible fatherhood programs on low-income fathers, and found that programs improved fathers' nurturing behavior and engagement with their children, although it did not impact co-parenting, contact with the children, financial support for the children, fathers' earnings, or fathers' social-emotional well-being.

While little is known about the mechanisms within Responsible Fatherhood programs that may lead to change, research suggests that persistent contact with fathers may be an effective engagement strategy (Cullen, Cullen, Band, Davis, & Lindsay, 2011; Ghate, Shaw, & Hazel, 2000). In order to increase the impact of the R3 Academy, we sought to understand the benefit of adding a one-on-one component to the workshop. To better reinforce skills taught in class, one third of our clients were assigned a Phone Coach, who called them weekly to practice the skills they were learning in the workshop. To help better meet the economical and physical needs of clients, one third were assigned a weekly call with a Case Manager who assessed their needs and helped them access local resources available. The remaining one third of clients participated in the R3 Academy alone.

Retention and engagement are known challenges in the field of Responsible Fatherhood (Zaveri et al. 2015). We anticipated that the addition of phone services and case management would help improve program retention and promote a deeper understanding of the curricula for the men enrolled. Understanding whether these are successful strategies could be a valuable contribution to the field of Responsible Fatherhood.

This report will first present the primary and secondary research questions for the study. The section following will provide a description of the intervention and counterfactual conditions, after which we will discuss the overall study design and the analyses used. Results will then be presented and finally a discussion of the implications of the results will be given.

B. Primary research question(s)

The study conducted by Healthy Relationships California (HRC), as part of their New Pathways to Fathers and Families Grant project (2015-2020), sought to understand whether the addition of a weekly phone call with participants, either offering skills coaching or case management, resulted in improved father involvement, co-parenting, and economic outcomes at posttest when compared to receiving just the fatherhood workshop alone. More specifically, our primary research questions were as follows:

1. What is the impact of weekly Phone Coaching calls relative to the control on father involvement, as measured by two subscales for caregiving and support, at posttest?
2. What is the impact of weekly Case Management calls relative to the control on father involvement, as measured by two subscales for caregiving and support, at posttest?
3. What is the impact of weekly Phone Coaching calls relative to the control on co-parenting at posttest?
4. What is the impact of weekly Case Management calls relative to the control on co-parenting at posttest?
5. What is the impact of weekly Phone Coaching calls relative to the control on economic outcomes at posttest?
6. What is the impact of weekly Case Management calls relative to the control on economic outcomes at posttest?

C. Secondary research question(s)

Beyond our primary research questions, we also sought to understand more about the effects of the R3 Academy as a whole, as well as the longer-term outcomes of phone coaching and case management one year after the intervention was completed. Our secondary research questions were as follows:

1. Do fathers who attended the R3 Academy experience statistically significant changes in parenting outcomes between pre- and post-test?
2. Do fathers who attended the R3 Academy experience statistically significant changes in economic outcomes between pre- and post-test?
3. Do fathers who attended the R3 Academy experience statistically significant changes in co-parenting outcomes between pre- and post-test?
4. Do fathers who received Phone Coaching calls and remained in the sample at post-test experience statistically significant changes on father involvement between the post-test and the one-year follow-up compared to the control?

5. Do fathers who received Case Management calls and remained in the sample at post-test experience statistically significant changes on father involvement between the post-test and the one-year follow-up compared to the control?
6. Do fathers who received Phone Coaching calls and remained in the sample at post-test experience statistically significant changes on the relationship between co-parents between the post-test and the one-year follow-up compared to the control?
7. Do fathers who received Case Management calls and remained in the sample at post-test experience statistically significant changes on the relationship between co-parents between the post-test and the one-year follow-up compared to the control?
8. Do fathers who received Phone Coaching calls and remained in the sample at post-test experience statistically significant changes on economic outcomes between the post-test and the one-year follow-up compared to the control?
9. Do fathers who received Case Management calls and remained in the sample at post-test experience statistically significant changes on economic outcomes between the post-test and the one-year follow-up compared to the control?

II. Intervention and counterfactual conditions

All participants in the study attended the 24-hour R3 Academy workshop series, discussed in more detail below. One third of participants were randomly selected to the counterfactual condition which consisted of the R3 Academy only. One third were randomly assigned to participate in weekly calls with a case manager, while the final one third participated in weekly calls with a communication skills coach. The following section details the intervention and counterfactual conditions as intended; more specifically, the components, content, dosage, and delivery intended (see Table II.2 for a summary of this information). Also discussed is HRC's staff training in order to administer the intervention and counterfactual conditions. Finally, the research questions asked for the implementation study of the program are presented.

A. Description of program as intended

All clients (both in the intervention conditions and counterfactual conditions) participated in the R3 Academy, a 24-hour workshop series that includes first the 16-hour *Raising World Class Kids*™ course (focused on parenting skills as well as co-parenting relationship skills) and then the 6-hour *Jobs and Money (JAM) Session* course (which focuses on money management and job search skills). Besides the 22 curricula hours, two hours of the workshop were devoted to data collection. The R3 Academy was offered in two formats—either 12 two-hour sessions that occurred once a week, or 6 four-hour sessions that occurred once a week. Twelve week classes covered one session a week, while six week classes covered two sessions. (For topics covered in R3 Academy class sessions see Appendix A.)

All clients received a weekly reminder text and email about the date of the next class. Clients in the intervention conditions also received two reminder calls to attend the class, one prior to the first class session, and one prior to the final class session.

The intervention focused on the addition of weekly phone services to the R3 Academy.

One third of participants were assigned to the Case Management condition. These participants received a weekly call from an assigned Case Manager. Case Management services provided included:

- Assessment of needs, to be completed by the participant identifying and selecting target areas of interest/concern;
- Participation in Case Management planning;
- Referral by Case Manager to appropriate community systems;
- Information and assistance navigating community systems;
- Monitoring of access to and success of referrals, with re-evaluation of needs and re-referral as indicated.

One third of clients were assigned to the Phone Coaching condition. These clients received a weekly call from an assigned Phone Coach, who reviewed with them the skills taught in the class and engaged in skills practice with them. Each trained Phone Coach worked from an established protocol that focused on the communication and problem-solving skills learned in the R3 Academy, and helped participants take the skills from the classroom and apply them to their daily lives. The Phone Coaches' protocol was synced with class content, so each week, they reviewed the skills taught in the previous class session. Figure II.1 provides information on which groups received which elements of the intervention. Table II.1 shows the topics covered during phone coaching calls.

Figure II.1. Components Delivered to Randomized Participants

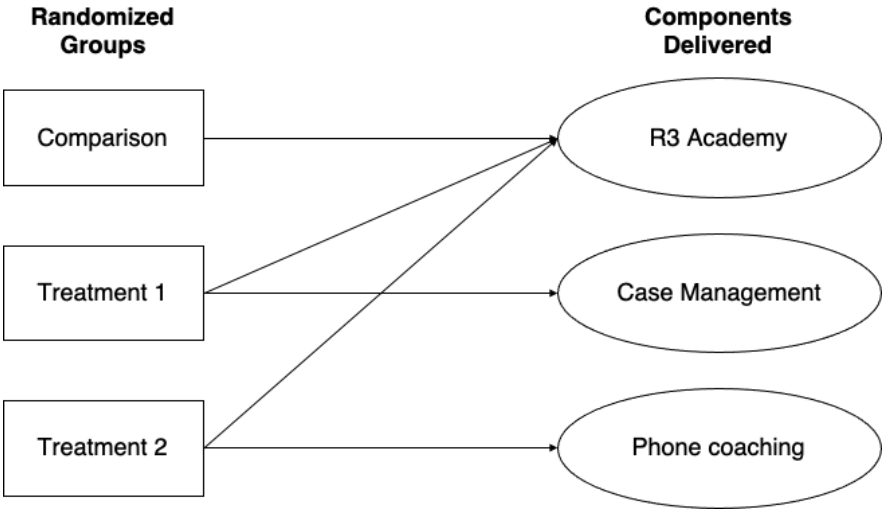


Table II.1. Topics Covered in R3 Academy Phone Coaching

Session Number	Curriculum	Topics Covered
1	None	No call this week
2	Raising World Class Kids (RWCK)	Introduction; parenting styles
3	RWCK	Power listening; method of choice
4	RWCK	Who owns the problem
5	RWCK	XYZ messages; confrontation cycle
6	RWCK	Conflict management tips; 6 steps to a win-win
7	RWCK	Values collisions
8	RWCK	Responsibility; discipline vs. punishment; natural consequences; logical consequences

Session Number	Curriculum	Topics Covered
9	RWCK	Going back and cleaning it up
10	Jobs and Money Session (JAM): Making Money Work (MMW)	Financial literacy defined; 6 steps to a financial recovery
11	JAM: Job Search Success (JSS)	Identifying your skills; developing a professional first impression; skills worksheet; developing a resume; prepare and nail job interview
12	JAM: JSS	No call this week

HRC Staff assigned as case managers served only in that role and not as phone coaches. Likewise, HRC phone coaches never served as case managers. This was done to prevent any accidental crossover from serving in both roles. Some phone coaches also worked as R3 Academy instructors, but instructors never provided phone coaching to fathers in the classes they taught—only to fathers attending other instructors' classes.

B. Intended Content:

The R3 Academy includes the 16-hour *Raising Worlds Class Kids*™ course as well as the 6-hour *Jobs and Money (JAM) Session* course. For these curricula, all clients received a *Raising World Class Kids* workbook, an *AP 4 Parent's Guide*, a *Communication Skills for a World Class Relationship* booklet, and a *JAM Session* workbook.

Phone Coaches followed a Phone Coaching playbook that outlined which skills to discuss for each week of the course. The skills covered included:

- Parenting Skills and Power Listening Lite
- Full Power Listening and Method of Choice
- Who Owns the Problem?
- XYZ-Message and the Confrontation Cycle
- 6 Steps to a Win-Win
- 5 Options for Values Collisions
- Logical Consequences and Going Back and Cleaning It Up
- Financial Literacy and Job Search Success

C. Planned Dosage and Implementation Schedule:

The intervention conditions took place on a weekly basis. For 6-week workshops, clients received 5 calls, beginning after the first class and ending one week prior to the last class session.

For the 12-week workshops, clients received 10 calls, beginning after the second class, and ending one week prior to the last class session. Case Management calls typically lasted 15-30 minutes. Phone coaching calls lasted 30-45 minutes for the 12-week format, and 45-60 minutes for the 6-week format.

Both the 6-week and 12-week workshop formats were made available to HRC instructors so they could select a format that would best fit the needs and schedules of fathers in their area. Sites located in San Diego used the 6-week workshop format, while all other sites used the 12-week format. The implementation study examined whether differences existed between these formats (see page 29).

D. Intended Delivery:

Participants were called weekly by their assigned Case Managers and Phone Coaches. Case Managers and Phone Coaches arranged for calls to take place at a time that fit into participants' schedules. Participants needed access to a phone for these calls, and all participants had access to a phone, either a cell phone or a landline.

R3 Academy clients attended the workshop series at 56 sites across Alameda, Los Angeles, Sacramento, San Bernardino, San Diego, San Joaquin, and Yolo counties in California. These sites included churches, schools, family resource centers, and family education offices.

E. Target Population:

The program is intended to be delivered to fathers or father figures with children under the age of 21. All men who fit these criteria were allowed to participate.

F. Education and training of staff:

All Case Managers and Phone Coaches had a bachelor's degree or a minimum of 2 years of experience in case management, counseling, training, or communication. Most of the staff had worked with Healthy Relationships California in the past.

All staff participated in multiple training sessions prior to calling clients, including reviewing scripts, role playing, and learning to use nFORM (the grant required data management website). Table II.3 summarizes staff training provided. Staff had daily access to supervisors as well as weekly meetings with the supervisor and their fellow team members. Meetings included reviewing cases, assessing needs, and additional training as identified.

Table II.2. Description of intended intervention and counterfactual components and target populations

Component	Curriculum and content	Dosage and schedule	Delivery	Target Population
Intervention				
Relationship skills and parenting workshop	Raising World Class Kids™ (RWCK) curriculum; Parenting Skills; Child Development; Communication Skills, Relationship Skills with Co-parent	Offered in two 16-hour formats: (1) eight 2-hour sessions occurring once a week, or (2) four 4-hour sessions occurring once a week.	Workshops provided by HRC's team of experienced instructors in 56 sites located in the counties of Alameda, Los Angeles, Sacramento, San Bernardino, San Diego, San Joaquin, and Yolo	Fathers or father figures with children under age 25
Economic stability workshop	Jobs and Money (JAM) Session curriculum; Skills-focused Resume Development; Job Search Strategies; Job Interview Skills; Examining Spending Habits; Prioritizing Needs vs. Wants; Creating a Spending Plan; Debt-reduction Strategies	Offered in two formats: (1) 6 hours, taught as 3 two- hour sessions occurring once a week, or (2) 1 two- hour session followed by a four-hour session the next week.	Workshops provided by HRC's team of experienced instructors	Fathers or father figures

Component	Curriculum and content	Dosage and schedule	Delivery	Target Population
Weekly Case Management Phone Calls	Assessment of case management needs; Referrals if needed; Follow-ups on referrals given	Offered in two formats: (1) For 12-week classes, 10 weekly calls beginning after the second week (15-30 minutes); (2) for 6-week classes, 5 weekly calls beginning after the first week (15-30 minutes).	Calls made by HRC's trained Case Management staff.	Fathers or father figures randomly assigned to the Case Management condition
Weekly Skills Coaching Phone Calls	Reviewing skills taught in class in the RWCK and JAM Session curricula	Offered in two formats: (1) For 12-week classes, 10 weekly calls beginning after the second week (30-45 minutes); (2) for 6-week classes, 5 weekly calls beginning after the first week (45-60 minutes).	Calls made by HRC's trained Phone Coaching staff.	Fathers or father figures randomly assigned to the Phone Coaching condition
Reminder Calls	Reminder to attend the next class session	Received one call prior to the first class session, and one call prior to the last class session.	Calls made by HRC's trained Registration Team.	Fathers or father figures randomly assigned to the Intervention Conditions
Counterfactual				
Relationship skills and parenting workshop	Raising World Class Kids™ curriculum; Parenting Skills; Child Development; Communication Skills, Relationship Skills with Co-parent	Offered in two 16-hour formats: (1) eight 2-hour sessions occurring once a week, or (2) four 4-hour sessions occurring once a week.	Workshops provided by HRC's team of experienced instructors	Fathers or father figures
Economic stability workshop	Jobs and Money (JAM) Session curriculum; Skills-focused Resume Development; Job Search Strategies; Job Interview Skills; Examining Spending Habits; Prioritizing Needs vs. Wants; Creating a Spending Plan; Debt-reduction Strategies	Offered in two formats: (1) 6 hours, taught as 3 two-hour sessions occurring once a week, or (2) 1 two-hour session followed by a four hour session the next week.	Workshops provided by HRC's team of experienced instructors	Fathers or father figures
Weekly Reminder Calls	Reminder to attend the next class session	Received a call once weekly.	Calls made by HRC's trained Registration Team members.	Fathers or father figures randomly assigned to Control Group

Table II.3. Staff training and development to support intervention and counterfactual components

Component	Education and initial training of staff	Ongoing training of staff
Intervention		
Weekly Case Management Calls	Case Managers are male and female and hold at least a bachelor's degree or a minimum of 2 years of experience in case management or counseling. Most Case Managers are bilingual in English and Spanish. Case Managers participated in multiple training sessions including reviewing case management protocols and community resources and roleplaying prior to working with clients.	Case Managers meet weekly with their Supervisor to review cases, assess needs, and participate in additional training as the Supervisor identifies. They also participate in an annual refresher of their nFORM training.
Weekly Phone Coaching Calls	Phone coaches are male and female and hold at least a bachelor's degree or a minimum of 2 years of experience in case management, counseling, training, or communication. Most Phone Coaches are bilingual in English and Spanish. Phone Coaches participated in multiple training sessions including participation in RWCK and JAM Session workshops, reviewing scripts and roleplaying prior to working with clients. Phone Coaches were also trained in rating clients' comprehension and cooperation during the calls.	Phone Coaches meet weekly with their Supervisor to review cases, assess needs, and participate in additional training as the Supervisor identifies. They also participate in an annual refresher of their nFORM training.
Counterfactual		
Weekly Reminder Calls	HRC's Registration Team members are male and female. All but one member is bilingual in English and Spanish. They have been trained in nFORM and professional phone etiquette.	Team members receive an annual refresher of their nFORM training.
All Conditions		
R3 Academy Workshop Sessions	HRC's Facilitators are male and female. but one is bilingual in English and Spanish. Facilitators were required to attend the R3 Academy curricula classes before they could be trained as instructors. Curricula training lasted at least 16 hours and included training and practice in all skills taught, as well as practice teaching with critique offered by the curricula developer and fellow trainees.	All Facilitators meet monthly with HRC's Senior Staff to evaluate program delivery and undergo training on specific issues arising in their work in the field. Facilitators receive an annual refresher of their nFORM training.

B. Description of counterfactual condition as intended

A. Intended Components

One third of clients were assigned to the counterfactual condition. Clients in this condition attended the R3 Academy workshop. Like all clients, they received a weekly text and email to remind them of the next class session. In addition, clients in the counterfactual condition received a weekly reminder call reminding them of the date and time of the next class session. Counterfactual clients did not receive Case Management or Phone Coaching calls.

B. Intended Dosage:

Clients received a weekly reminder call the day before their next class session, beginning prior to the first class—6 calls for those in the 6-week workshops, and 12 calls for those in the 12-week workshops.

C. Intended Content:

As with the intervention groups, the counterfactual group participated in the 16-hour *Raising World Class Kids*™ course and the 6-hour *JAM Session* course. Reminder calls were brief and only contained information about the date, time, and location of the next class session.

D. Intended Delivery:

Members of HRC's Registration Team made the weekly reminder calls. All were trained in phone etiquette and protocol.

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

This section presents the research questions we examined to understand the implementation of weekly phone call services for the R3 Academy in California. These additional implementation questions sought to describe whether intervention participants received the expected number of phone calls, whether Phone Coaching clients received the expected course content in their calls, whether clients engaged with the additional phone services offered, and whether class format (6 or 12-weeks) affected phone call participation. Understanding these issues sheds light on the impact study results, as a lack of participation in or engagement with Phone Coaching and Case Management could affect the intervention's impact. If Phone Coaching clients did not receive the expected content, they would not experience the expected benefits. Further, understanding how class format affected intervention participation provides an idea of what class format might be preferable in the future.

Table III. Research questions for Implementation Study

Implementation element	Research question
Intervention Group Questions	
Fidelity	Did the Phone Coaching clients receive the course content expected?
Dosage	How many phone calls did the Phone Coaching clients participate in on average? How many phone calls did the Case Management clients participate in on average?
Engagement	How willing were Phone Coaching clients to engage in their weekly calls? Did Phone Coaching clients comprehend the skills taught? How many Case Management clients received referrals? How many referrals were followed up on?
Context	Was there a difference in Phone Coaching participation for the 6-week classes versus the 12-week classes? Was there a difference in Case Management participation for the 6 week classes versus the 12-week classes?
Control / Comparison Group Questions	
Fidelity	Did all control group members receive their expected weekly reminder calls?

III. Study design

In this section, we will first describe how the sample was formed and the research design for our study. Next, we will discuss the data collection process for both the implementation and impact analysis studies.

A. Sample formation and research design

Sample formation

Fathers and father figures were recruited to the sample from across California in the counties of Alameda, Los Angeles, Sacramento, San Bernardino, San Diego, San Joaquin, and Yolo. Workshops were held at multiple sites in each county. In total, workshops took place in 56 sites, including 34 churches, 10 schools, 3 family education offices, and 9 additional locations (including community centers, family resource centers, retail locations, and private residences). Instructors recruited participants from a variety of organizations including churches, schools, child support agencies, employment assistance centers, TANF offices, WIC agencies, Head Start, Healthy Start, child welfare agencies, and family resource centers.

To be included in the sample, participants needed to be fathers or father figures age 18 and over with children under the age of 21 years old. There were no additional requirements. All program materials were available in English and Spanish.

Research Design and Random assignment process

This study was a randomized controlled trial. The unit of random assignment was individual clients. The HRC Research Manager randomly assigned clients to the intervention conditions or the counterfactual condition. The timing of the random assignment changed on May 1, 2018. Originally, HRC planned to randomly assign clients after they attended their first class session. However, in the beginning of the grant period, this was not possible. A requirement of the New Pathways for Fathers and Families grant was that grantees use the Information, Family Outcomes, Reporting, and Management (nFORM) website. Client information such as random assignment was recorded in nFORM, as well as demographic data, contact info, and some pretest and post test data. Initially nFORM was designed so that the local evaluation assignment was locked in the record permanently after the client took the Entrance Survey. To accommodate this, initially all clients enrolled in the R3 Academy were randomly assigned the day before classes began so the assignment could be recorded prior to the Entrance Survey. As consent and baseline data collection both occurred at the first class, random assignment happened prior to both. Unfortunately, some enrolled clients never attended, which affected baseline response rates.

Later in the grant period, nFORM was altered to allow local evaluation assignments to be entered at any time. When this became possible, timing of the random assignment was changed to the day after the first class, and only clients who attended the course were randomly assigned. Clients were allowed to join the class at the second session if they missed the first. The day after the second class, random assignment was conducted for any clients who had joined at the second session.

As described above, the HRC Research Manager randomly assigned clients to the intervention conditions or the counterfactual condition. Random assignment was conducted with the randomization.com website. This website generates a list of numbered random assignments roughly equally distributed between the three groups. The generated list was then used to assign clients in the order they were listed on the classroom roster.

Randomization was conducted separately for each R3 Academy cohort. Sixty-eight cohorts participated in the study. Clients were intended to have a 1/3 chance of being in the phone coaching group and a 1/3 chance of being in the case management group. Ultimately, there were 448 clients assigned to the case management condition, 440 assigned to the phone coaching group, and 431 assigned to the control group. Thus, random assignment appears to have worked as intended since the groups' sizes are approximately even.

Consent process

The Brigham Young University Institutional Review Board approved the study design and data collection plans on 05/13/2016. Supplemental review approvals occurred on 12/15/2017 for phone interviews and for continuation of the project in May of 2017, 2018, and 2019.

Participants signed consent forms immediately prior to completing the nFORM Entrance survey and local evaluation Qualtrics pretest survey at the first class session. During the length of the study, only 7 men declined to participate.

There were no differences between groups regarding when the consent process was offered. The consent form itself was available in English and Spanish. All participants were offered the same incentives if they chose to participate—a $50 Kohls card for attending the final class session, a $50 Kohls card for attending at least 2/3 of total class sessions, and a chance at winning a prize upon completion of the one-year follow-up survey.

Clients were not notified of their random assignment until after the first class (for 6-week sessions) and after the second class (for 12-week sessions). There may have been spillover effects if participants talked about additional services or made in-class comments that were reflective of these additional services. Such spillover, however, cannot be measured in our data. Clients assigned to the intervention conditions received their first calls from Case Managers and Phone Coaches at this time, who informed them of their random assignment. Control group members were not told explicitly that they were in the control group. As discussed previously in the section on random assignment, at the beginning of the grant project, clients were randomly assigned prior to giving consent to the study.

B. Data collection

In this section, we will discuss our data collection methods, beginning first with describing the data collection for the implementation evaluation and then for the impact evaluation.

1. Implementation analysis

Below is a description of the data sources used to address the implementation study research questions. To see this information captured in a table, please see the Appendix Table C.1.

Fidelity

Phone Coaching fidelity (or whether Phone Coaching clients received the program content expected on their calls) was measured using the Phone Coach fidelity check surveys recorded in Qualtrics. This survey asks Phone Coaches to select which curricula session(s) they discussed during their phone call. For each curricula session, Phone Coaches are presented with a list of all topics that should be covered, and they

mark which ones were discussed with their client. Phone Coaches were asked to complete this measure within 24 hours of each call.

Dosage

For both the Phone Coaching and Case Management groups, dosage was examined by reviewing the individual service contacts recorded in nFORM. Phone Coaches and Case Managers were asked to report every call in nFORM immediately, if possible, and no later than 24 hours after each call. All Phone Coaches and Case Managers were trained in service contact recording prior to making their first calls. Individual service contact records included the following information:

- Date of contact
- Phone Coach or Case Manager's name
- Contact Method (phone)
- Length of contact
- Whether participant answered the phone
- Purpose of Call (Phone Coaching or Case Management)

Case management also included this information:

- Client Needs and Issues Discussed
- Referrals Made
- Did the client follow through on the referral?

Contacts were not counted in the analyses if: (1) the participant did not answer the phone or (2) if no needs or issues were discussed.

Engagement

Phone Coaching engagement data came from the Phone Coach fidelity check survey. The Phone Coach fidelity check survey asked phone coaches to report on clients' level of comprehension and cooperation for every topic discussed (see Appendix D for more information on rating). Comprehension addressed whether clients understood the topic discussed and was measured as minimal, moderate, or high. Cooperation measured whether clients were willing to engage in the call, and was measured as noncompliant, resistant, guarded, or engaged. The Phone Team's supervisor trained Phone Coaches in how to rate clients appropriately. The fidelity check survey was completed within 24 hours of every call.

Case Management engagement data came from nFORM, and focused on the number of referrals given, and whether clients followed up on their referrals. Referrals were recorded within 24 hours of every call. nFORM flagged referrals to remind case managers to follow up with clients at their next call and to record whether or not referrals were acted upon. If clients had not acted on the referral, case managers continued to follow up on future calls.

Context

The evaluation examined if there was a difference in intervention dosage, measured as percentage of respondents that participated in phone coaching or case management within assignment category, for the

6-week classes versus the 12-week classes. Data came from individual service contacts in nFORM as well as nFORM's data on the length of each session series.

Control/Comparison Group

Control group fidelity was measured by examining whether the control group received their expected number of reminder calls. This data came from the individual service contacts in nFORM, which were recorded by HRC's Registration Team after they made the reminder calls. All members of the team were trained in recording these contact records in the nFORM system. Individual service contact records for reminder contacts included:

- Date of contact
- Registration Team member's name
- Contact Method (phone)
- Length of contact
- Whether participant answered the phone
- Purpose of Call (Reminder)

2. Impact analysis

Data was collected from the clients at enrollment, the first class, the last class, and one year after the last class:

- Basic demographic information on the clients was collected with the Applicant Characteristics Survey in nFORM. After clients requested to join an R3 Academy, HRC's Registration Team called them and administered the survey over the phone.
- Baseline data was collected during the first class session. Clients completed two surveys online on Chromebooks, first the nFORM Entrance Survey, and then the Local Evaluation pretest (administered as an online Qualtrics survey). In rare cases of technology failure, both surveys were administered on paper. Clients who missed the first class session were still permitted to participate in the R3 Academy if they arrived at the second class session with a completed paper survey. Directors mailed blank surveys to everyone who missed the first class. Brigham Young University (BYU) research assistants were responsible for inputting survey data collected on paper into nFORM.
- At the end of the final class session, clients again completed surveys online on Chromebooks, including both the nFORM Exit Survey, and the Local Evaluation posttest questions through Qualtrics. Again, in rare cases of technology failure, the surveys were administered on paper. If clients missed the final class session, BYU research assistants called them to complete the surveys by phone. Unfortunately, our data on this issue is limited and we cannot analyze the accuracy of data collected through different modalities.
- One year after the final class session, clients were invited to participate in a one-year follow-up Qualtrics survey. Initially, clients were invited to attend a one-year reunion event where the data was collected on Chromebooks. However, the reunions were subject to poor attendance. Instead, we shifted to trained research assistants at BYU calling clients at the one-year mark to collect surveys over the phone. This also had a low success rate. In December 2018, reunions were reinstituted with greater efforts made to invite clients to attend. Clients who did not attend were contacted by BYU

research assistants and HRC staff members to complete the survey by phone. From March 2020 to June 2020, data collection shifted completely to phone calls again during the COVID-19 pandemic. At this time, clients were also sent personalized links to complete the one-year survey online.

Surveys were collected the same way for intervention groups and the counterfactual group at each time point. See Appendix Table C.2 for data collection instruments.

IV. Analysis methods

This section describes the construction of the sample used for analysis, the outcome measures, and the baseline equivalence of the treatment and comparison/control groups. Table IV.1a. provides the samples by intervention status, while Figure C.1. in Appendix C provides the CONSORT diagram for each stage of data collection.

A .Analytic sample

In total 1,323 individuals were randomized and 194 individuals were found ineligible: 179 individuals who identified as female and 15 with children over 21. Of the randomized individuals, 888 were assigned to a treatment group (448 to case management, 440 to phone coaching) and 431 were assigned to the comparison. Four were assigned to 360 Services, which included both case management and coaching. This treatment group was discontinued starting January 26, 2017 and these data are not included in the analysis. Of the 1,319 respondents eligible for inclusion in the data, 1,240 consented to data collection (94.0%). Of this group, 1,009 contributed a baseline survey: 336 in case management, 339 in phone coaching, and 337 in the comparison group. Those that did not contribute a baseline survey did not attend the first class and could not be reached via telephone, text, or email for survey completion. Because of low attrition, only respondents that provided data on all outcome measures were included. Not all respondents provided complete information for each outcome variable. At baseline, 998 respondents completed the economic provisions scale, 996 completed the father involvement scale, and 978 completed the coparenting scale.

At post-test, which took place immediately upon conclusion of the R3 Academy, 759 respondents provided data on all outcome measures, while 19 respondents did not. We dropped these 19 respondents because of low attrition rates. A second follow-up took place approximately one year after the respondent completed the R3 Academy. All individuals who consented to participate were eligible for the one-year follow-up survey. At one-year follow-up, 658 respondents provided data for at least one outcome, while 653 provided data for all outcomes. We dropped the five respondents that did not contribute to all outcomes.

Differences in completion rates across groups were small. We used the cautious boundary (CB) and optimistic boundary (OB) of the What Works Clearinghouse (WWC) standards to assess whether the RCT could be classified as a low or high attrition study. Looking at Table IV.1a, which reports case management-comparison group attrition, at post-test, the overall attrition was 40.4% and the differential attrition was 1.4%, which indicates that the study is a low-attrition RCT under the CB. At one-year follow up, overall attrition was 49.5% and the differential attrition was 1.2%, which is right at the CB for a low attrition RCT, and below the OB for a low attrition RCT.

Table IV.1b. reports attrition for phone coaching and comparison. At post-test, the overall attrition rate was 41%, with a differential attrition rate of 2.4%, which means the study is a low attrition RCT under the

CB. Finally, the overall attrition at one-year follow up was 50.3% and the differential attrition was 3.4%, which is above the CB but below the OB for a low attrition RCT. Overall, we assess that we have low attrition RCTs using these standards for all outcomes, across all time points.

Analysis of the service provision data on case managements and phone coaching indicated that no individuals assigned to control received any portion of the intervention. Moreover, no individuals assigned to a treatment condition that received the other treatment condition were identified in the nFORM service contacts data.

Table IV.1a. Individual sample sizes by intervention status (Case Management vs. Comparison)

Number of individuals	Intervention sample size	Comparison sample size	Total sample size	Total response rate	Intervention response rate	Comparison response rate
Assigned to case management	448	431	879	n.a.	n.a.	n.a.
Contributed a baseline survey	336	337	673	76.6%	75.0%	78.1%
Contributed to first follow-up survey (post-test at end of R3 Academy)	264	260	524	59.6%	58.9%	60.3%
Contributed to second follow-up (one-year follow-up)	220	224	444	50.5%	49.1%	51.9%

n.a. = not applicable.

Table IV.1b. Individual sample sizes by intervention status (Phone Coaching vs. Comparison)

Number of individuals	Intervention sample size	Comparison sample size	Total sample size	Total response rate	Intervention response rate	Comparison response rate
Assigned phone coaching	440	431	871	n.a.	n.a.	n.a.
Contributed a baseline survey	339	337	676	77.6%	77.0%	78.1%
Contributed to first follow-up survey (post-test at end of R3 Academy)	254	260	514	59.0%	57.8%	60.2%
Contributed to second follow-up (one-year follow-up)	209	224	433	49.7%	47.5%	51.9%

n.a. = not applicable.

B. Outcome measures

For the primary research questions, participants were assessed on changes in their scores on father involvement, coparenting, and economic provision between pre-test (prior to the first class) and post-test (immediately after the last class concluded). Individuals that did not complete each item of a scale were coded as missing.

Father involvement was measured with developmentally appropriate measures of paternal engagement with children from the Fatherhood Research and Practice Network (FRPN). Respondents were asked to

indicate the frequency of engaging in between 9 and 11 activities with their youngest child—the focal child of this analysis. The sample was randomly split in half for age-specific exploratory and confirmatory factor analyses on the measures. These analyses indicated that there were two subscales in the overall scale and age specific Cronbach's alphas are provided in Table IV.2. The first addresses age-appropriate caregiving behaviors, while the other focuses on emotional and social support. Age-specific measures were combined into a measure for all fathers with children 21 years of age and younger. Because there were slight differences in the number of questions asked for each age group—scores were standardized with higher scores indicative of greater involvement. The scale for caregiving ranges from 0 to 40 and ranges from 0 to 12 for support.

Coparenting relationship quality was assessed with an 11-item scale from FRPN. Respondents were asked to provide their level of agreement, on a five-point Likert scale, with items about how mothers and fathers work together to parent the focal child. Items were summed to generate a scale ranging from 0 to 44.

Economic provision was assessed with a five-item scale, in nFORM, primarily addressing job skills. Individual items were measured on a four-point Likert scale indicating the respondent's level of agreement with each statement. Items were summed together to generate a scale ranging from 0 to 20, with higher scores indicative of greater job skills/economic provision.

For each scale, change scores were analyzed by subtracting the pre-test score from the post-test score.

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

Outcome measure	Description of the outcome measure	Source	Timing of measure
Father Involvement—Caregiving Subscale[1]	This measure comes from the Fatherhood Research & Practice Network (FRPN) Father Engagement Scale. Items are scored on a 0= never to 4= everyday/almost everyday scale for frequency. The caregiving subscale includes the following items: **1.** How often have you fed/given a bottle to child?[a] **2.** How often have you praised child?[a,b,c,d] **3.** How often have you watched or cared for child? [a,b,c,d] **4.** How often have you put child to sleep?[a] **5.** How often have you played with child? [a,b,c,d] **6.** How often have you talked to child? [a,b,c,d] **7.** How often have you hugged child? [a,b,c,d] **8.** How often have you had meals with child? [b,c,d] **9.** How often have you watched TV with child?[,c,d] Cronbach's alpha 1 month-12 months: 0.82 13 months-5 years: 0.80 6 years-11 years: 0.83 12 years-21 years: 0.84	Local evaluation	Post-test (immediately after intervention ends)

Outcome measure	Description of the outcome measure	Source	Timing of measure
Father Involvement—Support Subscale[1]	This measure comes from the Fatherhood Research & Practice Network (FRPN) Father Engagement Scale. Items are scored on a 0= never to 4= everyday/almost everyday scale for frequency. The support subscale includes the following items. **1.** How often have you encouraged child?[b,c,d] **2.** How often have you told child you love them?[b,c,d] **3.** How often have you taught child to make good choices?[b] **4.** How often have you talked to child?[c,d] **5.** How often have you read to child?[a] **6.** How often have you sung to child?[a] **7.** How often have you told stories to child?[a] Cronbach's alpha 1 month-12 months: 0.87 13 months-5 years: 0.82 6 years-11 years: 0.84 12 years-21 years: 0.84	Local follow-up survey	Post-test (immediately after intervention ends)
Co-parenting Relationship Quality	The outcome measure consists of 11 items scored from FRPN on a Likert scale (0= strongly disagree, 1= disagree, 2= neutral, 3= agree, 4= strongly agree). All fathers, regardless of coparental relationship status, were asked how well parents work together on the following dimensions: **1.** Mother contradicts father's decisions (RC) **2.** Mother makes negative comments about the way father parents (RC) **3.** Mother undermines father (RC) **4.** Mother and father discuss way to meet child's needs **5.** Mother and father share information about the child **6.** Mother and father make joint decisions about child **7.** Mother tries to understand father's perspective **8.** Mother respects father's decisions about parenting **9.** Mother makes it hard to spend time with child (RC) **10.** Mother makes it hard to talk with child (RC) **11.** Mother tells child what they are allowed to say to father (RC) Cronbach's alpha: 0.81	Local evaluation	Post-test (immediately after intervention ends)

Outcome measure	Description of the outcome measure	Source	Timing of measure
Economic provisions	Economic provisions were measured with five items scored on a Likert scale (0= strongly agree, 1= agree, 2= disagree, 3= strongly disagree). These items measure statements about employment and include: **1.** I have good job skills **2.** I know where to find job openings **3.** I know how to apply for a job **4.** I feel confident in my ability to conduct an effective job search **5.** I feel confident in my interviewing skills Cronbach's alpha: 0.82	nFORM	At post-test (immediately after intervention ends)

[a]measured among children 1 month to 12 months old

[b]measured among children 13 months to 5 years old

[c]measured among children 6 years to 11 years old

[d]measured among children 12 years to 21 years old

[1] Age-specific measures are standardized using FRPN scoring procedures for combination into a single measure

RC= reverse coded.

For the secondary research questions, we focused on three of the four items used for the primary research questions. Each was assessed approximately one year after the class concluded. The scales used for secondary questions were: father involvement—caregiving, father involvement—support, and coparenting quality. All scales were measured in the same manner as they were at post-test. Cronbach's alpha was reassessed at one-year follow-up and all scales continued to demonstrate high reliability (see Table IV.3).

Because analyses focused on the durability of the skills learned during the R3 Academy and in treatment conditions, change scores were produced by subtracting the one-year follow-up score from the post-test score.

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

Outcome measure	Description of outcome measure	Source	Timing of measure
Father Involvement—Caregiving Subscale[1]	This measure comes from the Fatherhood Research & Practice Network (FRPN) Father Engagement Scale. Items are scored on a 0= never to 4= everyday/almost everyday scale for frequency. The caregiving subscale includes the following items: **1.** How often have you fed/given a bottle to child?[a] **2.** How often have you praised child?[a,b,c,d] **3.** How often have you watched or cared for child?[a,b,c,d] **4.** How often have you put child to sleep?[a] **5.** How often have you played with child? [a,b,c,d] **6.** How often have you talked to child? [a,b,c,d] **7.** How often have you hugged child? [a,b,c,d] **8.** How often have you had meals with child? [b,c,d] **9.** How often have you watched TV with child?[,c,d] Cronbach's alpha 1 month-12 months: 0.82 13 months-5 years: 0.80 6 years-11 years: 0.83 12 years-21 years: 0.84	Local evaluation	One-year follow-up
Father Involvement—Support Subscale[1]	This measure comes from the Fatherhood Research & Practice Network (FRPN) Father Engagement Scale. Items are scored on a 0= never to 4= everyday/almost everyday scale for frequency. The support subscale includes the following items. **1.** How often have you encouraged child?[b,c,d] **2.** How often have you told child you love them? [b,c,d] **3.** How often have you taught child to make good choices? [b] **4.** How often have you talked to child?[c,d] **5.** How often have you read to child?[a] **6.** How often have you sung to child?[a] **7.** How often have you told stories to child?[a] Cronbach's alpha 1 month-12 months: 0.87 13 months-5 years: 0.82 6 years-11 years: 0.84 12 years-21 years: 0.84	Local evaluation	One-year follow-up

Outcome measure	Description of outcome measure	Source	Timing of measure
Co-parenting Relationship Quality	The outcome measure consists of 11 items scored from FRPN on a Likert scale (0= strongly disagree, 1= disagree, 2= neutral, 3= agree, 4= strongly agree). All fathers, regardless of coparental relationship status, were asked how well parents work together on the following dimensions: **1.** Mother contradicts father's decisions **2.** Mother makes negative comments about the way father parents **3.** Mother undermines father **4.** Mother and father discuss way to meet child's needs **5.** Mother and father share information about the child **6.** Mother and father make joint decisions about child **7.** Mother tries to understand father's perspective **8.** Mother respects father's decisions about parenting **9.** Mother makes it hard to spend time with child **10.** Mother makes it hard to talk with child **11.** Mother tells child what they are allowed to say to father Cronbach's alpha: 0.81	Local evaluation	One-year follow-up

[a]measured among children 1 month to 12 months old

[b]measured among children 13 months to 5 years old

[c]measured among children 6 years to 11 years old

[d]measured among children 12 years to 21 years old

[1] Age-specific measures are standardized using FRPN scoring procedures for combination into a single measure

C. Baseline equivalence and sample characteristics

Baseline equivalence was assessed between each treatment and the control group on the analytic sample of respondents that provided data on all outcome measures. Equivalence was assessed on the four key outcome variables at baseline: father involvement—caregiving, father involvement—support, co-parenting quality, and economic outcomes. Further, equivalence was assessed on the following key demographic characteristics: class language (% in English), income, number of children under 21, employment status, educational attainment, racial/ethnic identity, age, foreign born status, partner status, and residential status relative to the child.

To assess equivalence for continuous variables we calculated Hedge's G and for dichotomous variables we calculated Cox's index. Groups were considered equivalent if the standardized effect size difference between control and treatment groups was less than or equal to 0.05. Because no effect size was greater than this benchmark and no statistically significant differences were found between intervention and comparison groups, the groups were considered equivalent and no statistical adjustments on the models were necessary (results reported in Table IV.4).

The sociodemographic characteristics of the two intervention and control groups are provided in Table IV.4. Overall, more than 80% of participants in this program indicated that Spanish was their primary language and took all assessments in Spanish. The sample consisted primarily of low-income fathers, with more than 40% of participants making less than $2,000 per month ($24,000 per year) and more than 70% making less than $3,000 per month ($36,000 per year). One reason for this may be the educational profile of participants. More than 40% did not complete high school and nearly 75% had a high school diploma/GED or less education. Only 12% of eligible fathers had at least a Bachelor's degree.

The modal number of children under 21 for each respondent was two and very few fathers had three or more children. Most fathers had some employment for wages at the time of enrollment. About half of the fathers identified as Hispanic or Latino, approximately 45% identified as White, and the remainder as a member of another racial/ethnic identity or multiple racial/ethnic identities. Fathers that identified themselves as Hispanic/Latino on the ethnicity question or via self-identification on the racial identification question were recoded as Hispanic or Latino. Fathers tended to be middle-aged, most were married, and resided with the focal child. On average, focal children were between 8 and 9 years of age at baseline.

Table IV.4a. Summary statistics of key baseline measures and baseline equivalence for case management and comparison, for individuals/couples completing (n=673)

Baseline measure	Case Management mean (standard deviation)	Comparison mean (standard deviation)	Case Management v. Comparison mean difference	p-value of difference
Father involvement—caregiving	29.43 (7.61)	28.69 (7.07)	1.07	0.37
Father involvement—support	8.28 (3.13)	8.39 (3.00)	0.11	0.86
Coparenting	28.06 (9.67)	28.27 (9.45)	0.21	0.18
Economic provisions	9.07 (2.73)	9.15 (3.04)	0.58	0.56
Spanish language primacy (%)	81.4% (n.a.)	80.3% (n.a.)	1.1%	0.57
Income				0.20
$1000 or less/month	13.5% (n.a.)	12.0% (n.a.)	1.5%	
$1001–$2000/month	27.9% (n.a.)	26.7% (n.a.)	1.2%	
$2001–$3000/month	32.5% (n.a.)	36.1% (n.a.)	3.6%	
$3001–$4000/month	13.7% (n.a.)	12.5% (n.a.)	0.8%	
$4001 or more/month	12.3% (n.a.)	12.8% (n.a.)	0.5%	0.20
Educational attainment				0.53
Less than high school	44.4% (n.a.)	41.1% (n.a.)	0.3%	
High school graduate/GED	32.5% (n.a.)	32.0% (n.a.)	0.5%	
Some college/Associate's	12.3% (n.a.)	13.9% (n.a.)	1.6%	
Bachelor's degree or higher	10.8% (n.a.)	13.0% (n.a.)	2.2%	
Number of children				0.24
1	27.4% (n.a.)	27.4% (n.a.)	0.0%	
2	59.0% (n.a.)	55.9% (n.a.)	3.1%	
3	6.0% (n.a.)	4.6% (n.a.)	1.4%	

Baseline measure	Case Management mean (standard deviation)	Comparison mean (standard deviation)	Case Management v. Comparison mean difference	p-value of difference
4+	7.5% (n.a.)	12.1% (n.a.)	4.6%	
Currently employed	81.7% (n.a.)	82.8% (n.a.)	1.1%	0.57
Racial/ethnic identity				0.99
Hispanic/Latino	50.8% (n.a.)	49.6% (n.a.)	1.2%	
White	44.7% (n.a.)	45.2% (n.a.)	0.5%	
Other racial/ethnic identity	4.5% (n.a.)	5.2% (n.a.)	0.7%	
Father's age	3.35 (1.04)	3.31 (1.02)	0.04	0.39
Father is foreign-born	85.4% (n.a.)	84.3% (n.a.)	1.1%	0.66
Marital status				0.19
Married	84.3% (n.a.)	81.9% (n.a.)	2.6%	
Separated/previously married	6.6% (n.a.)	6.3% (n.a.)	0.3%	
Cohabitating	5.2% (n.a.)	6.2% (n.a.)	1.0%	
Single	3.9% (n.a.)	5.6% (n.a.)	1.3%	
Does not reside with child	13.0% (n.a.)	13.7% (n.a.)	0.7%	0.38
Age of focal child	8.49 (5.35)	8.53 (5.28)	0.04	0.88

n.a. = not applicable.

Table IV.4b. Summary statistics of key baseline measures and baseline equivalence for case management and comparison, for individuals/couples completing (n=676)

Baseline measure	Phone coaching mean (standard deviation)	Comparison mean (standard deviation)	Phone coaching v. Comparison mean difference	p-value of difference
Father involvement—caregiving	29.45 (6.77)	28.69 (7.07)	1.07	0.34
Father involvement—support	8.40 (3.00)	8.39 (3.00)	0.01	0.81
Coparenting	29.50 (8.96)	28.27 (9.45)	1.23	0.19
Economic provisions	8.95 (2.62)	9.15 (3.04)	0.20	0.57
Spanish language primacy (%)	78.2% (n.a.)	80.3% (n.a.)	2.1%	0.50
Income				0.26
$1000 or less/month	12.6% (n.a.)	12.0% (n.a.)	0.6%	
$1001–$2000/month	28.1% (n.a.)	26.7% (n.a.)	1.4%	
$2001–$3000/month	30.2% (n.a.)	36.1% (n.a.)	5.9%	
$3001–$4000/month	17.9% (n.a.)	12.5% (n.a.)	5.4%	
$4001 or more/month	11.2% (n.a.)	12.8% (n.a.)	1.6%	
Educational attainment				0.51
Less than high school	39.0% (n.a.)	41.1% (n.a.)	2.1%	
High school graduate/GED	35.3% (n.a.)	32.0% (n.a.)	3.3%	
Some college/Associate's	15.1% (n.a.)	13.9% (n.a.)	1.2%	

Baseline measure	Phone coaching mean (standard deviation)	Comparison mean (standard deviation)	Phone coaching v. Comparison mean difference	p-value of difference
Bachelor's degree or higher	10.6% (n.a.)	13.0% (n.a.)	2.4%	
Number of children				0.27
1	29.7% (n.a.)	27.4% (n.a.)	2.3%	
2	55.6% (n.a.)	55.9% (n.a.)	0.3%	
3	5.2% (n.a.)	4.6% (n.a.)	0.6%	
4+	9.5% (n.a.)	12.1% (n.a.)	2.6%	
Currently employed	83.2% (n.a.)	82.8% (n.a.)	0.4%	0.54
Racial/ethnic identity				0.97
Hispanic/Latino	50.6% (n.a.)	49.6% (n.a.)	1.0%	
White	43.3% (n.a.)	45.2% (n.a.)	1.9%	
Other racial/ethnic identity	6.1% (n.a.)	5.2% (n.a.)	0.9%	
Father's age	3.28 (1.02)	3.31 (1.02)	0.03	0.339
Father is foreign-born	85.0% (n.a.)	84.3% (n.a.)	0.7	0.68
Marital status				0.18
Married	79.0% (n.a.)	81.9% (n.a.)	2.9%	
Separated/previously married	9..2% (n.a.)	6.3% (n.a.)	2.9%	
Cohabitating	6.3% (n.a.)	6.2% (n.a.)	0.1%	
Single	5.3% (n.a.)	5.6% (n.a.)	0.3%	
Does not reside with child	10.7% (n.a.)	13.7% (n.a.)	3.0%	0.12
Age of focal child	8.95 (5.35)	8.53 (5.28)	0.42	0.62

n.a. = not applicable.

V. Findings and Estimation Approach

A. Implementation evaluation

1. Key findings

This section discusses the key findings regarding the implementation of the program in both the comparison and treatment groups. The key findings for implementation are:

1. Participants in case management and phone coaching tended to participate in the majority of intended phone calls, but few engaged in all of the intended calls.
2. There were essentially no differences in dosage between participants assigned to 6- and 12-week programs.
3. There was high fidelity across all program elements.
4. There was moderate engagement among phone coaching participants, as measured by phone coach evaluations. Similar levels of engagement were observed among case management participants, as defined by their follow-up on provided referrals.

Dosage

Table V.0.a provides information on dosage among those assigned to either the case management or phone coaching groups. On average, those in the case management group received 65.4% of intended calls, while those in phone coaching received 63.3% of intended calls. Overall, nearly 85 percent in case management and 80 percent in phone coaching participated in at least one call and more than 60 percent received at least half their scheduled calls in both groups. However, only 4 percent and 7 percent of individuals in the two conditions participated in all calls, respectively. The precipitous drop may be related to the number of calls, feelings among phone call participants that calls were redundant, or large time commitments required of participants who when they enrolled in the R3 Academy had not expected to be assigned to a treatment group that entailed significant additional contact hours.

Context

One reason for the low number of participants completing all calls may be differences in program lengths. Some participants completed six-week courses, while others completed 12-week courses. Focusing on attendance and phone participation (see Table V.0.b), there are differences in call participation among the case management group. T-tests used to focus on proportional differences between groups were introduced to address potential statistically significant differences between groups. The mean percentage of calls participated in for those in six-week courses was 66.4 percent, compared to 60.6 percent for those attending 12-week courses ($p < .01$). No other contextual differences were found.

Fidelity

There are two questions regarding fidelity in the implementation of the program. The first addresses whether phone coaching participants received the coaching content as expected. Table V.0.b reports fidelity across the nine phone coaching content areas. Overall, the majority of phone coaching participants received the content as designed. Among the parenting-associated calls, the percentage of participants that received all components of the call, which ranges between two and four total components, is between 91 and 98 percent, depending on content area. These results indicate high fidelity in this area. The percentage of participants that received all components of the JAM call, focused on economic and job-related materials, was lower. 79 percent of participants received all components of this call. Notably, this call includes five components and participants failing to complete an at-home challenge included as part of the phone coaching call tended to be the reason why all components of this call were not completed. Indeed, this step not being completed was responsible for more than 90 percent of the total incompletes on the JAM call.

The second fidelity question considers whether comparison group participants received all of their weekly reminder calls. Service contact data shows that, on average, participants received 82.9% of their scheduled weekly calls. 98 percent received at least one call, 72 percent received at least 80 percent of the total expected calls, and nearly 70% all of their scheduled calls. Again, these results indicate relatively high fidelity within the program.

Table V.0.a. Dosage and fidelity for class attendance and phone calls among participants

	Mean	At least once	At least 50%	At least 80%	100%
Class attendance					
Comparison	76.9%	95.4%	82.6%	64.2%	36.1%
Case Management	72.7%	94.6%	77.1%	60.4%	32.4%
Phone coaching	71.9%	92.7%	75.6%	60.4%	32.1%
Overall	73.8%	94.2%	78.5%	61.6%	33.6%
Phone calls					
Comparison[1]	82.9%	98.0%	81.8%	72.4%	68.6%
Case Management	65.4%	84.6%	66.0%	25.9%	3.8%
Phone coaching	63.3%	80.1%	62.3%	30.3%	7.1%
Overall (treatment only)	64.9%	82.5%	64.4%	27.5%	5.4%

[1]Refers to weekly reminder calls only

Table V.0.b. Context of attendance and phone participation by course length

	6-week courses	12-week courses	p-value
Mean attendance	60.8%	58.7%	0.206
Mean calls–case management	66.4%	60.6%	<.001
Mean calls—phone coaching	65.9%	64.4%	0.197

Engagement

The final set of questions about the implementation of the program focuses on engagement among participants in phone coaching and phone services. Each element of the program can be found in Table II.1. Table V.0.c reports cooperation and comprehension across all phone coaching sessions. Cooperation addressed how willing phone coaching clients were to participate in their calls, on a four-point scale (0= non-compliant, 1= resistant, 2= guarded, 3= engaged). On average, clients tended to be somewhat guarded in their approach across all sessions, with average scores nearly at that mark on the scale. Comprehension was measured on a three-point scale (0= minimum, 1= moderate, 2= high). On average, comprehension tended to be moderate across all phone session components.

Engagement in the phone services group was assessed using data on referrals. Overall, 338 of the 448 participants (75.4%) in the phone services condition received at least one referral. Of the group that received referrals, 196 followed-up on these referrals (58.0%).

Table V.0.c. Fidelity and engagement in phone coaching

Phone session	Not all components	All components	Cooperation	Comprehension
Parenting styles	8%	92%	1.99	0.89
Power listening	6%	91%	1.98	0.88
Who owns the problem	5%	95%	1.98	0.91
Confrontation cycle	2%	98%	1.99	0.85
6 steps to a win-win	2%	98%	1.99	0.93
Values collision	6%	94%	1.98	0.91
Consequences	4%	96%	1.97	0.86
Review	6%	94%	1.99	0.95
JAM	21%	79%	1.99	0.98

B. Primary impact evaluation

1. Key findings

This section discusses the key findings regarding program impact in both the comparison and treatment groups. The key findings for the impact analyses are:

1. Compared to participants that received only the R3 academy, phone coaching participants had significantly larger improvements in caregiving, coparenting, and economic provision between pre- and post-test.
2. Compared to participants that received only the R3 academy, case management participants had significantly larger increases in economic provision.
3. There are significant effects of add-on services for most outcomes and the findings are robust.

The primary analysis focused on changes between pre- and post-test for four outcomes: father's caregiving, father's support, coparenting quality, and economic provision. The analyses included the covariates in Table V.1. to provide more precise impact estimations. Because the program took place across multiple sites in California, supplementary analyses focused on the possible clustering within location (see Appendix G for information on this analysis). These analyses indicated that clustering was significant for three of the four outcomes (support, coparenting, and economic provision). As a result, clustering was accounted for in all analyses. In order to provide corrected standard errors, ordinary least squares regression analyses were run with a correction for site clusters. Statistical significance was determined through two-tailed tests and were determined as significant at three levels: $p < .01$, $p < .05$, or $p < .10$.

We used an intent-to-treat (ITT) framework where all individuals with valid data on outcome variables and covariates were included in the analysis, regardless of their participation in the program. In the initial analysis, missing data was evaluated on outcome measures and covariates (Enders, 2010). These evaluations indicated that the data was missing completely at random or missing at random. The sociodemographic profile of individuals with missing data on outcome variables was also assessed. This analysis revealed few differences in the sociodemographic profiles of individuals that did or did not provide data on the outcomes. American born respondents were approximately 6% less likely to provide

information on caregiving and support than foreign born fathers, while foreign born fathers were about 11% less likely to provide data on economic provision. Child age was positively associated with the likelihood that respondents did not provide information on caregiving and support—though the odds were quite small. Non-residential fathers were slightly less likely (5% and 8%, respectively) to provide information on the caregiving and support scales. No other statistically significant differences were observed. In order to preserve sample size, regression imputation methods were used to account for missing data on baseline covariates, but not outcomes, using 20 imputed data sets (What Works Clearinghouse, 2019). Evaluation of imputed data versus non-imputed data revealed no substantive differences in results.

Table V.1. Covariates included in impact analyses

Covariate	Description of the covariate
Spanish language primacy	Respondent took course in Spanish (1= Spanish, 0= English)
Father's income	Father's reported income at baseline, 0= $1000 or less/month (reference), 1= $1001-$2000/month, 2= $2001-$3000/month, 3= $3001-$4000/month, 4= $4001 or more/month
Father's educational attainment	Father's highest educational attainment at baseline, 0= less than high school (reference), 1= High school graduate or GED, 2= Some college, Associate's degree, or post-high school vocational training, 3= Bachelor's degree or more
Number of children 21 or younger	Number of children 21 or younger father reports at baseline, 0= 1 child (reference), 1= 2 children, 2= 3 children, 3= 4 or more children
Currently employed	Father reports he is employed for income at baseline, 0= no (reference), 1= yes
Father's age	Father's reported age at baseline, 0= 18-20 years old (reference), 1= 21-24 years old, 2= 25-34 years old, 3= 35-44 years old, 4= 45-54 years old, 5= 55-64 years old, 6= 65 years old or older
Marital status	Father's reported marital status at baseline, 0= currently married, 1= separated/divorced/widowed, 2= cohabiting, 3= single
Non-residential father	Father reports he does not reside with focal child at baseline 0= resides with child (reference), 1= does not reside with child
Age of focal child	Age of focal child at baseline, in years (continuous)

Estimated effects of changes in outcome scores from pre- to post-test are presented in Table V.2. This table provides mean changes by intervention group, comparison of these differences (case management versus comparison and phone coaching versus comparison). Standardized effect sizes, in standard deviation units in order to assess the magnitude of the effect, from the clustered OLS regression models are presented in the far right-hand column of Table V.2. All models include the full suite of control measures shown in Table V.1.

For *caregiving,* we found that there were increases in all groups. The mean increase for the comparison group was 5.11 points, 5.58 for the case management group, and 6.01 for the phone coaching group. Only the difference between the phone coaching and comparison groups was statistically significant ($p < .05$). To assess the magnitude of these changes, standardized scores (beta coefficients), measured in standard deviation units, were used. Under standard social science effect sizes (Cohen, 1988), significant effects less than or equal to 0.25 standard deviations are small, 0.26 to 0.50 are considered moderate in size, and greater than 0.51 are considered large. In the case of caregiving, the effect size was small in size. Individuals in the phone coaching group had scores that were, on average, 0.17 standard deviations greater than those in the comparison category.

For *support,* the differences were substantially smaller. The mean changes in support were relatively small at 0.02 points for the comparison group, 0.07 points for the case management group, and 0.20 for the phone coaching group. Differences between these groups were not statistically significant.

Differences between groups were observed for *coparenting quality*, however. While the increase in coparenting scores was, on average, very small for the comparison group (M= 0.02), the increases for the case management (M= 1.02) and phone coaching (M= 1.39) were larger. Regression analyses indicated that both increases were significantly greater than that experienced in the comparison group. Individuals assigned to case management had scores that were 0.08 standard deviations greater, on average, than those assigned to the comparison group (p< .10). Those assigned to phone coaching averaged an increase of 0.12 standard deviations in coparenting quality over those in the comparison group (p< .05).

Finally, there were substantial increases in *economic provisions*, as well. The mean increase for the comparison group, case management, and phone coaching groups were 1.34, 1.77, and 1.68 points, respectively. Regression analyses indicated that the increase among the case management group was significantly greater than among the comparison group. On average, individuals who received case management had economic provision scores 0.19 standard deviations greater than those in the comparison group (p< 0.01). The difference between the phone coaching and comparison groups was not statistically different, however.

There were also some associations among the covariates listed in Table V.1., as well. For caregiving, we found that changes were smaller as children got older—meaning that the program had a weaker correlation with caregiving if fathers reported on older children. A one standard deviation increase in focal child age was associated with a 0.12 standard deviation decrease in the change score for caregiving between pre- and post-test (p< .01). In contrast, paternal support scores increased more with focal child age. A one standard deviation increase in focal child age was associated with a 0.13 standard deviation increase in the support change score (p< .01).

We also found relationships between coparenting and income, language, employment, and educational attainment. For income, effects appeared to be largest for the poorest fathers. Fathers making between $1,001-$2,000 and $2,001-$3,000 per month had significantly smaller changes in coparenting quality than fathers making $1,000 or less per month. Scores in these two groups were 0.32 (p< .10) and 0.34 (p< .10) standard deviations lower, on average, than they were for the poorest fathers. The impact of the program for Spanish speakers was smaller than English speakers. The difference between these two groups was 0.35 standard deviations (p< .05). Employed fathers, however, experienced slightly better changes in their scores than unemployed fathers. The difference between these groups was 0.10 standard deviations (p< .10). Finally, fathers with a high school diploma/GED (B= –0.11, p< .05) and fathers with a college degree (B= –0.16, p< .10) both reported smaller changes in their coparenting than fathers without a high school diploma.

Finally, high school graduates reported a significantly greater increase in economic provisions than those without a high school diploma. The difference between these two groups was 0.20 standard deviations (p< .05). Foreign-born fathers, meanwhile, reported smaller increases than native-born fathers (B= –0.23, p< .05), as did fathers who were separated or previously married when compared to currently married participants (B= –0.23, p< .10). Child age was also associated with increases in change scores. A one standard deviation increase in focal child age was associated with a 0.10 standard deviation increase in the change score between pre- and post-test (p< .01).

Table V.2. Post-intervention estimated effects using data from post-test immediately after intervention to address the primary research questions (n= 778)

Outcome measure Comparison	Intervention mean change (standard deviation)	Comparison mean change (standard deviation)	Intervention compared to comparison mean difference (p-value of difference)	Effect size
Father Involvement—Caregiving				
Case management vs. comparison	(4.57)	5.11 (4.97)	0.46 (0.291)	0.09
Phone coaching vs. comparison	6.01 (4.33)	5.11 (4.97)	0.90 (0.027)	0.17*
Father Involvement—Support				
Case management vs. comparison	0.07 (3.41)	0.02 (3.37)	0.05 (0.622)	0.01
Phone coaching vs. comparison	0.20 (3.38)	0.02 (3.37)	0.18 (0.475)	0.08
Coparenting quality				
Case management vs. comparison	1.02 (8.57)	0.04 (9.76)	0.98 (0.054)	0.08+
Phone coaching vs. comparison	1.39 (8.54)	0.04 (9.76)	1.35 (0.016)	0.12*
Economic provision				
Case management vs. comparison	1.77 (2.14)	1.34 (1.91)	0.43 (0.005)	0.19**
Phone coaching vs. comparison	1.68 (2.38)	1.34 (1.91)	0.34 (0.147)	0.09

Source: West Coast Dads & Kids Initiative, 2017-2020, comparison between baseline data and post-test taken immediately after intervention.

Notes: p-values are included in parentheses. Effect sizes are calculated using Hedge's g. See Table IV.2 for a more detailed description of each measure and Chapters IV and V.B for a description of the impact estimation methods.

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

C. Sensitivity analyses

1. Key findings

To address the robustness of our findings, we conducted several sensitivity analyses. These results are summarized in Table V.3. Benchmark results, also reported in Table V.2. are provided in Table V.3. for comparison. Our first robustness check focused on individuals that attended at least one class or participated in at least one phone call, if assigned to a treatment group. Missing data was handled using the same methods as with the benchmark models, imputing data only on missing baseline covariates. Notably, the differences between the benchmark results, presented in the far left-hand column and in the treated only analysis are generally similar, with a few notable exceptions. For support, the difference between the phone coaching and control groups becomes statistically significant at $p< .10$. For coparenting, the difference between the case management and comparison groups increases in magnitude and become statistically significant ($p< .10$). Likewise, the difference between phone coaching and comparison grow in magnitude, from 0.12 standard deviations to 0.17, and in statistical significance, from $p< .10$ to $p< .01$. Finally, the difference in economic provision scores between the phone coaching and comparison groups increases slightly in magnitude ($B= 0.14$) and in significance level.

A second strategy to deal with missing data, this time on the outcome measures, is to carry forward the last observation (LOCF). In this case, scores at baseline are carried forward to post-test and change scores

are entered as zero. Differences between the LOCF and the benchmark models are very small and indicate that change scores are not sensitive to missing data on outcomes at post-test.

Finally, we used multiple imputation methods to account for missing data on covariates and outcome measures with between 5 and 30% missingness. The results of these models indicate significant differences for both the case management and phone coaching groups in reference to the comparison group for caregiving, a significant difference between the phone coaching and comparison group for support, and significant increases for both treatment groups versus the comparison group for both coparenting and economic provision.

Overall, these analyses indicate that our benchmark results are largely robust, although somewhat sensitive to missing data on outcome measures. Generally speaking, the models agree on four main points: (1) fathers assigned to phone coaching had significantly greater improvements in caregiving than fathers assigned to the comparison group; (2) assignment (phone coaching and case management) category had little impact on changes in support scores; (3) fathers in the phone coaching group reported larger improvements coparenting between pre- and post-test than fathers in the comparison group; and (4) fathers in the case management category experienced more improvement economic provision than fathers in the comparison group.

Table V.3. Differences in mean changes between intervention and comparison groups estimated using alternative methods

Outcome	Intent-to-treat benchmark (OLS Regression)	Treated only (OLS Regression)	LOCF (OLS Regression)	Imputed models (OLS Regression)
Father involvement—caregiving				
Case management vs. comparison	0.09	0.07	0.08	0.08+
Phone coaching vs. comparison	0.17*	0.17*	0.15*	0.17*
Father involvement—support				
Case management vs. comparison	0.01	0.03	0.02	0.02
Phone coaching vs. comparison	0.08	0.10+	0.07	0.09+
Coparenting				
Case management vs. comparison	0.08	0.11*	0.08	0.14*
Phone coaching vs. comparison	0.12+	0.17**	0.11+	0.16**
Economic provision				
Case management vs. comparison	0.19**	0.22**	0.17**	0.19**
Phone coaching vs. comparison	0.09+	0.14*	0.09	0.16**
N	778	668	1,009	996

Source: West Coast Dads & Kids Initiative, 2017-2020. Comparison between baseline data and post-test taken immediately after intervention.

Notes: Standardized coefficients presented. Treated only indicates respondent attended at least one class or participated in one phone call, if in treatment group. LOCF= Last Observation Carried Forward

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

D. Additional analyses

1. Key findings

In addition to the primary impact evaluation, several additional questions were examined. Table V.4a. reports results for the first three secondary research questions, which addressed mean changes in parenting, coparenting, and economic outcomes between pre- and post-test for fathers that attended R3. Overall, we found that there were statistically significant increases in all four and that effect sizes were small to moderate in magnitude. The sample included participants that provided post-test data and completed a one-year follow up. Among this group, caregiving scores increased an average of 2.59 points (g= 0.165), support by 1.84 points (g= 0.151), coparenting by 1.69 points (g= 0.136) and economic provision by 0.81 points (g= 0.294). All differences were statistically significant.

Table V.4a. Post-intervention and baseline means for fathers attending R3 Academy using data from post-test to address the secondary research questions (n= 778)

Outcome measure	Intervention mean or % (standard deviation)	Comparison mean or % (standard deviation)	Intervention compared with comparison mean difference (p-value of difference)	Effect size
Father involvement—caregiving	28.26 (7.16)	30.85 (7.37)	2.59 (p= 0.003)	0.165
Father involvement—support	10.41 (3.04)	12.25 (2.98)	1.84 (p= 0.011)	0.151
Coparenting	37.48 (9.38)	39.29 (7.68)	1.69 (p= 0.005)	0.136
Economic provision	8.21 (2.81)	9.02 (2.67)	0.81 (p< 0.001)	0.294

Source: West Coast Dads & Kids Initiative, 2017-2020, comparison between baseline data and post-test taken immediately after intervention.

Notes: p-values are included in parentheses. Effect sizes are calculated uses Hedge's g. See Table IV.3 for a more detailed description of each measure and Chapters IV and IV.B for a description of the impact estimation methods.

Results in Table V.4b. address the remaining six secondary research questions. These models are descriptive models that assess changes in father involvement and coparenting between the end of the R3 Academy (post-test) and the one-year follow-up. Missing data for these analyses was handled in the same manner as our benchmark models. Namely, missing data was imputed for baseline (posttest) covariates, but not for outcomes at the one-year follow-up. Sensitivity analyses revealed that the results are robust to different missing data strategies (LOCF and imputation on the dependent variable).

Between post-test and one-year follow up, fathers in both the case management and comparison groups reported a decline in their caregiving scores. For the case management group, fathers averaged a 0.44-point decline between post-test and follow-up, while the comparison group averaged a 0.43-point decline. This difference was not statistically significant. In contrast, the phone coaching group averaged a 0.35-point increase in their caregiving scores between post-test and one-year follow-up. The difference between scores was statistically significant in regression models including all covariates listed in Table V.1. Scores in the phone coaching group were, on average, 0.10 standard deviations greater (p< .01) at one-year follow-up than they were in the comparison group.

Support scores increased slightly for all groups between post-test and one-year follow-up. The average change in score between the two time periods for the comparison group was 0.69 points, 0.80 points for

the case management group, and 0.89 points for the phone coaching group. There were no statistically significant differences between intervention and comparison groups, however.

Finally, there were marked increases in coparenting between post-test and follow-up for all groups. For the comparison group, scores increased by 2.78 points, on average, between these two time points. The intervention groups reported larger increases of 4.14 points and 7.68 points for the case management and phone coaching groups, respectively. Results from regression analyses indicated that the difference between phone coaching and comparison was statistically significant (B= 0.13, p< .05).

Table V.4b. Post-intervention estimated effects using data from one-year follow-up to address secondary research questions

Outcome measure-Comparison	**Intervention mean change (standard deviation)**	**Comparison mean change (standard deviation)**	**Intervention compared to comparison mean difference (p-value of difference)**	**Regression Coefficient**
Father Involvement—Caregiving (n= 653)				
Case management vs. comparison	–0.44 (8.12)	–0.43 (7.93)	0.01 (0.984)	0.00
Phone coaching vs. comparison	0.35 (6.76)	–0.43 (7.93)	0.78 (0.000)	0.10**
Father Involvement—Support (n= 653)				
Case management vs. comparison	0.80 (3.89)	0.69 (4.05)	0.11 (0.368)	0.02
Phone coaching vs. comparison	0.89 (3.38)	0.69 (4.05)	0.20 (0.276)	0.05
Coparenting (n= 648)				
Case management vs. comparison	4.14 (5.46)	2.78 (5.34)	1.36 (0.142)	0.04
Phone coaching vs. comparison	7.68 (7.12)	2.78 (5.34)	4.90 (0.017)	0.13*

Source: West Coast Dads & Kids Initiative, 2017-2020, comparison between post-test taken immediately after intervention and one-year follow-up.

Notes: p-values are included in parentheses. Effect sizes are calculated using standardized regression coefficients (*B*). See Table I V.2 for a more detailed description of each measure and Chapters IV and V.B for a description of the impact estimation methods. Regression coefficients are standardized.

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

VI. Discussion

A. Implementation Analysis

The implementation of the program was assessed across multiple measures of fidelity, dosage, engagement, and context. Overall, attendance at classes and participation in phone services was relatively high. Participants, on average, attended over 75 percent of classes and phone participants averaged about 60 percent on their scheduled phone calls. There were precipitous declines in phone call participation and few participants engaged in all calls. Moreover, phone call engagement tended to be somewhat low, as measured by comprehension and cooperation. Measures of fidelity, however, suggest that case managers and phone coaches showed high levels of fidelity with the program, suggesting that participants may have been unenthused about engaging with the phone content of the program. For example, a comparatively

small number of participants engaged in the homework given as part of the JAM session component of the course materials which impacted their engagement with phone coaching regarding that part of the program content.

B. Impact Analysis

The results of the impact analysis indicate that, overall, there were some significant differences in outcomes between intervention and comparison groups. Fathers who received phone coaching had significantly greater increases in their caregiving than comparison group fathers between pre- and post-test. Similar patterns were observed around coparenting quality. Meanwhile, there were no significant differences in changes across time between fathers in the case managements and comparison conditions. While increases in mean support scores were observed for all groups, there were no differences between pre-post based on assignment category. The results around economic provisions highlighted a third pattern. Namely, that fathers in the case management category reported more significant increases in these scores than their counterparts in the comparison group. In contrast, no statistically significant differences between the phone coaching and comparison group were observed.

Secondary descriptive analyses considered the question of long-term outcomes of the program in the areas of father involvement and coparenting quality. This was analyzed with data collected at one-year post program exit and comparing it to the outcomes collected at program exit. Notably, fathers in the case management and comparison groups, on average, experienced declines in caregiving scores between post-test and one-year follow-up. In contrast, scores tended to increase for fathers still in the sample at post-test who were in the phone coaching group. For the remaining two outcomes, scores increased for all conditions among those still in the sample at post-test—albeit slightly in most cases. We found that fathers remaining in the sample at post-test who were assigned to the phone coaching group had significantly higher scores for caregiving and coparenting at one-year follow-up than similar fathers assigned to the comparison condition. No statistically significant differences were observed between case management and comparison at one-year follow-up.

The differences in patterns across outcomes may be explained by the content that participants received in each condition. Case management explicitly focused on the needs of participants by identifying areas of concern around their day-to-day lives, well-being, and basic necessities. Much of this was done through referral, planning, help navigating systems, and providing assistance, as necessary. The explicit focus on these areas may help explain why fathers assigned to this condition reported significantly larger increases on the economic provisions outcomes compared to those in the other categories. In contrast, fathers assigned to the phone coaching condition reviewed skills learned during the R3 Academy and were given opportunities to practice them in weekly phone calls. This skill practice tended to focus on relationships with coparenting partners and children, helping participants apply course content to their specific circumstances, and providing a review of each week's material. Such content may help explain why fathers in this condition improved more, on average, in caregiving and coparenting—and the apparent persistence of these effects over time.

The lack of differences, particularly with respect to supportive fathering are a bit more difficult to explain. One potential reason for a lack of variability among fathers is that R3 Academy course content may have more readily addressed issues around emotionally supporting children and being available when their children need them. This content may not have readily been integrated into many conversations in phone coaching in the same way caregiving and coparenting were emphasized. Additional research will need to consider why we saw no differences in this area versus other areas of our evaluation.

C. Limitations

Our evaluation has a number of limitations worth noting. First, an initial group of participants that were assigned to a third condition, which was dropped, was not included in this analysis. This condition combined both case management and phone coaching for individuals. Given the varied effects of these two conditions across outcomes, it is possible that this group may have experienced improved outcomes relative to both intervention groups and the comparison group. Second, our study faced a problem common to many programs and analyses of this type—namely attrition. Fathers were difficult to contact, evaluate, and include at all time points across the study. Continued efforts to address attrition and retention in this population are necessary. These problems are compounded by concerns about immigration status, access to the internet, and other data collection issues which raised problems throughout the study. Third, data was collected in a variety of methods—via paper, computer, over the phone, and via secured links to respondents. How these factors may have affected responses was not assessed in this report. However, it is possible that various data collection methodologies, employed to reduce attrition as much as possible, could have influenced responses. Fourth, fathers provided self-report data and reports from coparents were not included in our analysis. Many studies have shown that fathers often overestimate their participation in parenting and coparenting—which could influence their responses throughout the study. Thus, response bias is a potential issue in our study (Dyer et al., 2014). Fifth, individuals in control and treatment conditions all attended R3 Academy classes together. Individuals assigned to the comparison group may have experienced spillover effects from those assigned to phone coaching or case management. Such spillover effects could not be assessed in this data. To the extent that the effects of treatment were positive, individuals assigned to comparison could have inadvertently benefited from their participation in classes with those who received treatment. Finally, we did not evaluate potential moderators and mediators of the relationships presented here. Although we included numerous control measures in our models, it is possible that effects varied by race/ethnicity, language, nativity status, residential status, and relationship status. Future work would do well to consider such relationships.

The role of implementation cannot be overlooked in explaining some of the smaller effects of this program, however. Somewhat low levels of participation in phone coaching and services may depress some of the effects seen here. For example, cooperation levels were low and phone coaches indicated that participants had moderate levels of comprehension in calls. Effects may be larger for those participants that were more engaged in these components of the program. At the same time, however, program participants show clear and demonstrable improvements in all assessed components. Course content provided in the R3 Academy may be particularly relevant for understanding the value of the overall program.

D. Conclusion

Overall, we found that fathers in treatment groups were more involved in caregiving and support, more engaged coparents, and improved in their ability to provide economically, relative to the comparison category. However, the effects of phone coaching and case management were somewhat mixed. Phone coaching appeared to have positive effects on relational aspects of parenting and coparenting between pre- and post-test. Meanwhile, case management had stronger economic benefits. Overall, however, these additional aspects of the program appeared to have benefits above and beyond course materials. Thus, our results suggest that added-on services, beyond course content, have benefits and targeted services could be utilized, depending upon a program's emphasis, goals, and mission.

VII. References

Avellar, S., Covington, R., Moore, Q., Patnaik, A., & Wu, A. (2018). Effects of four Responsible Fatherhood programs for low-income fathers: Evidence from the Parents and Children Together Evaluation. OPRE Report Number 2018-50. Washington, DC: Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.https://www.acf.hhs.gov/opre/resource/effects-four-responsible-fatherhood-programs-low-income-fathers-evidence-parents-and-children-together-evaluation

Center for Disease Control and Prevention. (n.d.) Marriage and divorce. http://www.cdc.gov/nchs/fastats/marriage-divorce.htm Accessed May 2015.

Census.gov: American FactFinder, American Community Survey, *5-Year Estimate: 2009-2013*: Poverty tables, U.S. and California. Accessed May 2015.

Cohen, J. (1988). Statistical Power Analysis for the Behavioral Sciences (2nd Edition). Hillsdale, NJ: Lawrence Earlbaum Associates.

Cullen, S.M., Cullen, A., Band, M., Davis, L. & Lindsay, G. (2011) Supporting fathers to engage with their children's learning and education: an under-developed aspect of the Parent Support Adviser pilot. *British Educational Research Journal*, 37(3), 485–500.

Ghate, D., Shaw, C. & Hazel, N. (2000). *Engaging Fathers in Preventive Services: Fathers and Family Centres*. Joseph Rowntree Foundation, York.

Holmes, E.K., Egginton, B.R., Hawkins, A.J., Robbins, N.L. and Shafer, K. (2020). Do Responsible Fatherhood Programs Work? A Comprehensive Meta-Analytic Study. *Fam Relat.* doi:10.1111/fare.12435.

Koestner, R., Franz, C.E., & Weinberger, J. (1990). The family origins of empathic concern: A 26-year longitudinal study. *Journal of Personality and Social Psychology, 58,* 709-717.

Lam, C.B., McHale, S.M., & Crouter, A.C. (2012). Parent-child shared time from middle childhood to late adolescence: Developmental course and adjustment correlates. *Child Development, 83*,(6): 2089-2103.

Lieberman, M., Doyle, A.B., & Markiewicz, D. et al., (1999). Attachment to mother and father: Links to peer relations in childhood. *Child Development*, 70, 202-213.

The National Center for Education Statistics. (1997). The Condition of Education, 1997. Washington, DC: U.S. Department of Education, Office of Education Research and Improvement.

Rohner, R.P. & Veneziano, R.A. (2001). The importance of father love: history and contemporary evidence. *Review of General Psychology*, 5(4), 382-405.

Zaveri, H., Baumgartner, S., Dion, R., and Clary, L. (2015). Parents and Children Together: Design and implementation of Responsible Fatherhood programs. OPRE Report Number 2015-76. Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services. https://www.acf.hhs.gov/opre/resource/parents-and-children-together-design-and-implementation-of-responsible-fatherhood-programs

VIII. Appendices

A. R3 Academy Class Sessions

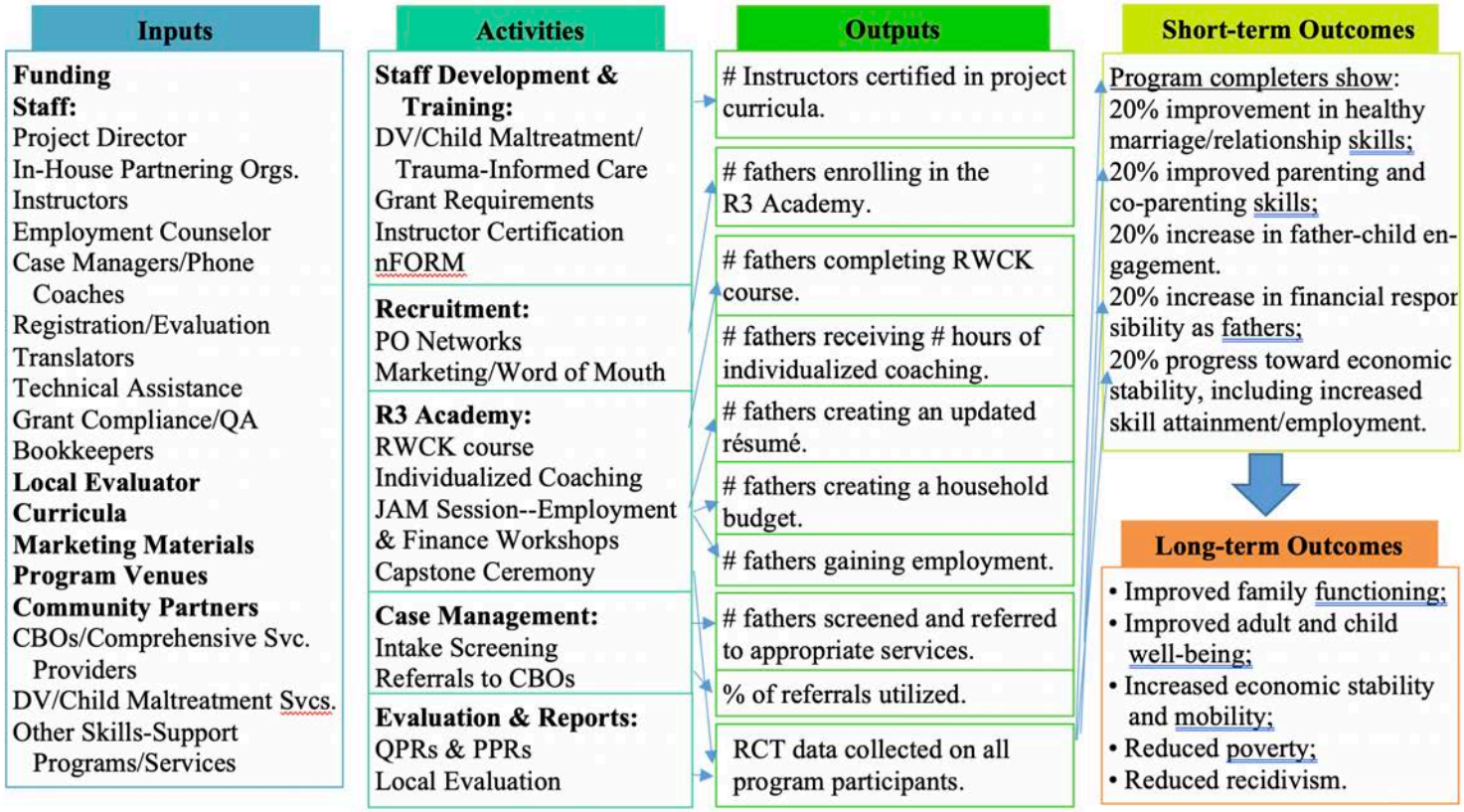
Table A.1. Topics Covered in R3 Academy Class Sessions

Session Number	Curriculum	Topics Covered
1	None	Entrance Survey
2	Raising World Class Kids (RWCK)	Becoming a World Class Parent
3	RWCK	Building Blocks for Successful Communication
4	RWCK	How Parents and Kids Can Speak and Be Heard
5	RWCK	Making Your Family a Winning Team
6	RWCK	Bringing Out Character in Your Kids
7	RWCK	What Kids Need from Those Who Love Them
8	RWCK	Understanding Your Child and Supporting School Success
9	RWCK	Making Good Choices in the Real World
10	Jobs and Money Session (JAM): Making Money Work (MMW)	Financial Literacy and Making Money Work
11	JAM: Job Search Success (JSS)	Skills Assessment, Resume Writing
12	JAM: JSS	Job Interviews; Exit Survey

B. LOGIC MODEL (OR THEORY OF CHANGE) FOR THE PROGRAM

Figure B.1. Logic Model

Goals: Increase the frequency and quality of father-child interactions, and thereby improve overall family relationships and child outcomes, by providing fathers with comprehensive Relationship Education and employment-support programs.
Assumptions: 1) Communication and problem-solving are key skills that many fathers do not develop early in life; 2) Teaching these to fathers is a cost-effective way to address societal problems; 3) Evidence-informed curricula help fathers improve relationships with children, partner, and the workplace; 4) Strengthening father-child relationships has large impact on key social factors in society.[1]
Target Populations: Low-income/unemployed/under-employed fathers, including TANF recipients, in 7 West Coast cities.



[1] Hundreds of research studies supporting these assumptions in *Healthy Marriages, Healthy Society* and other resources at: www.RelationshipsCA.org.

C. DATA AND STUDY SAMPLE

Description of Implementation data

For a full text description of the table below, see pages 15-16.

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

Implementation element	Research question	Data source	Timing/frequency of data collection	Party responsible for data collection
Intervention Group Questions				
Fidelity	Did the Phone Coaching clients receive the course content expected?	Phone Coach fidelity check survey (Qualtrics)	Completed within 24 hours of each contact	Phone Coaches
Dosage	How many phone calls did the Phone Coaching clients participate in on average?	Individual service contacts in nFORM	Completed within 24 hours of each contact	Phone Coaches
Dosage	How many phone calls did the Case Management clients participate in on average?	Individual service contacts in nFORM	Completed within 24 hours of each contact	Case Managers
Engagement	How willing were Phone Coaching clients to engage in their weekly calls?	Phone Coach fidelity check survey (Qualtrics)	Completed within 24 hours of each contact	Phone Coaches
Engagement	Did Phone Coaching clients comprehend the skills taught?	Phone Coach fidelity check survey (Qualtrics)	Completed within 24 hours of each contact	Phone Coaches
Engagement	How many Case Management clients received referrals?	Referrals in nFORM	Completed within 24 hours of each contact	Case Managers
Engagement	How many referrals were followed up on?	Referrals in nFORM	Completed within 24 hours of each contact	Case Managers
Context	Was there a difference in Phone Coaching participation for the 6 week classes versus the 12 week classes?	nFORM data on session series length; individual service contacts in nFORM	Session series data entered when new class launched; service contacts completed within 24 hours of each contact	HRC Staff; Case Managers and Phone Coaches
Context	Was there a difference in Case Management participation for the 6 week classes versus the 12 week classes?	nFORM data on session series length; individual service contacts in nFORM	Session series data entered when new class launched; service contacts completed within 24 hours of each contact	HRC Staff; Case Managers and Phone Coaches
Control/Comparison Group Questions				
Fidelity	Did all control group members receive their expected weekly reminder calls?	Individual service contacts in nFORM	Completed within 24 hours of each contact	HRC staff

Description of Impact Data

1. Overall, the sample consisted of men with at least one child aged 21 or younger (due to nForm restrictions). All outcome measures were restricted to fathers with children of this age. The analytic sample varied according to the research question and are listed, below.

Primary research questions

Q1: The analytic sample consisted of all individuals that were intended to be treated in the Phone Coaching group (ITT). In other words, if they were randomly assigned, they stayed assigned in the analytic sample—regardless of their participation in the program.

Q2: The analytic sample consisted of all individuals that were intended to be treated in the Case Management group (ITT). In other words, if they were randomly assigned, they stayed assigned in the analytic sample—regardless of their participation in the program.

Q3: The analytic sample consisted of all individuals that were intended to be treated in the Phone Coaching group (ITT). In other words, if they were randomly assigned, they stayed assigned in the analytic sample—regardless of their participation in the program.

Q4: The analytic sample consisted of all individuals that were intended to be treated in the Case Management group (ITT). In other words, if they were randomly assigned, they stayed assigned in the analytic sample—regardless of their participation in the program.

Q5: The analytic sample consisted of all individuals that were intended to be treated in the Phone Coaching group (ITT). In other words, if they were randomly assigned, they stayed assigned in the analytic sample—regardless of their participation in the program.

Q6: The analytic sample consisted of all individuals that were intended to be treated in the Case Management group (ITT). In other words, if they were randomly assigned, they stayed assigned in the analytic sample—regardless of their participation in the program.

Secondary research questions

Q1: The analytic sample consisted of individuals that participated in the R3 Academy and completed at least some portion of the baseline survey and individuals that completed survey items for all outcome variables at post-test.

Q2: The analytic sample consisted of individuals that completed some portion of the baseline data and individuals that completed both surveys for father involvement at post-test and one-year follow-up.

Q3: The analytic sample consisted of individuals that completed some portion of the baseline data and individuals that completed both surveys for father involvement at post-test and one-year follow-up.

Q4: The analytic sample consisted of individuals that completed some portion of the baseline data and individuals that completed both surveys for co-parenting quality at post-test and one-year follow-up.

Q5: The analytic sample consisted of individuals that completed some portion of the baseline data and individuals that completed both surveys for co-parenting quality at post-test and one-year follow-up.

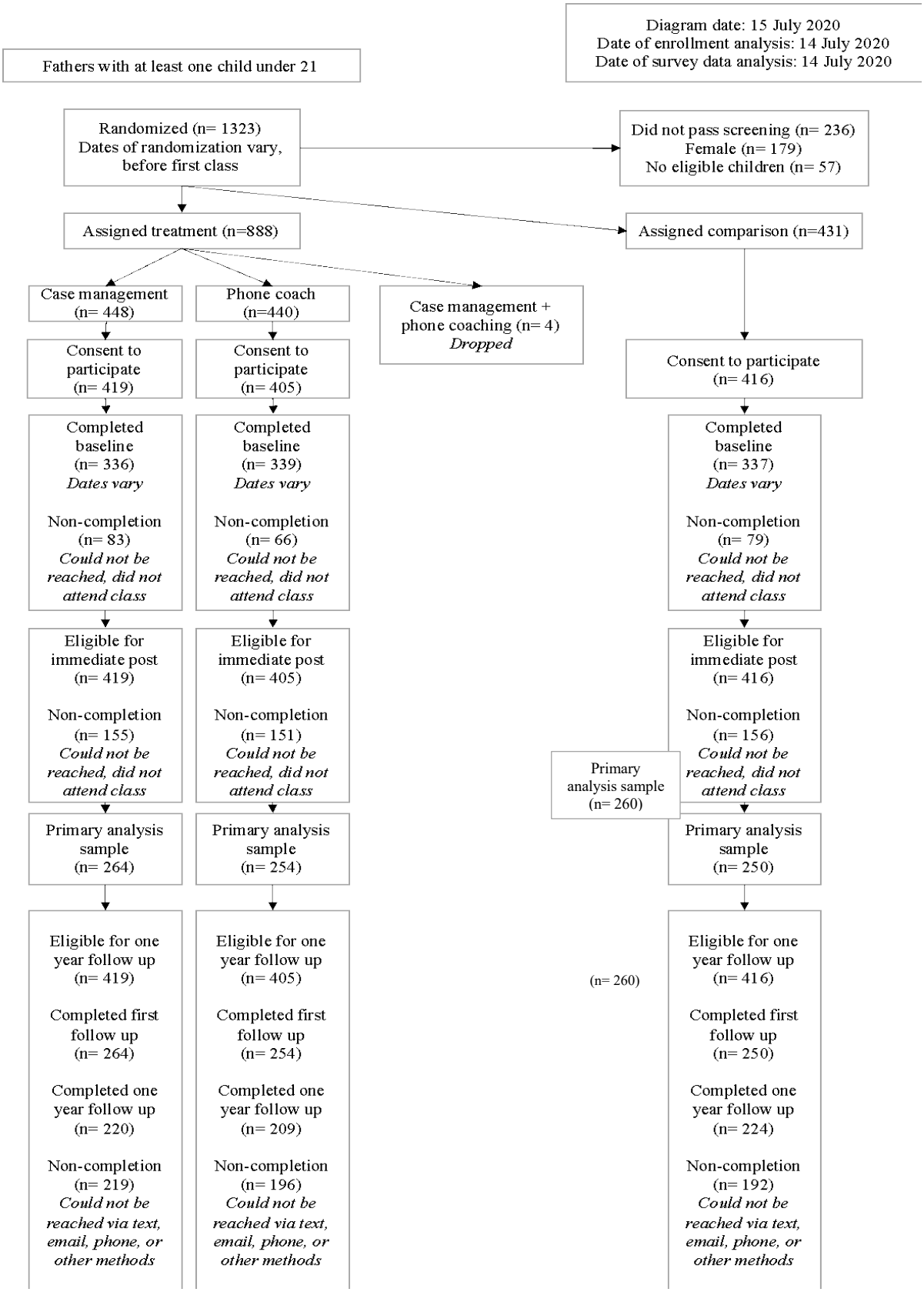
Q6: The analytic sample consisted of individuals that completed some portion of the baseline data and individuals that completed both surveys for economic outcomes at post-test and one-year follow-up.

Q7: The analytic sample consisted of individuals that completed some portion of the baseline data and individuals that completed both surveys for economic outcomes at post-test and one-year follow-up.

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

	Data source	Timing of data collection	Mode of data collection	Party responsible for data collection	Start and end date of data collection
Intervention	Intervention group study participants	Enrollment (demographic characteristics)	Telephone survey (nFORM Applicant Characteristics Survey)	Program staff	September 2016 through May 2019
		First Class Session (baseline)	In-person online survey (nFORM; Qualtrics); paper survey if technology fails or client misses first class	Program Staff	September 2016 through June 2019
		Last Class Session (posttest)	In-person online survey (nFORM; Qualtrics); paper survey if technology fails; telephone survey if class is missed	Program Staff for in class collection, Evaluation Staff for phone survey	February 2017 through July 2019
		1 year follow-up	In-person online survey (Qualtrics); telephone survey if reunion event is not attended	Program staff for in-person collection, Evaluation Staff for phone survey	February 2018 through April 2020

Figure C.1. CONSORT Diagram of HRC Impact Evaluation



D. Phone Coaching Fidelity Checks

Phone coaches were asked to rate participants on comprehension of the material taught and on willingness to cooperate with the phone coach. The following table describes ratings on these measures of fidelity.

Table D.1. Phone Coach Fidelity Check Rating Criteria

Topic	Rating	Criteria
Comprehension	Minimal	By end of call, practices or applies majority of skills incorrectly. Cannot describe concepts covered.
	Moderate	By end of call, practices and applies most skills accurately but still requires some correction. Describes majority of concepts taught accurately but not all.
	High	By the end of call, correctly practices and applies skills. Can accurately describe concepts taught.
Cooperation	Noncompliant	Refuses to participate in phone coaching. Will not answer questions or engage in skills practice.
	Resistant	Declines to participate in skills practice or discussion, but ultimately does participate after encouragement from coach.
	Guarded	Willing to participate in skills practice and discussion, but hesitant or fearful to share real life scenarios or to respond to exercises.
	Engaged	Eager/comfortable participating in skills practice and discussion with coach.

Note: Prior to engaging with clients, Phone Coaches were trained on this criteria by HRC's Research Manager and by the Phone Team Supervisor. Phone coaches met regularly with their supervisor and could consult her if clarification was needed after a client interaction.

E. ATTRITION RATES AND BASELINE EQUIVALENCE OF THE RCT DESIGN

Group equivalence was monitored quarterly by the local evaluator. At each time period, there was no statistically significant non-equivalence, as measured by t-tests, ANOVAs, and chi-square analyses. Because equivalence was achieved at each time-point, no adjustments were necessary throughout. Table IV.4. reports equivalences in the final sample and is discussed in section IV.C.

F. DATA PREPARATION

Data Preparation and cleaning was completed using the following steps:

2. Data from nFORM and the local evaluation, which was collected through Qualtrics, were merged together using unique individual IDs assigned at baseline. Confirmation of successful merging was obtained using Stata 16.0.
3. Duplicate observations and responses were checked using built-in checks in Stata 16.0. 47 duplicate IDs were identified in the data in the merger of baseline and post-test data. 31 of these were data entries that were started and then ended for various reasons. In these cases, data was merged together to create a single entry. In 16 cases, the ID was entered multiple times, but only one entry included any responses. These 16 cases were eliminated. Data was rechecked for duplicate entries and none were identified. The same process was repeated when baseline/post-test and one-year follow-up data. Five duplicate IDs were identified in this merger. In each case, the ID was entered multiple times, but only one entry included any responses. These cases were removed.
4. Data were assessed for inclusion criteria. Inclusion criteria included respondent gender and child age. Mothers (n= 179) and fathers without a child 21 or younger were removed from the analysis (n= 57).

5. All crosstabs were checked on outcome and relevant variables during coding and recoding. All items for scales and variables were checked for valid values. Most questions included discrete choice options. Child age, father age, and race/ethnicity (if the respondent chose "other") allowed for open response. Child age was encoded numerically and all values were assigned discrete values in years. The same process was undertaken for father's age. Finally, individuals were assigned to race/ethnicity categories based on their responses. 47% of respondents that identified as "other" reported that they were Latino or Hispanic. These individuals were reassigned to this category, along with those that identified as Latino/Hispanic in the specific question in nFORM.
6. No variables showed coding errors in non-valid values. However, all categories with less than 5% of valid values were collapsed into larger categories for comparison. This was done for the income, education, and race/ethnicity variables.
7. Data was checked for consistency in reporting on child at baseline, post-test, and one-year follow-up. This was done by checks on child name/initials, child age, and child sex. The focal child in the analyses was the youngest child of the respondent. In 17 cases, it could not be verified that fathers were reporting on children across time periods. Because consistency could not be identified in the models, these respondents were removed from the analysis.
8. All variables were recoded so that higher scores reflected higher levels of involvement, coparenting quality, and economic provisions. Any scale items which required reverse coding were coded as such.
9. Exploratory and confirmatory factor analyses were conducted on all outcome measures. Following Kline (2015), the sample was randomly split in half for scale assessment. Items in the EFA with factor scores of 0.40 or greater were analyzed in the CFA. CFA scores of 0.50 or higher were retained for final inclusion in the scales. Cronbach's alpha was calculated for each scale. Because the father involvement scales were age specific, these were standardized following the recommendations of FRPN for combination and analysis of fathers across child age. Because of the lack of reliability of mean imputation methods and the restrictions on the use of multiple imputation techniques, if a respondent did not answer all questions in the outcome scale, they were removed from the analysis.
10. All continuous measures were assessed for normality and transformed. None showed problematic skewness.
11. Missing data patterns were evaluated for common missing data patterns, etc. on all measures using the evaluation methods outlined in Enders (2010). Variables were assessed using Little's test of missing completely at random (MCAR). Variables which were not missing at random were evaluated for missing at randomness (MAR). All variables were either MCAR or MAR, allowing for imputation on covariates. Sociodemographic profiles of respondents with missing data were analyzed in relation to the outcomes. Small differences were observed by foreign born status, child age, and residential status.

G. IMPACT ESTIMATION

Baseline equivalence was assessed using baseline equivalence commands in Stata 16.0. The Table1 command allows for evaluation of baseline equivalence on various items. Continuous items were evaluated using ANOVAs with Kruskal-Wallis tests (for analysis of more than two groups) for continuous variables, Pearson's chi-square was used for both categorical and dichotomous variables and checked for robustness using Fisher's exact test. Effect sizes for baseline equivalence were calculated using Stata 16.0. Hedges *g* was calculated for continuous variables and Cox's index for

categorical/continuous variables. All effect sizes were between 0.00 and 0.05, meaning that statistical adjustment was not required.

Condition crossover was assessed using data from phone calls, enrollment rosters, and workshop attendance to determine if individuals in the control condition received contact outside class, beyond reminder calls. None were identified.

H. SENSITIVITY ANALYSES AND ALTERNATIVE MODEL SPECIFICATIONS

All analyses were conducted using linear regression analyses including a variable for control-treatment group. Models were adjusted using a clustered regression because an intercepts-only multi-level regression to assess the interclass correlation (ICC) indicated that the ICC was problematic, following the recommendation of Hoffmann (2016) the design effect threshold was two, using the design effect test. ICCs and design effect scores are reported in Table H.1. below. Our analytic sample includes all individuals with intention-to-treat. In other words, we included everyone that was assigned to a condition. Because we are not permitted to impute on outcome variables, we can only include those individuals that were non-compliant with the full program, but still completed the survey at post-test.

Table H.1. Interclass Correlations and Design Effects at Post-test

Measure	ICC	Design effect
Father involvement		
Caregiving	0.003	2.387
Support	0.034	5.383
Coparenting	0.027	4.351
Economic	0.025	4.223

The following sensitivity analyses were conducted:

1. To address the possibility of differences between intention-to-treat and the treated by conducting a sensitivity analysis on only the treated (defined as attending at least one class).
2. Last-observation-carried-forward (LOCF) was used to assess the possibility that attrition affected results. This method uses scores from pre-test and carries them forward to post-test assuming no change in outcome scores.
3. Imputed data were used on the outcomes to assess potential sensitivity to missing data. Models with variables where variables with at least 5% missing are imputed will be compared to main models and discrepancies between the two will be reported.

All sensitivity analyses were performed on primary and secondary research questions.