

OPRE Report  
2011-17

# *The* **BUILDING STRONG FAMILIES PROJECT**

**BSF's Effects on  
Couples Who  
Attended Group  
Relationship Skills  
Sessions: A Special  
Analysis of  
15-Month Data**



May 2011





OPRE Report Number:  
2011-17

Contract Number:  
233-02-0056

Mathematica Reference Number:  
08935.190

**Submitted to:**  
Nancye Campbell, Project Officer  
Seth Chamberlain, Project Officer  
Office of Planning, Research, and Evaluation  
Administration for Children and Families  
U.S. Department of Health and Human  
Services  
[www.acf.hhs.gov/programs/opre/index.html](http://www.acf.hhs.gov/programs/opre/index.html)

**Submitted by:**  
Robert G. Wood, Project Director  
Mathematica Policy Research  
P.O. Box 2393  
Princeton, NJ 08543-2393  
Telephone: (609) 799-3535  
Facsimile: (609) 799-0005

**BSF's Effects on Couples Who  
Attended Group Relationship  
Skills Sessions: A Special  
Analysis of 15-Month Data**

May 2011

Robert G. Wood  
Quinn Moore  
Andrew Clarkwest



This report is in the public domain. Permission to reproduce is not necessary. Suggested citation: Wood, Robert G., Quinn Moore, and Andrew Clarkwest (2011). *BSF's Effects on Couples Who Attended Group Relationship Skills Sessions: A Special Analysis of 15-Month Data*, OPRE Report # 2011-17, Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

**Disclaimer**

The views expressed in this publication do not necessarily reflect the views or policies of the Office of Planning, Research and Evaluation, the Administration for Children and Families, or the U.S. Department of Health and Human Services.



## **ACKNOWLEDGMENTS**

We would like to thank the many people who have contributed in significant ways to this report. First, we wish to acknowledge the support of the staff at the Office of Planning, Research, and Evaluation (OPRE) at the Administration for Children and Families, U.S. Department of Health and Human Services. Special thanks go to our OPRE project officers, Nancye Campbell and Seth Chamberlain, who have provided thoughtful guidance throughout this effort. We also wish to thank Scott Stanley, Paul Amato, and Charles Michalopolous who provided very helpful comments on our research design before we began our analysis.

At Mathematica Policy Research, we would like to thank Peter Schochet who provided input on the analysis plan, as well as Alan Hershey and Ken Fortson who provided very useful comments on earlier drafts of the report. We would also like to thank Melissa Clark for providing statistical code for performing our matching procedures. Xiaofan Sun provided excellent programming assistance. Cindy George carefully edited the report and Linda Heath did an excellent job in producing it.

The Authors



## **CONTENTS**

Introduction .....	1
BSF Group Sessions: The Core of Participation .....	2
Methods for Estimating Effects for Couples Who Attended Group Sessions .....	4
Impacts of BSF on Couples Who Attended Group Sessions .....	6
Discussion .....	9
References .....	11
Appendix A: Technical Documentation for the BSF Treatment-on-the-Treated Analysis .....	A.1
Appendix B: Supplemental Tables .....	B.1





**TABLES**

1.	Curricula Used by BSF Programs.....	3
2.	Impact of BSF on Key Outcomes at 15-Month Follow-up for All Couples.....	6
3.	Estimated Impact of BSF on Key Outcomes at 15-Month Follow-up for Couples Who Attended at Least One Group Session .....	7
4.	Estimated Impact of BSF on Key Outcomes at 15-Month Follow-up for Couples Who Attended at Least Half of Group Sessions .....	8

**FIGURES**

1.	The BSF Program Model .....	2
2.	Percentage of Couples Attending BSF Group Sessions .....	3
3.	Initial Characteristics of BSF Group Couples by Level of Attendance at Group Sessions .....	4



## Introduction

Building Strong Families (BSF), a program of relationship skills education for unwed parents, has been found in a rigorous random assignment evaluation to have limited effects on couples who signed up for the program (Wood, McConnell, et al. 2010). Averaging results across the eight local programs that participated in the evaluation, BSF had no effect on the couples' relationship quality or on the likelihood that they would remain romantically involved or get married 15 months after they enrolled in the program. When impacts were examined separately for the eight programs, only one was found to have a consistent pattern of positive effects on couples' relationships, while another was found to have negative effects.

These results, however, leave us with an unanswered question of wide interest, because not all couples randomly assigned to receive BSF services actually participated. The core BSF service was group workshops on relationship skills, and across all evaluation sites about 45 percent of the couples assigned to the program group never attended even one workshop session. BSF was a voluntary program and voluntary programs, particularly those serving low-income families, often have low participation rates (McCurdy and Daro 2001; Myers et al. 1992; Garvey et al. 2006).<sup>1</sup> Even so, it is natural to ask whether BSF had any effects on the couples who did attend group sessions.

This question is not answered in the BSF 15-month impact report, which is based on analyses of all couples who signed up for BSF, including those assigned to the program group who never attended a group session (Wood, McConnell, et al. 2010). The impacts reported thus represent the average effect on all program applicants of being *offered* BSF services and not the effect of actually attending group sessions. Such “intent-to-treat” (ITT) impact estimates are widely used in large-scale evaluations, for two reasons. First, they preserve the key strength of a random assignment research design—specifically that one can be confident that the program and control groups were similar at baseline, and that differences in outcomes that emerge (and that are large enough to be unlikely to be the result of chance) can be attributed to the program. If some sample members who were randomly assigned are excluded from the analysis, one can no longer be certain that the two research groups are similar and that differences in outcomes between them represent the effect of the program. Second, ITT estimates address a policy-relevant research question: What is the effect of offering a program in the “real world,” where one can anticipate that not everyone in the target population will participate in all program services? Nonparticipation may limit a program's ability to affect outcomes in the target population the program seeks to help, despite extensive efforts to promote attendance, and it is appropriate to reflect this limitation in estimates of a program's impact.

There are reasons, however, to address the question of whether BSF improves outcomes for those who participate, using other analytical methods. Despite the value of ITT estimates, they cannot be used to estimate impacts on those who received services, since the program group couples who actually receive services are not selected randomly. If policymakers or program leaders believe that improvements in implementation or program targeting could substantially raise participation rates, knowing whether participation yields impacts could help determine whether making such

---

<sup>1</sup> See Dion et al. (2010) for a detailed discussion of participation patterns in the BSF program and the challenges associated with achieving high rates of attendance at group sessions among low-income, unmarried couples.

improvements is worth the effort. Estimating the impact on participants who received particular program services, also known as “treatment-on-the-treated” (TOT) impacts, requires using quasi-experimental research techniques—in other words, techniques that do not rely solely on the study’s random assignment design. After providing some further detail on the BSF model, this report explains two such methods for estimating TOT impacts and their results. Both methods are used to estimate TOT impacts on key relationship outcomes for the eight BSF program sites combined, and not for individual program sites, because (as discussed in Appendix A) TOT estimation introduces sample size limitations that make estimation of site-specific impacts infeasible.

The analysis finds no strong evidence of effects on couples who attended group sessions. Among those who attended at least one group session, there were no statistically significant effects on the key relationship outcomes. Among the smaller group of couples who attended at least half of the group sessions offered, there was no strong evidence of effects, with one exception. BSF appears to have increased the likelihood that these couples would be living together (married or unmarried) at the 15-month follow-up—with an impact on this outcome of 7 to 10 percentage points.

### BSF Group Sessions: The Core of Participation

The core component of BSF was curriculum-based group education on relationship skills (Figure 1). The BSF model did not require a specific curriculum, but it did require that the group sessions use a curriculum that covered key topics such as communication, conflict management, and marriage. Three curricula were developed for the eight BSF programs by experts who tailored their existing curricula for married couples to the needs of unmarried parents (Table 1). These curricula varied in the total hours of group sessions offered and the specified ideal group size. Group sessions usually met weekly but the formats differed. Sessions ranged in length from two to five hours, with shorter sessions typically held on weeknights and longer sessions held on weekends. Depending on the format and the number of hours of instruction offered, the curriculum could take as little as six weeks or as much as five months to complete. The program in Oklahoma City chose the *Becoming*

Figure 1. The BSF Program Model

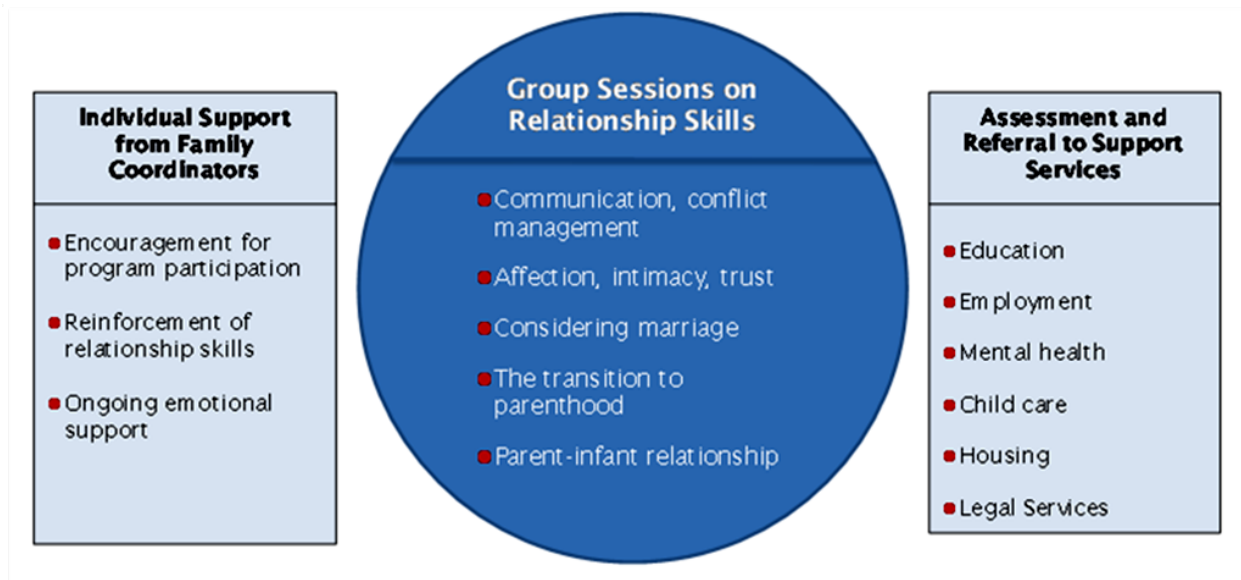


Table 1. Curricula Used by BSF Programs

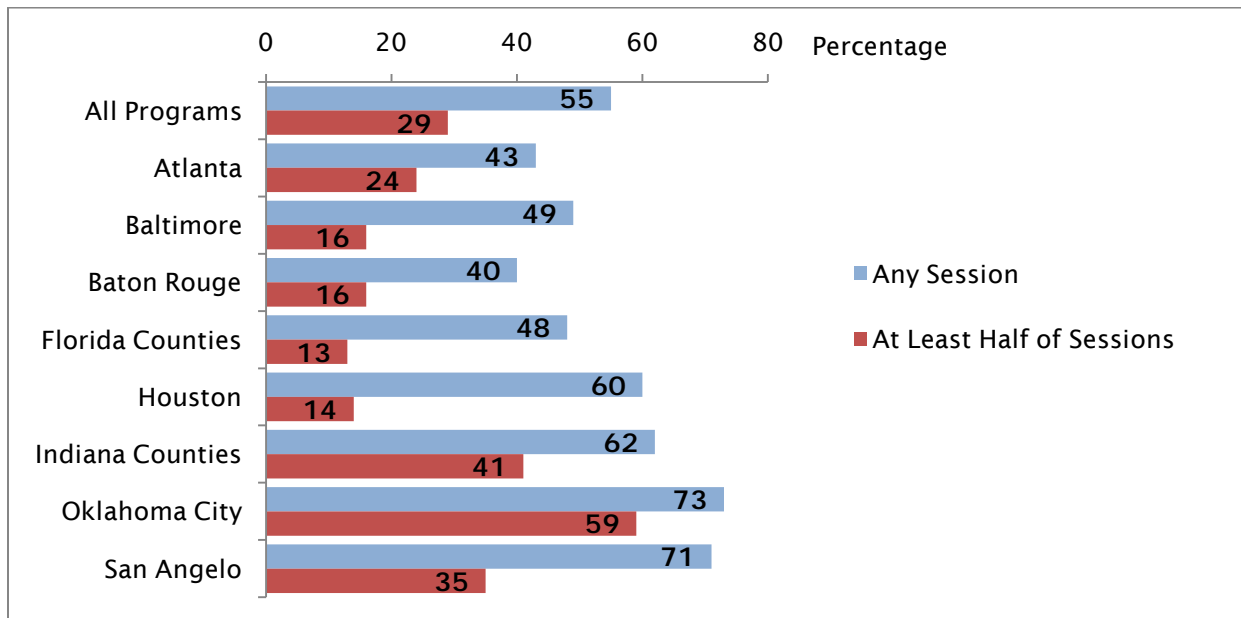
Curriculum	Developers	Total Hours of Group Sessions Offered	Group Size	Programs Using Curriculum
Becoming Parents for Low-Income, Low-Literacy Couples	Pamela Jordan	30	10 to 15 couples	Oklahoma City
Love’s Cradle	Mary Ortwein and Bernard Guerney	42	6 to 8 couples	San Angelo, Houston
Loving Couples, Loving Children	John and Julie Gottman	42	4 to 6 couples	Atlanta, Baltimore, Baton Rouge, Florida Counties, Indiana Counties

*Parents* curriculum and the San Angelo and Houston programs chose *Love’s Cradle*. The other five programs chose *Loving Couples, Loving Children*.

Attendance at group sessions was not required, and a substantial number of BSF couples did not attend. Across the eight programs, 55 percent of couples assigned to the program group attended at least one group session (Figure 2). Frequent attendance was less common. Only 29 percent of couples in the program group completed at least half of the curriculum. Rates of attendance varied substantially across the eight programs. In Oklahoma City and San Angelo, more than 70 percent of BSF couples attended at least one group session together. In contrast, only 43 percent of couples in Atlanta and 40 percent of couples in Baton Rouge attended even one group session.

The BSF program model included other services in addition to group relationship skills education (Figure 1). In particular, it included a family coordinator whose role was to reinforce relationship skills, provide emotional support, and encourage participation in and completion of the

Figure 2. Percentage of Couples Attending BSF Group Sessions



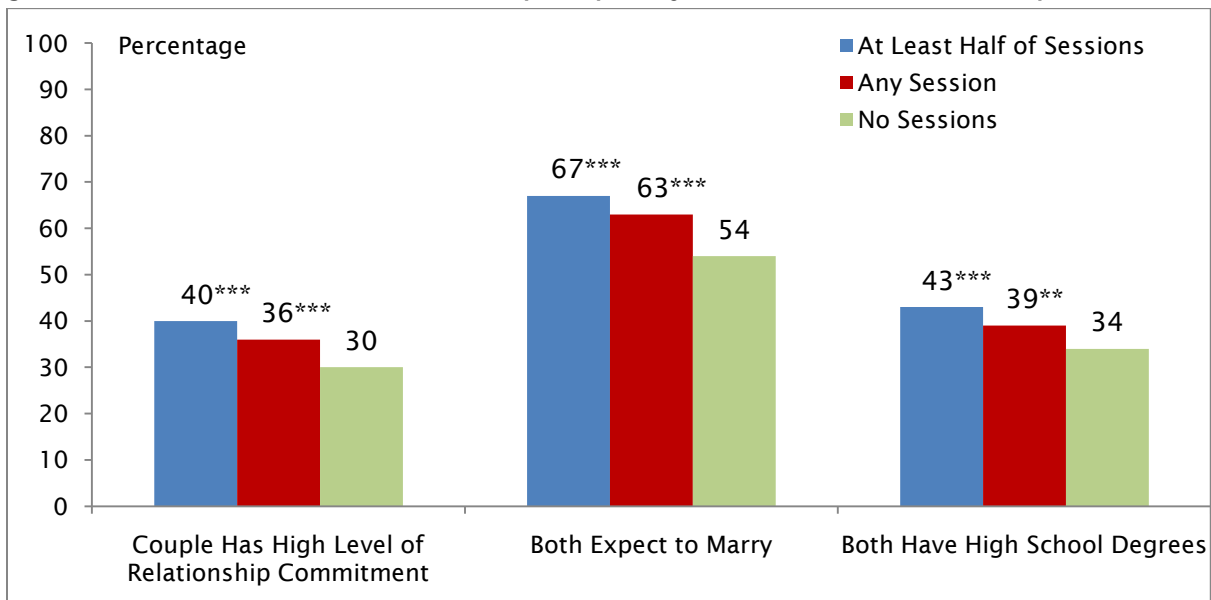
Source: BSF management information system data.

group sessions. The family coordinator also assessed family members’ needs and referred them for appropriate support services, such as education, employment, and mental health services. Many BSF couples who did not attend group sessions received one of these additional services from the program. Overall, 93 percent of BSF couples received some service from the program. Thus, attempting to distinguish effects of receiving services of any kind from the effects of the BSF program offer would not be productive, since the large majority of couples received some service from the program. This analysis instead focuses on whether participating in group sessions at a specified level affected key outcomes. It includes two measures of participation: (1) whether couples attended at least one group session, and (2) whether they attended at least half of the scheduled sessions.

### Methods for Estimating Effects for Couples Who Attended Group Sessions

The central challenge in estimating BSF impacts on participants is identifying an appropriate comparison group. This is difficult because attendance at group sessions was not determined randomly, but resulted from couples’ own choices and situations. Descriptive analysis provides evidence that couples who did attend group sessions were different from those who did not; they had higher levels of relationship commitment initially, were more likely to rate their chances of marriage as high, and had higher levels of education (Figure 3). Because of these initial differences, it is inappropriate to compare outcomes for the control group as a whole to outcomes for program group couples who attended group sessions. Calculating an accurate impact estimate requires comparing BSF couples who attended group sessions to those in the control group who *would have* attended if they had been offered access to the program. The challenge, of course, is figuring out which control group couples these are.

Figure 3. Initial Characteristics of BSF Group Couples by Level of Attendance at Group Sessions



Source: BSF management information system data and BSF baseline information forms.

Note: High level of commitment defined as being in the top third of all BSF couples on the relationship commitment scale. See Wood, Moore, et al. (2010) for a discussion of how this scale was constructed.

\*\*\*/\*\*/\* Statistically significantly different from the “no sessions” group at the .01/.05/.10 level.

## Methods Used in This Analysis

This analysis uses the following two quasi-experimental research methods to address this challenge:<sup>2</sup>

1. A *traditional matching approach*, in which individual couples who attended BSF group sessions are matched to similar control couples and the outcomes of these two groups are compared to estimate program effects
2. A *“likely attender” approach*, in which a group of “likely attenders” is identified within each research group and the outcomes of these two groups are then compared to estimate impacts

Both of these methods involve estimating a predicted probability of attending group sessions, or a propensity score, based on couples’ characteristics at the time they applied for BSF. In the traditional approach to propensity score matching, program group couples who attended sessions are matched to individual couples in the control group who have similar propensity scores. This method should generate two research groups that are similar in their observed initial characteristics. However, it is possible that the groups still differ on unmeasured characteristics, such as their level of motivation to improve their relationships.

The “likely attender” approach uses propensity scores to identify couples in both research groups who are *most likely* to attend group sessions if they are offered to them. Thus, these TOT estimates are not based on couples in the BSF program group who actually attended, but on couples whose baseline characteristics indicate that they would be likely to attend. This approach preserves the BSF study’s experimental framework, because the predicted probability of participating is based entirely on initial characteristics and is not influenced by couples’ later choices. Therefore, one can be confident that the two groups of likely attenders are similar on both measured and unmeasured characteristics. However, if the propensity score model cannot accurately predict who is a likely attender, and the likely attenders in the program group are a substantially different set of couples from those who actually attended, then the results will not yield an accurate estimate of the effects of BSF on those who attended group sessions.<sup>3</sup>

Although these two approaches differ, both rely on the extent to which the propensity models identify sample members who would choose to attend group sessions if offered the opportunity. If the predictive power of the model is high, then the two approaches will yield similar results that are likely to reflect BSF’s effects on those who attended group sessions. Conversely, if the propensity model has little predictive power, these approaches tend to yield different results, neither of which is likely to represent the program’s effects for attenders (Schochet and Burghardt 2007). Therefore, an examination of the degree to which results from these two methods are similar can suggest how much confidence one can have in the estimates.

---

<sup>2</sup> Appendix A includes a more detailed description of the methods used in this analysis.

<sup>3</sup> Although the likely attender method does preserve the experimental framework, it is referred to here as a quasi-experimental approach to estimating effects on couples who attended group sessions because the research groups are based on those who are predicted to attend, not those who actually attended. Therefore, these results provide an accurate estimate of BSF’s impacts for those couples who appear likely to attend sessions, but not necessarily those who actually attended them.

## An Alternative Method: The Bloom Adjustment

Another commonly used approach to estimating TOT impacts is the Bloom adjustment, which inflates the experimentally obtained ITT estimates by the inverse of the proportion of program group members who actually received the intervention (Bloom 1984). The adjustment is based on the assumption that the impact of the program on nonparticipants is zero. The Bloom adjustment is not well suited for this analysis, because it cannot be used to estimate the effects of different types or amounts of services. However, if one assumes that BSF had no effect on couples who did not attend group sessions but who received other help from the program (such as support from a family coordinator or referrals to support services), then one can use a similar adjustment to estimate the effects of BSF on couples who attended group sessions. As discussed in greater detail in Appendix A, this adjustment yields estimates of the effects of BSF on couples who attended at least one group session that are similar to those found using the traditional matching and likely attender methods described above.

## Impacts of BSF on Couples Who Attended Group Sessions

### Results for the Eight Programs Combined

As reported before, results based on standard ITT impact estimation techniques indicate that BSF had no effect on key measures of relationship status or quality when results were averaged across the eight local programs (Wood, McConnell, et al. 2010). BSF and control group couples had almost identical rates of romantic involvement, co-residence, and marriage 15 months after random assignment (Table 2). Similarly, combined data for the eight programs show no impacts on key measures of relationship quality, such as relationship happiness, conflict management, and fidelity.

**Table 2. Impact of BSF on Key Outcomes at 15-Month Follow-up for All Couples**

Outcome	BSF Group	Control Group	ITT Impact	Effect Size
<b>Relationship Status</b>				
Still romantically involved (%)	76	77	-1	-0.04
Living together, married or unmarried (%)	62	62	0	0.00
Married (%)	17	18	-1	-0.06
<b>Relationship Quality</b>				
Relationship happiness	8.37	8.32	0.06	0.04
Support and affection	3.46	3.45	0.01	0.03
Use of constructive conflict behaviors	3.26	3.23	0.03	0.05
Avoidance of destructive conflict behaviors	2.76	2.75	0.01	0.01
Neither reports infidelity (%)	75	73	2	0.06
<b>Sample Size</b>	<b>2,217</b>	<b>2,207</b>		

Source: BSF 15-month follow-up survey, conducted by Mathematica Policy Research.

Notes: See Wood, Moore, et al. (2010) for a discussion of how these relationship measures were constructed. Impacts are adjusted using a pooled regression controlling for the couple’s baseline relationship and demographic characteristics. Impact estimates are calculated based on a weighted average of program-level impacts in which all programs are weighted equally. See Appendix A for more details. The difference between BSF and control group means may not equal the estimated impact due to rounding.

\*\*\*/\*\*/\* Statistically significant at the .01/.05/.10 level.

ITT = Intent to treat.



For couples who attended at least one group session, the TOT estimation techniques similarly find no overall effects on key relationship outcomes. Analyses using both the traditional matching and likely attender methods show no statistically significant effects on relationship status or quality among the 55 percent of couples who attended at least one group session (Table 3). In addition, effect sizes remain relatively small (the largest is 0.12) when the analysis is restricted to couples who attended a group session.

When impacts are estimated for couples who attended at least half of the scheduled group sessions, a few statistically significant impacts emerge, although not a consistent pattern. Most notably, the estimates show that BSF had an impact on the likelihood that these couples would be living together, either married or unmarried, at the time of the 15-month follow-up. As shown in Table 4, analysis using the traditional matching method finds a 10-percentage-point impact of BSF on this measure among couples who attended at least half of the sessions. The likely attender method finds a 7-percentage-point impact on this measure. There is no impact on the likelihood that participating couples would be romantically involved or married at the time of the 15-month follow-up or, in most cases, on relationship quality. The one exception is the likelihood of remaining faithful to each other. Using the traditional matching method, BSF has a marginally statistically significant impact on fidelity; using the likely attender method it does not (Table 4).

**Table 3. Estimated Impact of BSF on Key Outcomes at 15-Month Follow-Up for Couples Who Attended at Least One Group Session**

Outcome	Using Traditional Matching Method				Using “Likely Attenders” in Both Research Groups			
	BSF Group	Comparison Group	TOT Impact	Effect Size	BSF Group	Comparison Group	TOT Impact	Effect Size
<b>Relationship Status</b>								
Still romantically involved (%)	78	77	1	0.02	78	78	-0	-0.00
Living together, married or unmarried (%)	65	62	3	0.08	65	64	2	0.05
Married (%)	18	19	-1	-0.03	20	21	-1	-0.04
<b>Relationship Quality</b>								
Relationship happiness	8.34	8.29	0.05	0.04	8.37	8.33	0.03	0.02
Support and affection	3.46	3.43	0.03	0.06	3.46	3.46	-0.00	-0.00
Use of constructive conflict behaviors	3.26	3.23	0.03	0.06	3.26	3.24	0.01	0.03
Avoidance of destructive conflict behaviors	2.73	2.74	-0.01	-0.01	2.74	2.75	-0.01	-0.01
Neither reports infidelity (%)	77	74	4	0.12	77	75	2	0.08
<b>Sample Size</b>	<b>1,276</b>	<b>823</b>			<b>1,276</b>	<b>1,252</b>		

Source: BSF 15-month follow-up survey, conducted by Mathematica Policy Research.

Notes: See Wood, Moore, et al. (2010) for a discussion of how these relationship measures were constructed. Impacts are adjusted using a pooled regression controlling for the couple’s baseline relationship and demographic characteristics. Impact estimates are calculated based on a weighted average of program-level impacts in which all programs are weighted equally. See Appendix A for more details. The difference between BSF and comparison group means may not equal the estimated impact due to rounding.

\*\*\*/\*\*/\* Statistically significant at the .01/.05/.10 level.

TOT = Treatment on the treated.

**Table 4. Estimated Impact of BSF on Key Outcomes at 15-Month Follow-Up for Couples Who Attended at Least Half of Group Sessions**

Outcome	Using Traditional Matching Method				Using “Likely Attenders” in Both Research Groups			
	BSF Group	Comparison Group	TOT Impact	Effect Size	BSF Group	Comparison Group	TOT Impact	Effect Size
<b>Relationship Status</b>								
Still romantically involved (%)	84	80	4	0.17	82	81	1	0.05
Living together, married or unmarried (%)	72	62	10**	0.27	72	65	7**	0.20
Married (%)	22	20	2	0.05	22	23	0	-0.00
<b>Relationship Quality</b>								
Relationship happiness	8.45	8.26	0.18	0.14	8.43	8.35	0.09	0.06
Support and affection	3.49	3.44	0.04	0.11	3.48	3.47	0.01	0.02
Use of constructive conflict behaviors	3.31	3.24	0.07	0.13	3.29	3.25	0.04	0.07
Avoidance of destructive conflict behaviors	2.74	2.71	0.03	0.05	2.76	2.75	0.02	0.03
Neither reports infidelity (%)	82	75	7*	0.26	83	80	4	0.14
<b>Sample Size</b>	<b>694</b>	<b>493</b>			<b>694</b>	<b>707</b>		

Source: BSF 15-month follow-up survey, conducted by Mathematica Policy Research.

Notes: See Wood, Moore, et al. (2010) for a discussion of how these relationship measures were constructed. Impacts are adjusted using a pooled regression controlling for the couple’s baseline relationship and demographic characteristics. Impact estimates are calculated based on a weighted average of program-level impacts in which all programs are weighted equally. See Appendix A for more details. The difference between BSF and comparison group means may not equal the estimated impact due to rounding.

\*\*\*/\*\*/\* Statistically significant at the .01/.05/.10 level.

TOT = Treatment on the treated.

### Results by BSF Curriculum Group

As discussed in Appendix A, sample size limitations make it infeasible to examine TOT impact estimates for all eight local BSF programs separately. Therefore, to explore variation in TOT impacts across programs, the analysis team estimated BSF’s effects among couples who attended group sessions for three groups of local BSF programs based on the curriculum they used. For this analysis, the local programs were grouped as follows: (1) the programs that used *Loving Couples*, *Loving Children* (those operating in Atlanta, Baltimore, Baton Rouge, Florida counties, and Indiana counties); (2) the programs that used *Love’s Cradle* (those operating in Houston and San Angelo); and (3) the programs that used *Becoming Parents* (the Oklahoma City program).

In general, the TOT results for these three groups of programs are broadly similar to ITT impact estimates, suggesting the ITT results do not substantially underestimate BSF’s true impact for these groups of programs. In addition, the pattern of positive effects found in ITT estimates for the Oklahoma program (the only program that used the *Becoming Parents* curriculum) persists when TOT estimation techniques are used. However, these results are less often statistically significant because of the smaller sample sizes available for TOT estimation. These results are presented and discussed in more detail in Appendix A.

## Discussion

This analysis examined whether BSF had effects on couples who received the core program service of group relationship skills education. As reported in the 15-month impact report, BSF had no effect on the key relationship outcomes for the full set of couples enrolled in the BSF evaluation. Based on combined data from the eight programs, BSF and control group couples had, on average, almost identical relationship outcomes 15 months after they applied for the program. A substantial portion of BSF applicants—45 percent across the eight programs—never attended a group relationship skills session. This analysis examined whether BSF had effects on the 55 percent of couples who did attend group sessions.

In general, the analysis finds no strong evidence of effects on these participating couples. Among those who attended at least one group session, there were no statistically significant effects on the eight key relationship outcomes examined. Among the smaller group of couples who attended at least half of the group sessions offered (29 percent of those offered services), there was no strong evidence of effects on relationship outcomes, with one exception. BSF appears to have increased the likelihood that these couples would be living together (married or unmarried) at the 15-month follow-up—with an impact on this outcome of 7 to 10 percentage points.

Two cautions should be noted about interpreting these results. First, they were generated using quasi-experimental estimation techniques. Therefore, they provide less rigorous evidence of program effectiveness than the results reported in the BSF 15-month impact report, which were based on the evaluation's random assignment research design. Second, the TOT estimates were generated using only the outcomes of couples who attended group sessions (or appeared likely to attend based on their initial characteristics) and not the outcomes of all BSF applicants. BSF couples who attended group sessions differ in clearly observable ways from those who did not. In particular, the approximately 3 in 10 BSF couples who attended at least half the group sessions are older, less economically disadvantaged, and have higher levels of relationship commitment than couples who attended sessions less frequently or not at all. The BSF program model may have been more successful with couples with these characteristics, so one cannot assume that if all BSF applicants had attended group sessions at this level, the program would have had the same effect on co-residence for all couples that it did for those who chose to attend a substantial number of group sessions.



## References

- Benjamini, Y., and Y. Hochberg. "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing." *Journal of the Royal Statistical Society, Series B (Methodological)*, vol. 57, no. 1, 1995, pp. 289–300.
- Bloom, Howard S. "Accounting for No-Shows in Experimental Evaluation Designs." *Evaluation Review*, vol. 8, April 1984, pp. 225-246.
- Dion, M. Robin, Sarah A. Avellar, Elizabeth Clary. "The Building Strong Families Project: Implementation of Eight Programs to Strengthen Unmarried Parent Families." Washington, DC: Mathematica Policy Research, 2010.
- Garvey, Christine, Wrenetha Julion, Louis Fogg, Amanda Kratovil, and Deborah Gross. "Measuring Participation in a Prevention Trial with Parents of Young Children." *Research in Nursing & Health*, vol. 29, no. 3, May 2006, pp. 212-222.
- Imbens, Guido W. and Jeffrey M. Wooldridge. "Recent Developments in the Econometrics of Program Evaluation." *Journal of Economic Literature*, vol. 47, March 2009, pp. 5-86.
- McCurdy, Karen, and Deborah Daro. "Parent Involvement in Family Support Programs: An Integrated Theory." *Family Relations*, vol. 50, no. 2, April 2001, pp. 113-121.
- Myers, H., K. Alvy, A. Arrington, M. Richardson, M. Marigna, R. Huff, M. Main, and M. Newcomb. "The Impact of a Parent Training Program on Inner-City African-American Fathers." *Journal of Community Psychology*, vol. 20, no. 2, 1992, pp. 132-147.
- Schochet, Peter Z. "An Approach for Addressing the Multiple Testing Problem in Social Policy Impact Evaluations." *Evaluation Review*, vol.33, December 2009, pp. 539-567.
- Schochet, Peter Z., and John Burghardt. "Using Propensity Scoring to Estimate Program-Related Subgroup Impacts in Experimental Program Evaluations." *Evaluation Review*, vol. 31, no. 2, April 2007, pp. 95-120.
- Schochet, Peter Z., and Hanley Chiang. "Technical Methods Report: Estimation and Identification of the Complier Average Causal Effect Parameter in Education RCTs." Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education, April 2009.
- Wood, Robert G., Sheena McConnell, Quinn Moore, Andrew Clarkwest, and JoAnn Hsueh. "Strengthening Unmarried Parents' Relationships: The Early Impacts of Building Strong Families." Princeton, NJ: Mathematica Policy Research, May 2010.
- Wood, Robert G., Quinn Moore, Andrew Clarkwest, JoAnn Hsueh, and Sheena McConnell. "Strengthening Unmarried Parents' Relationships: The Early Impacts of Building Strong Families. Technical Supplement." Princeton, NJ: Mathematica Policy Research, May 2010.



## **APPENDIX A**

### **TECHNICAL DOCUMENTATION FOR THE BSF TREATMENT-ON-THE-TREATED ANALYSIS**





## TECHNICAL DOCUMENTATION FOR THE BSF TREATMENT-ON-THE-TREATED ANALYSIS

The main body of this report provides a brief description of the methods used to estimate the effects of the Building Strong Families (BSF) program for couples who attended group sessions and a presentation of the overall results of these treatment-on-the-treated (TOT) analyses. This appendix provides a more detailed explanation of the TOT methodology used in this analysis, as well as additional results not presented in the main body of the report. The first section of the appendix discusses alternative approaches to the propensity score-based methods to TOT impact estimation presented in this report. The next section discusses propensity scores and how they are calculated for the BSF TOT analyses. Next follows a description of the criteria for determining which analyses of BSF programs or groups of programs are presented in the report. That discussion is followed by details of (1) how the impact analyses were carried out, (2) how the impact estimates were adjusted for multiple comparisons, and (3) the results of sensitivity tests produced using alternative levels of frequent attendance. The final section describes results for groups of programs defined by the curriculum used.

### Alternative Approaches to Estimating TOT Impacts

The TOT impact estimates presented in this report were produced using propensity score-based techniques. As described in the main body of the report, the analysis relied on two TOT estimation techniques: (1) a *traditional matching approach*, in which each BSF couple that attended a group session was matched to a similar control group couple based on their propensity scores and (2) a *“likely attender” approach*, in which the set of couples within each research group who appeared most likely to attend based on their propensity scores were identified and their outcomes were compared to measure effects.

A simple alternative to estimating TOT impacts would be to compare the outcomes of treatment group members who attended sessions to those of all control group members. However, this approach would not produce valid estimates because couples who attended group sessions are not representative of the full set of couples who applied for BSF. Table A.1 illustrates some of the important ways in which attendees in the BSF group differ from the full set of control group members. Prior to enrollment, couples who went on to attend at least one session, on average, were more likely to be married or living together full time, had higher levels of relationship commitment, were older, and earned higher incomes. Those differences are even larger if all control group couples are compared to BSF group couples who attended at least half of group sessions. Based on their more favorable relationship and demographic characteristics, couples who attended group sessions would be expected to have better relationship outcomes than would the control group as a whole, even in the absence of the group education intervention. Therefore, in a comparison of the outcomes of BSF group attenders to all control group couples, it would not be possible to distinguish differences due to program services from differences in other pre-existing traits.

Another approach to estimating TOT impacts is the Bloom correction, which involves a simple adjustment to the ITT impact estimates. Specifically, it inflates the experimentally obtained ITT estimates by the inverse of the proportion of treatment group members who actually received

Table A.1. Baseline Characteristics of BSF Couples Who Attended Group Sessions Compared with the Characteristics of All Control Group Couples (percentages unless otherwise indicated)

	BSF Group Couples		All Control Group Couples
	Who Attended at Least One Group Session	Who Attended Half of Group Sessions	
<b>Relationship Characteristics</b>			
Couple's Relationship Status			
Married to each other	8.1	9.0*	6.8
Unmarried, cohabiting full-time	61.2**	64.0***	57.3
Unmarried, not cohabiting full-time	30.7***	27.0***	36.0
Relationship Quality (average scale value; range = 1-4)			
Couple Interaction	3.36	3.37*	3.33
Commitment	3.19**	3.24***	3.15
Both Partners Expect to Marry	63.1***	66.7***	57.9
Baby Born Before BSF Application	42.5	41.9	44.0
Either Partner Has a Child from a Prior Relationship	47.7	45.5	47.0
Pregnancy Intendedness			
Intended by both partners	25.2	27.0	24.1
Wanted by both partners, but considered mistimed	52.4	50.8	53.2
Unwanted by at least one partner	22.4	22.2	22.7
<b>Socioeconomic and Demographic Measures</b>			
Race/Ethnicity			
Both partners are Hispanic	26.3	26.3	25.5
Both partners are Black, non-Hispanic	46.7	45.0	46.7
Both partners are White, non-Hispanic	11.4	11.8	11.8
All other couples	15.7	16.9	16.1
High School Diploma Receipt (excluding GEDs)			
Both partners have diploma	38.7	42.8**	37.6
One partner has diploma	36.5	36.5	36.9
Neither partner has diploma	24.7	20.7**	25.5
Average Age (in years)			
Mother's age	24.0***	24.3***	23.4
Father's age	26.5***	26.9***	25.8
Couples' Total Earnings in Past Year (\$)	21,353***	23,567***	19,831
Either Partner Received TANF or Food Stamps in Past Year	46.1	43.1	45.6
<b>Mental Health, Attitudes, and Religiosity</b>			
Either Partner Has Psychological Distress <sup>a</sup>	40.2	38.7	38.4
Both Partners Agree with the Statement, "It is better for children if parents are married"	61.3	60.5	59.7
Attendance at Religious Services			
Both attend more than monthly	26.7*	26.6	23.9
One attends more than monthly	29.6	28.1	28.4
Neither attends more than monthly	43.7**	45.3	47.8
<b>Sample Size</b>	<b>1,276</b>	<b>694</b>	<b>2,207</b>

Sources: BSF baseline information form and BSF eligibility form.

Note: The eight programs are weighted equally for these calculations.

<sup>a</sup> Psychological distress is assessed using the Kessler-6 scale.

\*\*\*/\*\*/\* Difference between BSF and control group means are statistically significant at the .01/.05/.10 levels.

GED = general equivalency diploma; TANF = Temporary Assistance for Needy Families.

services (Bloom 1984).<sup>4</sup> This calculation is based on the assumption that all impacts observed for the full treatment group were generated by the effect on those who actually received services from the program, and that the impacts of the program on nonparticipants (who received no services) are zero.

Although the Bloom adjustment is a rigorous way to examine impacts for couples who received any services, it is not well suited for this analysis because it cannot be used to estimate the effects of different types or amounts of services. Instead, it can adjust for the proportion of couples who received no service at all from the program. Although only 55 percent of couples attended group relationship skills sessions (the core program service), more than 90 percent received at least some type of service from the program, including assistance from a family coordinator or referral to a support service. This high participation rate means that TOT estimates produced using the Bloom adjustment that adjust for the proportion of couples who received no service from BSF are almost identical to the ITT estimates presented in the main report (Table A.2).

**Table A.2. Bloom-adjusted TOT Estimates of BSF's Impacts on Key Relationship Outcomes**

Outcome	Bloom-adjusted TOT Estimates					
	ITT Estimates		Received Any BSF Service		Attended at Least One Group Session	
	Impact	Effect Size	Impact	Effect Size	Impact	Effect Size
<b>Relationship Status</b>						
Still romantically involved (%)	-1.34	-0.057	-1.43	-0.061	-2.40	-0.102
Living together, married or unmarried (%)	0.00	0.000	0.00	0.000	0.00	0.000
Married (%)	-1.12	-0.038	-1.20	-0.041	-2.00	-0.068
<b>Relationship Quality</b>						
Relationship happiness	0.06	0.040	0.06	0.042	0.10	0.066
Support and affection	0.01	0.029	0.01	0.030	0.02	0.048
Use of constructive conflict behaviors	0.01	0.010	0.01	0.011	0.02	0.017
Avoidance of destructive conflict behaviors	0.03	0.048	0.03	0.051	0.05	0.081
Neither reports infidelity (%)	1.78	0.056	1.90	0.060	3.18	0.100

Sources: BSF 15-month follow-up survey and BSF management information system records.

Notes: TOT impacts were created by multiplying the ITT impact estimates by the inverse of the participation rate. Consistent with how impacts were calculated, the participation rates used for these adjustments were calculated weighting sites equally and are specific to the samples used to generate the impacts. For example, the participation rates used to adjust impacts for relationship happiness are for couples who were still together at followup. Impacts for receiving any BSF services were adjusted using participation rates that ranged from 94 to 96 percent. Impacts for attending at least one group session were adjusted using participation rates that ranged from 56 to 61 percent.

\*\*\*/\*\*/\* Statistically significant at the .01/.05/.10 level.

ITT = Intent to treat.

TOT = Treatment on the treated.

<sup>4</sup> When this kind of correction is used, the standard error of the impact estimate for participants should be inflated to account for the estimation error in the no-show rate. Therefore, Bloom-adjusted TOT estimates will have slightly higher *p*-values than corresponding ITT estimates, and may have lower levels of statistical significance in some cases. The difference is negligible in practice (Schochet and Chiang 2009).

If one makes the stronger assumption that BSF had no effect on the couples who did not attend group sessions but who received other help from the program, then one can use a similar adjustment to estimate the effects of BSF on couples who attended group sessions. In particular, one can adjust the ITT impacts by the inverse of the proportion of couples who attended group sessions. Table A.2 presents BSF's impacts adjusted in this way. This method yields results that are quite similar to the results from the traditional matching and likely attender methods. In particular, estimated impacts on couples who attended at least one group session are small (all effect sizes are no larger than 0.10) and statistically insignificant (Table A.2). If one made the even stronger (and less plausible) assumption that BSF had no effect on couples who attended fewer than half the group sessions and adjusted the impacts accordingly, estimated impacts (both positive and negative) would be larger (ranging from a 4 percentage point negative impact on romantic involvement to a 6 percentage point positive impact on fidelity) but remain statistically insignificant (not shown).

Since the Bloom adjustment is not well suited for estimating effects of particular types or amounts services, alternative analytical methods were required. As described in the main body of the report, the analysis team used two propensity score-based methods—a traditional propensity score matching approach and a likely attender subgroup approach—that permit estimation of impacts for different levels of services received. Specifically, the main results include the impact of attending at least one group session (“any attendance”) and the impact of attending at least half of group sessions (“frequent attendance”), estimates that are not possible using the Bloom correction. These approaches attempt to avoid bias by producing comparable treatment and control group subsamples for analysis. The section that follows describes the processes for producing the propensity scores that are crucial for both methods.

## Developing the Propensity Score Model

### Objectives of the Propensity Score Estimation

For this analysis, propensity scores were generated using statistical models that predicted the likelihood of attendance at group sessions based on pre-intervention characteristics of the couples. The credibility of the TOT estimates depends on how well the probability of participation can be estimated. Abstractly, the aim of the traditional propensity score matching approach is to compare the outcomes of BSF group couples who attended group sessions to those of the control group couples who would have attended sessions if the sessions had been offered to them. Those two groups would be comparable, on average, in both observable and unobservable characteristics. Because of the random assignment design, such a subset of fully comparable couples does exist in the control group. But because members of the control group were not offered services, it is not possible to know for certain which of them are fully comparable to the program group couples who attended the sessions. Instead, the analysis team developed models that predict the probability of attendance. When using the propensity scores in a traditional matching approach, models with greater predictive power provide more confidence that the treatment group attendees are truly comparable to the control group couples to whom they are matched.

The predictive power of the propensity score model is also important in the likely attender approach, though for slightly different reasons. The likely attender approach does not compare outcomes of actual treatment group attenders to those of control group couples with similar propensity scores, but rather compares outcomes of BSF group couples with propensity scores above a particular cutoff to those of control group couples with propensity scores above that same cutoff. Because the method compares outcomes of couples identified in the same way in both groups, and because selection is based solely on baseline characteristics, differences in outcomes

between the two groups that are too large to have occurred at random can be attributed fully to differences in receipt of the BSF intervention. The likely attender method does not have the same risk of bias due to unobserved differences between two groups that is present with traditional propensity score matching. However, if a substantial proportion of likely attenders are not actual attenders, the estimated TOT impacts will be attenuated. At the limit, if the models have no predictive power, the actual participation rates of likely attenders will be no different from those of the program group sample as a whole, and the TOT impact estimates will be expected to look like the ITT estimates.

### **Data Used in the Propensity Score Estimation**

The propensity score estimation required data on attendance at group sessions and pre-random assignment characteristics of couples. Data on attendance at group sessions were obtained from records from automated data systems kept by the local BSF programs. The Becoming Parents curriculum, which was used by the Oklahoma City BSF program, offered a total of 30 hours of group relationship skills education. The other programs used either the Loving Couples, Loving Children curriculum or the Love's Cradle curriculum, each of which offered a total of 42 hours of group sessions. Therefore, for measures of sessions attended as a percentage of sessions offered, the number of hours attended was divided by 30 for couples in Oklahoma City and by 42 for couples in the other programs.

In order to identify any important predictors of group session attendance, pre-intervention characteristics were selected from three sources: (1) the BSF eligibility form completed at the time of program application, (2) the baseline information form also completed at application, and (3) the 15-month survey. Although the 15-month survey contained predominantly post-enrollment outcomes, it did gather some additional pre-enrollment background data that were not gathered at sample intake, such as whether the sample members were immigrants, their criminal history, and their experiences of abuse during childhood.

### **Methods for Developing Propensity Score Models**

The first step in developing the propensity score model was to identify a wide range of possible predictors. This included creating various transformations of variables, including spouse-specific measures, couple averages, differences between partners' responses, binary versions of ordinal or cardinal measures, and interactions between variables. Other strategies included identifying patterns in responses that might be meaningful, such as missing responses to questions, and creating measures of contextual factors that might be predictive of attendance, such as season of enrollment.

The next step was to determine which variables from among that pool would be included in the propensity score models. Because traits that are not accounted for can only introduce bias into the impact estimates if they are associated both with receipt of treatment and with relationship outcomes, a subset of measures that have been demonstrated to be strongly associated with outcomes in prior analyses, such as baseline relationship status, relationship quality, race/ethnicity, and age, were included. Another set of measures that were strongly associated with both attendance indicators in the exploratory analyses were also added at this step.

After selecting this initial set of variables, the team selected additional predictors from the remaining pool of variables using a structured process designed to identify the variables most strongly predictive of attendance. First, a logistic regression model was estimated with an indicator of the designated level of attendance (any group sessions or at least half of group sessions) as the

dependent variable and the initial set of predictors and a set of indicators for the various BSF programs as independent variables. The residual from that model was then calculated and the correlations between that residual and each variable in the pool of candidate predictors were assessed. The candidate variable most strongly associated with the residual was then added to the prediction model and the process of calculating the residual and checking correlations with the remaining candidate variable was repeated. In order to identify all variables that reasonably added predictive power to the model, the process continued until three variables in a row were selected that had  $p$ -values above .25.<sup>5</sup> At that point, those final three variables were removed and all prior measures were retained.

This process was performed separately for predictors of attending any sessions and predictors of frequent attendance. Table A.3 presents the resulting lists of predictors selected to be included in each of the two propensity score models. The two lists overlap substantially, though some differences exist.

### Estimation of the Propensity Scores

Predicted participation estimates were based on data for BSF program group members in the analysis sample using a logistic regression model that can be represented as:

$$(1) \quad \Pr(\text{Participation}_i) = \Lambda(X_i\beta)$$

where  $\Lambda$  is the cumulative distribution function for the logistic distribution,  $X_i$  represents a set of pre-intervention characteristics for BSF group couple  $i$ , and  $\beta$  is a vector of parameters corresponding to those characteristics. The coefficients for each predictor were allowed to vary across programs. Results from that model were used to calculate each couple's probability of attending at all and of attending frequently (defined as attending half the scheduled sessions). This calculation was made for both program and control group members. Because coefficients varied by program, the influence of any particular variable in determining a couple's propensity score could vary depending on the local BSF program in which the couple was enrolled.

### Creation of Comparison Groups

The traditional matching and likely attender approaches used the same propensity scores. The central difference between the approaches is the way those scores were used to construct research groups. For the traditional matching approach, each BSF group couple who met the given attendance threshold was matched to the comparison group couple from the same site that had the

---

<sup>5</sup> The  $p$ -value 0.25 is more generous than is typically used in causal analyses. But because the concern here is with predictive value rather than causal attribution, there is relatively more concern for Type II than Type I error. Adding variables beyond those did little to improve the model's predictive power.

Table A.3. List of Predictor Variables Included in Propensity Score Models

Baseline Characteristic	Whether Characteristic Is Included in Model for	
	Attended at Least One Group Session	Attended at Least Half of Group Sessions
<b>Relationship Characteristics</b>		
Married or cohabiting full-time <sup>a</sup>	Yes	Yes
Relationship commitment – mother <sup>b</sup>	Yes	Yes
Relationship commitment – father <sup>b</sup>	Yes	Yes
Interaction quality – couple average <sup>c</sup>	Yes	Yes
Neither partner describes relationship as “steady”	Yes	Yes
Partners have at least one prior child together	Yes	Yes
Either partner has child from another relationship	No	Yes
Partners knew each other more than three years	Yes	No
Partners knew each other one year or less	Yes	No
<b>Socioeconomic and Demographic Measures</b>		
Both partners are Hispanic	Yes	Yes
Both partners are Black, non-Hispanic	Yes	Yes
All other non-White race/ethnicity	Yes	Yes
One partner has high school degree or GED	Yes	Yes
Both partners have high school degree or GED	Yes	Yes
Partners’ average age	Yes	Yes
Mother is employed, father is not	Yes	Yes
Mother sexually abused while a minor	Yes	No
Father’s earnings last year	No	Yes
Father lived with both parents up to age 18	Yes	Yes
Father ever jailed before random assignment	Yes	No
<b>Mental Health, Attitudes, and Religiosity</b>		
Father’s anxiety <sup>d</sup>	Yes	Yes
Mother’s pro-marriage attitudes <sup>e</sup>	Yes	Yes
Both mother and father strongly agree it is better for children if their parents are married	Yes	No
Mother’s frequency of religious attendance	Yes	Yes
<b>Study Enrollment Characteristics</b>		
Indicators for local BSF program	Yes	Yes
Enrolled during first trimester of pregnancy	Yes	Yes
Enrolled during second trimester of pregnancy	Yes	Yes
Enrolled between June and August	Yes	Yes
Enrolled between September and November	Yes	Yes
Father’s BIF date relative to mother’s <sup>f</sup>	Yes	No
Latest date BIF was completed by either partner	No	Yes
Any missing items on BIF	Yes	Yes

Sources: BSF baseline information form, BSF eligibility form, and BSF 15-month follow-up survey.

<sup>a</sup> Both partners report being married or living together full-time.

<sup>b</sup> Value on index of four items gauging aspects of relationship commitment.

<sup>c</sup> Value on index of five items gauging aspects of the quality of couple interaction.

<sup>d</sup> Index of scale gauging respondent’s level of nervousness, restlessness, and extent to which daily life is perceived to be an effort.

<sup>e</sup> Value on item gauging agreement with the statement that “It is better for children if their parents are married.”

<sup>f</sup> Days between father’s completion and mother’s completion of BIF (negative if mother completed first, positive if father did).

BIF = baseline information form; GED = general equivalency diploma.



most similar propensity score.<sup>6</sup> This matching was performed separately for the *any attendance* and *frequent attendance* analyses, using the propensity scores that were generated for that particular attendance level. Under this “nearest neighbor” matching approach, it is possible for the same comparison group couple to be matched to more than one treatment group couple. Among program group members in the analysis sample, 1,274 couples attended at least one session. A total of 833 control group couples were matched to one or more of the BSF couples who attended group sessions. The difference between the two counts resulted from control group couples who were matched to more than one program couple. When a control group couple was matched more than once, the couple received correspondingly greater weight in the analysis, so that the weighted total sample size was the same for the program and control groups. The corresponding sample sizes for the frequent attender analyses are 694 program group couples and 500 control group couples who were matched to them.

Although this matching approach can produce groups that are comparable on observed characteristics, it is not possible to guarantee that they are comparable on traits that are not observed. Differences in unobservable traits can lead to bias in the TOT estimates generated using the traditional matching approach. For example, couples who attend sessions may have unmeasured pre-intervention traits that lead both to session attendance and better relationship outcomes. An example of such a trait could be a couple’s motivation to improve their relationship. It is also possible that attendance results from relationship outcomes, since couples who break up early on in the study are likely to discontinue attending group sessions. Therefore, attenders could have better outcomes than non-attenders because of the fact that relationship trouble caused the latter group not to attend. Because unobserved traits of attenders that influence their attendance would also be expected to produce more positive relationship outcomes, any bias resulting from such unobserved traits in the traditional matching method is likely to be in the direction of making the estimates of the impact of group attendance somewhat larger than the actual impacts.

Whereas the traditional approach matched each attender couple from the BSF group to a single couple from the control group, the likely attender method created subgroups of couples with high propensity scores within both the BSF and control groups. This approach creates groups solely on the basis of pre-intervention characteristics, which avoids the threat of bias introduced by unobservable traits that is present in traditional matching approaches. The likely attender cutoff value was set such that the number of BSF group couples above the cutoff was the same as the actual number of attenders. The cutoff was set program-by-program, so the numbers of likely attenders and actual attenders were identical within each program.

Of course, predictive power was not perfect and not all likely attenders were actual attenders. But, as shown in Table A.4, rates of attendance among those identified as likely attenders were much higher than for all couples. Overall, 55.1 percent of couples attended at least one session. Among couples designated as likely attenders using the approach described above, that rate was 73.6 percent. Among couples not categorized as likely attenders, only about half as many (35.4 percent)

---

<sup>6</sup> Some propensity score matching methods involve matching more than one comparison group member to each program group attender. Although these methods provide larger comparison samples and thus greater statistical power than single match methods (such as the one used in the analyses presented in this report), the single match approach leads to more credible estimates because they minimize bias (Imbens and Wooldridge, 2009). Since minimizing bias in TOT estimates is the central concern for this analysis, the analysis team chose to use an approach based on “nearest neighbor” single match with replacement.



attended at least one session. A similar pattern is observed for attending half or more of all sessions. Overall, 28.9 percent of couples had that level of attendance, but among likely attenders the rate was more than double (61.9 percent). The results show that the model did have substantial predictive power, but that power is far from perfect, which would lead to the expectation that estimates generated using this method might underestimate the true impact of attendance at group sessions somewhat.

**Table A.4. Actual Attendance Rates of Program Group Couples, by Likely Attender Status (Percentages)**

Level of Attendance	All BSF Group Couples	BSF Couples Included in Likely Attender Analysis	BSF Couples Excluded from Likely Attender Analysis
Attended at Least One Session	55.1	73.6	35.4
Attended at Least Half of Sessions	28.9	61.9	17.1

Sources: BSF baseline information form, BSF eligibility form, BSF 15-month follow-up survey, BSF management information system records.

Note: Likely attenders consist of program group couples with the highest propensity scores. The cutoff value is set site by site such that the number of likely attenders in each site is equal to the number of actual attenders. The “All BSF Group Couples” attendance rates are for all randomized couples. The rates for likely attenders are calculated using the 15-month analysis sample, with nonresponse weights to make that sample representative of the full baseline sample.

If the propensity score methods work properly, they should lead to well-matched research groups. Tables A.5 and A.6 show that the BSF and comparison group couples used in the TOT analyses are well matched on observable characteristics. Differences in mean characteristics are small for samples used both in the traditional matching and likely attender analyses. That is true for both the any attendance and frequent attendance analyses. Across the four samples used in the TOT analysis and among the wide range of baseline characteristics examined, there are only three marginally significant between-group differences. Because of the way that groups were formed in the likely attender method, it is almost certainly true that the groups are similar on unobservable traits as well. As noted earlier, with the traditional matching method it is not possible to assume that the groups are equivalent on traits that are not observed.

### Standards for Determining When TOT Estimation Was Appropriate

As discussed above, if the research samples identified for the analysis samples are not sufficiently well matched, then using them to estimate TOT impacts may be misleading and inappropriate. The TOT analysis must also contend with issues related to sample size constraints, which seriously limit the ability to detect meaningful impacts for individual BSF programs. Program-level research samples are quite small in some cases. TOT analysis requires limiting these samples further, since the analyses focus only on estimating impacts for those couples who attended group sessions, or attended them regularly. Because group attendance rates were relatively low in many programs, restricting the analysis to couples who attended group sessions reduces sample sizes substantially, resulting in limited statistical power at the program level. TOT impact estimates need to be calculated with a reasonable degree of precision to be meaningful and informative; therefore, these estimates may not be appropriate in cases with very low statistical power.

Table A.5. Baseline Characteristics of BSF and Control Group Couples Included in the Analysis of Those Who Attended at Least One Group Session (percentages unless otherwise indicated)

	Traditional Matching Method		"Likely Attender" Method	
	BSF Group	Control Group	BSF Group	Control Group
<b>Relationship Characteristics</b>				
Couple's Relationship Status				
Married to each other	8.1	8.4	9.1	8.4
Unmarried, cohabiting full-time	61.1	58.3	63.7	61.4
Unmarried, not cohabiting full-time	30.8	33.4	27.2	30.2
Relationship Quality (average scale value; range = 1-4)				
Couple Interaction	3.36	3.33	3.36	3.35
Commitment	3.19	3.18	3.23	3.21
Both Partners Expect to Marry	63.0*	58.5	63.6	61.4
Baby Born Before BSF Application	42.4	43.0	43.0	42.3
Either Partner Has a Child from a Prior Relationship	47.7	47.0	47.4	47.1
Pregnancy Intendedness				
Intended by both partners	25.2	23.3	25.3	23.8
Wanted by both partners, but considered mistimed	52.5	54.5	53.1	54.0
Unwanted by at least one partner	22.4	22.3	21.6	22.2
<b>Socioeconomic and Demographic Measures</b>				
Race/Ethnicity				
Both partners are Hispanic	26.3	25.2	26.3	26.1
Both partners are Black, non-Hispanic	46.7	46.8	44.7	44.3
Both partners are White, non-Hispanic	11.3	11.9	12.2	12.9
All other couples	15.6	16.1	16.8	16.7
High School Diploma Receipt (excluding GEDs)				
Both partners have diploma	38.7	38.9	40.8	39.9
One partner has diploma	36.6	36.8	36.9	38.0
Neither partner has diploma	24.8	24.3	22.3	22.2
Average Age (in years)				
Mother's age	24.0	23.6	24.2	24.0
Father's age	26.5	26.0	26.9*	26.4
Couples' Total Earnings in Past Year (\$)	21,345	20,416	21,804	20,842
Either Partner Received TANF or Food Stamps in Past Year	46.1	45.4	44.8	44.6
<b>Mental Health, Attitudes, and Religiosity</b>				
Either Partner Has Psychological Distress <sup>a</sup>	40.3	39.8	40.1	41.4
Both Partners Agree with the Statement, "It is better for children if parents are married"	61.3	60.4	62.4	59.0
Attendance at Religious Services				
Both attend more than monthly	26.7	26.8	27.8	27.3
One attends more than monthly	29.6	29.2	30.2	30.3
Neither attends more than monthly	43.7	44.0	41.9	42.5
<b>Sample Size</b>	<b>1,276</b>	<b>833</b>	<b>1,276</b>	<b>1,250</b>

Sources: BSF baseline information form and BSF eligibility form.

Note: The eight programs are weighted equally for these calculations.

<sup>a</sup> Psychological distress is assessed using the Kessler-6 scale.

\*\*\*/\*\*/\* Difference between BSF and control group means are statistically significant at the .01/.05/.10 levels.

GED = general equivalency diploma; TANF = Temporary Assistance for Needy Families.

Table A.6. Baseline Characteristics of BSF and Control Group Couples Included in the Analysis of Those Who Attended at Least Half of Group Sessions (percentages unless otherwise indicated)

	Traditional Matching Method		"Likely Attender" Method	
	BSF Group	Control Group	BSF Group	Control Group
<b>Relationship Characteristics</b>				
Couple's Relationship Status				
Married to each other	9.0	9.2	11.2	11.0
Unmarried, cohabiting full-time	64.0	61.2	67.6	66.6
Unmarried, not cohabiting full-time	27.0	29.6	21.2	22.4
Relationship Quality (average scale value; range = 1-4)				
Couple Interaction	3.37	3.35	3.41	3.37
Commitment	3.24	3.24	3.32	3.29
Both Partners Expect to Marry	66.7	62.9	68.1	65.4
Baby Born Before BSF Application	41.8	42.5	40.8	41.5
Either Partner Has a Child from a Prior Relationship	45.5	41.7	44.9	40.5
Pregnancy Intendedness				
Intended by both partners	26.8	23.9	27.1	24.8
Wanted by both partners, but considered mistimed	50.9	54.8	52.9	52.5
Unwanted by at least one partner	22.3	21.3	20.0	22.7
<b>Socioeconomic and Demographic Measures</b>				
Race/Ethnicity				
Both partners are Hispanic	26.2	26.9	25.3	26.4
Both partners are Black, non-Hispanic	45.1	46.8	43.6	44.1
Both partners are White, non-Hispanic	11.8	11.7	12.7	13.9
All other couples	16.9	14.6	18.3	15.6
High School Diploma Receipt (excluding GEDs)				
Both partners have diploma	42.8	43.3	45.8	48.4
One partner has diploma	36.5	37.7	37.0	36.4
Neither partner has diploma	20.7	19.0	17.1	15.2
Average Age (in years)				
Mother's age	24.3	24.0	24.6	24.5
Father's age	26.9	26.4	27.6*	26.9
Couples' Total Earnings in Past Year (\$)	23,579	22,293	26,021	25,561
Either Partner Received TANF or Food Stamps in Past Year	43.2	40.9	40.9	41.2
<b>Mental Health, Attitudes, and Religiosity</b>				
Either Partner Has Psychological Distress <sup>a</sup>	38.6	38.3	35.2	38.9
Both Partners Agree with the Statement, "It is better for children if parents are married"	60.6	58.5	57.8	56.7
Attendance at Religious Services				
Both attend more than monthly	26.6	25.7	26.2	28.4
One attends more than monthly	28.2	27.4	28.7	25.1
Neither attends more than monthly	45.2	46.9	45.1	46.6
<b>Sample Size</b>	<b>694</b>	<b>500</b>	<b>694</b>	<b>711</b>

Sources: BSF baseline information form and BSF eligibility form.

Note: The eight programs are weighted equally for these calculations.

<sup>a</sup> Psychological distress is assessed using the Kessler-6 scale.

\*\*\*/\*\*/\* Difference between BSF and control group means are statistically significant at the .01/.05/.10 levels.

GED = general equivalency diploma; TANF = Temporary Assistance for Needy Families.

In response to these two challenges, the following two criteria were applied to determine whether it is appropriate to calculate TOT impacts:

- 1. Adequate statistical power.** For an impact analysis to be useful, it needs to have adequate statistical power to detect meaningful program effects. Included analyses were required to be able to detect an effect size of 0.30 as statistically significant at the 5 percent level. This standard for adequate statistical power is similar to the level of statistical power of the ITT impact estimates for the smallest BSF program, as presented in the 15-month BSF impact report.<sup>7</sup>
- 2. Evidence of well-matched research groups.** Confidence in the TOT estimates requires evidence that the research samples used to estimate them are well matched on initial characteristics. The standards used to assess this evidence follow standards for baseline equivalence established by the What Works Clearinghouse, an initiative of the Department of Education that assesses the rigor of research evidence. In particular, samples were required to be well matched on five baseline measures of relationship status and quality (marriage, full-time co-residence, less than full-time co-residence, relationship commitment scale, and relationship interaction scale) with no differences between the research groups on these measures that are larger than a quarter of a standard deviation.<sup>8</sup>

As shown in Table A.6, the overall TOT analysis averaged across all programs easily meets the adequate statistical power criterion; all these TOT analyses are able to detect effect sizes at least half as small as those required. The second criterion is also met as there are no large differences in the key baseline characteristics of any of the relevant research samples.

The inclusion criteria are generally not met for the TOT analyses related to individual BSF programs, primarily because the sample sizes for individual programs are too small to provide adequate statistical power (Table A.7). Only Atlanta and Oklahoma City have adequate statistical power to conduct a program-level TOT analysis of the effect of attending at least one group session. However, the Atlanta TOT sample does not meet the standard of having well-matched research groups for this analysis because there are large initial differences between the research groups on some relationship measures. In addition, only Oklahoma City has adequate statistical power for a program-level TOT analysis of the effect of attending at least half of group sessions. Therefore, the only program that is able to meet the inclusion criteria for all TOT analyses is the Oklahoma City program, which is the largest program and the program with the highest rate of attendance.

Because it is generally not possible to calculate meaningful and informative TOT impact estimates at the program level, analysis was conducted for groups of programs defined by the

---

<sup>7</sup> The minimum detectable effect size for the ITT analysis of the 15-month data for the San Angelo program, the smallest BSF program, was 0.29.

<sup>8</sup> Although most of the main relationship status and quality outcomes are defined for all couples, some are defined only for those who remain in a romantic relationship, or for those who remain in a relationship or in contact (Wood, Moore, et al. 2010). Therefore, research samples were required to be well matched for each of these three types of couples.

Table A.7. Minimum Detectable Effect Sizes and Whether Statistical Power is Adequate, by BSF Program and Analysis Sample

Program	At Least One Group Session Analysis			At Least Half of Group Sessions Analysis		
	MDE of Traditional Method	MDE of “Likely Attender” Method	Adequate Statistical Power?	MDE of Traditional Method	MDE of “Likely Attender” Method	Adequate Statistical Power?
Overall	0.10	0.11	Yes	0.13	0.15	Yes
BSF Program						
Atlanta	0.27	0.29	Yes	0.36	0.38	No
Baltimore	0.32	0.35	No	0.55	0.57	No
Baton Rouge	0.33	0.35	No	0.50	0.53	No
Florida counties	0.29	0.32	No	0.57	0.58	No
Houston	0.33	0.38	No	0.61	0.71	No
Indiana counties	0.30	0.34	No	0.36	0.50	No
Oklahoma City	0.19	0.22	Yes	0.21	0.24	Yes
San Angelo	0.35	0.40	No	0.50	0.53	No
BSF Curriculum						
<i>Becoming Parents</i>	0.19	0.22	Yes	0.21	0.24	Yes
<i>Love’s Cradle</i>	0.24	0.27	Yes	0.38	0.42	Yes
<i>LCLC</i>	0.13	0.15	Yes	0.20	0.21	No

Sources: BSF 15-month follow-up survey and BSF management information system records.

LCLC = Loving Couples, Loving Children; MDE=minimum detectable effect.

curriculum that they used. This system creates three groups of programs: (1) those that used the *Becoming Parents* curriculum (only Oklahoma City); (2) those that used the *Love’s Cradle* curriculum (Houston and San Angelo); and those that used the *Loving Couples, Loving Children* curriculum (Atlanta, Baltimore, Baton Rouge, Florida, and Indiana). These program-level groupings have the advantage of substantially increasing statistical power, which makes TOT analysis more feasible (Table A.7). For TOT analysis of the effect of attending at least one group session, all three curriculum groupings meet both inclusion criteria. For TOT analysis of the effect of attending at least half of group sessions, only *Loving Couples, Loving Children* and *Becoming Parents* meet the standards. The *Love’s Cradle* curriculum group has inadequate statistical power for this latter analysis.

## Estimating TOT Impacts

After using the propensity score models to determine research samples for each set of analyses, TOT impact estimates were calculated using methods similar to those used to calculate IIT estimates in the 15-month impact report (Wood, Moore, et al. 2010). Specifically, impact estimates were calculated using weighted least squares regression models. For the likely attender method, the analysis sample includes couples from the BSF and control groups who were likely to have had a given level of group session attendance based on their baseline characteristics. The analysis weights used were assigned based on the couple’s probability of survey nonresponse, as in the IIT analysis presented in the 15-month impact report (Wood, McConnell, et al. 2010). For the traditional propensity score matching method, these models were estimated on a sample that included BSF group couples who had attended a given level of group sessions and their matched comparison group counterparts. Because comparison group couples were only included in the analysis on the basis of their match to a BSF group couple who attended group sessions at a given level, they were assigned the analysis weight of the BSF group couple to whom they were matched. Comparison

group couples who were matched to multiple BSF group attendees received the sum of the weights of the couples to whom they were matched.

All other aspects of the TOT impact estimation—including calculation of pooled impacts by weighting BSF programs equally, and choice of covariates to control for characteristics measured in the baseline survey—were the same as those used in estimating ITT impacts. A more detailed discussion of these methodological details is provided in the technical documentation to the 15-month impact report (Wood, Moore, et al. 2010).

## Adjusting Results for Multiple Comparisons

Examining effects on numerous outcomes increases the chance of falsely identifying an impact as statistically significant (Schochet 2009). The BSF research design included several elements meant to minimize this possibility. These elements were also incorporated into the TOT analysis. Measures that were taken to minimize multiple comparison issues in the 15-month BSF impact analysis include using a small set of key outcomes, determining which sets of findings are most important on the basis of summary indices, and conducting sensitivity tests that adjust for multiple comparisons.<sup>9</sup>

The main focus of the BSF 15-month impact analysis is a small set of key relationship outcomes identified prior to beginning the analysis in two domains: (1) relationship status and (2) relationship quality. The three key relationship status outcomes are: (1) whether the couple was still romantically involved at follow-up, (2) whether they were living together (married or unmarried), and (3) whether they were married to each other. The five key relationship quality outcomes are (1) relationship happiness, (2) support and affection, (3) use of constructive conflict behaviors, (4) avoidance of destructive conflict behaviors, and (5) fidelity. These measures are described in more detail in the technical documentation to the main 15-month impact report (Wood, Moore, et al. 2010). Using a small set of outcomes within each domain makes it less likely that statistically significant findings will emerge by chance. Selecting the key outcomes before beginning analysis prevents focusing the assessment of program effectiveness on outcomes that happen to emerge as statistically significant (or the perception that this may have been the case).

The interpretation of findings on these key outcomes involved a careful assessment of whether statistically significant impact estimates were isolated or part of a stronger pattern within their domains. To that end, the impact analysis team constructed indices that summarize the outcomes in the relationship status and quality domains. The relationship status index was generated by summing the three main relationship status measures; the relationship quality index was constructed by normalizing each of the five main relationship quality measures and then summing the normalized values. If statistically significant impacts are found on these indices, one can have more confidence that BSF had effects within this outcome domain and it is less likely that any statistically significant impacts found on individual measures included in the index are due to chance rather than the BSF program.

For couples who attended at least one group session, the TOT impact estimates are not statistically significant for the relationship status and quality indices using either the traditional or likely attender methods (Table A.8). This is consistent with the results presented in the body of this

---

<sup>9</sup> See Wood, Moore, et al. (2010) for more detail about how the issue of multiple comparisons was addressed in the BSF 15-month impact analysis.

Table A.8. Sign and Statistical Significance of TOT Impacts on Relationship Status and Quality Indices, by Type of TOT Analysis

Program	Relationship Status Index	Relationship Quality Index
<b>Couples Who Attended at Least One Group Session</b>		
Traditional Matching Method	o	o
Likely Attender Method	o	o
<b>Couples Who Attended at Least Half of Group Sessions</b>		
Traditional Matching Method	+	o
Likely Attender Method	o	o

Sources: BSF 15-month follow-up survey BSF management information system records.

+ + +/+ +/+ Statistically significant positive impact at the .01/.05/.10 level.

o Impact is not statistically significant.

- - -/- -/- Statistically significant negative impact at the .01/.05/.10 level.

TOT = treatment on the treated.

report, which show no statistically significant effects on any of the individual relationship status or quality measures among couples who attended at least one group session. When impacts on the relationship status and quality indices are estimated for couples who attended at least half of the scheduled group sessions, there is some evidence of a statistically significant impact on relationship status, but none for relationship quality. The TOT impact estimate for the relationship status index is statistically significant at the 10 percent level using the traditional method, but is not statistically significant using the likely attender method. The impact analysis team also assessed whether significant findings on the key relationship status and quality measures were robust to statistical adjustments for multiple comparisons. These tests were conducted using the Benjamini-Hochberg method, which adjusts the thresholds at which p-values are considered statistically significant on the basis of the number of comparisons made in a given domain and the relative strength of each impact within the domain (Benjamini and Hochberg 1995).

The statistically significant positive effect of BSF on co-residence among couples who attended at least half the group sessions is robust to the Benjamini-Hochberg multiple comparison correction (Table A.9). However, the statistically significant effect on fidelity found using the traditional method for these couples is not.

**Table A.9. Statistical Significance of Key Outcomes Using Standard *p*-Value Thresholds and Thresholds Adjusted for Multiple Comparisons**

	Couple's Relationship Status			Couple's Relationship Quality <sup>a</sup>				
	Romantically Involved	Living Together (Married or Unmarried)	Married	Relationship Happiness <sup>b</sup>	Support and Affection	Use of Constructive Conflict Behaviors	Avoidance of Destructive Conflict Behaviors	Fidelity
<b>Couples Who Attended at Least One Group Session</b>								
<b>Traditional Method</b>								
Standard	o	o	o	o	o	o	o	o
Adjusted	o	o	o	o	o	o	o	o
<b>Likely Attender Method</b>								
Standard	o	o	o	o	o	o	o	o
Adjusted	o	o	o	o	o	o	o	o
<b>Couples Who Attended at Least Half of Group Sessions</b>								
<b>Traditional Method</b>								
Standard	o	++	o	o	o	o	o	+
Adjusted	o	+	o	o	o	o	o	o
<b>Likely Attender Method</b>								
Standard	o	++	o	o	o	o	o	o
Adjusted	o	+	o	o	o	o	o	o

Sources: BSF 15-month follow-up survey and BSF management information system records.

Notes: The adjustment for multiple comparisons used the Benjamini-Hochberg method, which adjusts the thresholds at which *p*-values are considered statistically significant on the basis of the number of comparisons made in a given domain and the relative strength of each impact within the domain.

<sup>a</sup> Details on the construction of relationship quality measures are provided in the technical supplement to the 15-month impact report (Wood, Moore, et al. 2010).

<sup>b</sup> Relationship happiness is measured only for couples who were still romantically involved at follow-up. In most cases, the initial characteristics of these couples in the two research groups were similar and comparing their outcomes was a valid measure of program impacts.

+++/++/+ Statistically significant positive impact at the .01/.05/.10 level.

o Impact is not statistically significant.

---/--/- Statistically significant negative impact at the .01/.05/.10 level.



## Sensitivity to Alternate Definitions of Frequent Attendance

One of the central goals of the TOT analysis is to assess the effect of BSF for couples who received a substantial amount of group relationship education. Making this assessment is more challenging than assessing TOT effects for those who received any amount of group relationship—while it is straightforward to identify couples who attended any group relationship education sessions, defining a “meaningful” level of group relationship education participation is subject to debate.

The main analysis defines frequent attendance as receiving at least half of the full curriculum. This corresponds to 15 hours of group relationship education in the Becoming Parents curriculum, and to 21 hours in the Loving Couples, Loving Children and Love’s Cradle curricula. This threshold for frequent attendance was selected before beginning the TOT analysis in order to avoid selecting definitions of frequent participation that happen to lead to statistically significant TOT effect estimates (or the perception that this may have been the case).

Sensitivity analyses of the main TOT results are based on two alternate definitions of frequent group attendance. Under the first alternate definition, the frequent attendance threshold is based on the number of hours of education received rather than the percentage of the curriculum received. Specifically, TOT impact estimates were calculated for couples who received at least 18 hours of group relationship education. This level of attendance is a midpoint between the attendance levels for Becoming Parents (15 hours) and Loving Couples, Loving Children/Love’s Cradle (21 hours) that are used in the primary definition of frequent attendance. Under the alternate definition, higher percentages of couples from programs using the Loving Couples, Loving Children and Love’s Cradle curricula are classified as frequent participants, while a lower percentage of couples in the Becoming Parents program are classified as frequent participants. Overall, 34 percent of BSF group couples are classified as frequent participants compared to 29 percent under the primary definition of frequent attendance.

The second alternate definition of frequent attendance sets a higher attendance threshold than was used in the primary definition. Specifically, TOT impact estimates were calculated for couples who received at least 75 percent of the full curriculum. This corresponds to 22.5 hours of group relationship education in the Becoming Parents curriculum, and to 31.5 hours in the Loving Couples, Loving Children and Love’s Cradle curricula. Under this definition, 22 percent of BSF group couples are classified as frequent participants.

As shown in Table A.10, results from the TOT analysis using alternate definitions are quite similar to those based on the primary definition.<sup>10</sup> When the definition of frequent attendance varies, the impact on living together remains statistically significant under all alternate specifications with the exception of the impact estimated using the likely attender method with the “at least 18 hours of attendance” threshold. No other relationship status impacts are statistically significant using either alternate TOT method with any definition of frequent attendance, with the exception of a positive, statistically significant finding on romantic involvement using the traditional method with the “at least three-quarters of group sessions” threshold.

---

<sup>10</sup> The TOT analyses using the two alternate thresholds meet the inclusion criteria related to adequate statistical power and well-matched research groups for the analysis pooled across BSF programs.

Table A.10. Statistical Significance of Key Outcomes Using Different Definitions of Frequent Group Session Attendance

	Relationship Status				Relationship Quality <sup>a</sup>					
	Index	Romantically Involved	Living Together (Married or Unmarried)	Married	Index	Relationship Happiness <sup>b</sup>	Support and Affection	Use of Constructive Conflict Behaviors	Avoidance of Destructive Conflict Behaviors	Fidelity
<b>Couples Who Attended at Least Half of Group Sessions</b>										
Traditional Method	+	o	++	o	o	o	o	o	o	+
Likely Attender Method	o	o	++	o	o	o	o	o	o	o
<b>Couples Who Attended at Least 18 Hours of Group Sessions</b>										
Traditional Method	o	o	++	o	o	o	o	o	o	o
Likely Attender Method	o	o	o	o	o	o	o	o	o	o
<b>Couples Who Attended at Least Three-Quarters of Group Sessions</b>										
Traditional Method	o	++	++	o	o	o	o	o	o	+
Likely Attender Method	o	o	+	o	o	o	o	o	o	o

Sources: BSF 15-month follow-up survey and BSF management information system records.

Notes: The adjustment for multiple comparisons used the Benjamini-Hochberg method, which adjusts the thresholds at which p-values are considered statistically significant on the basis of the number of comparisons made in a given domain and the relative strength of each impact within the domain.

<sup>a</sup>Details on the construction of relationship quality measures are provided in the technical supplement to the 15-month impact report (Wood, Moore, et al. 2010).

<sup>b</sup>Relationship happiness is measured only for couples who were still romantically involved at follow-up. In most cases, the initial characteristics of these couples in the two research groups were similar and comparing their outcomes was a valid measure of program impacts.

+++/++/+ Statistically significant positive impact at the .01/.05/.10 level.

o Impact is not statistically significant.

---/--/- Statistically significant negative impact at the .01/.05/.10 level.

When examining impacts on relationship quality, the TOT estimates are consistent across the three definitions of frequent attendance. The impact on the relationship quality index is not statistically significant using either TOT method with any definition of frequent attendance. Individual relationship quality measures are also not statistically significant, with the exception of positive significant findings for the fidelity outcome for the two definitions based on the percentage of curriculum received; the estimated impacts on fidelity are not significant using the 18-hour threshold.

## Estimating the Effect of an Additional Hour of Group Sessions

Another approach to estimating TOT impacts would be to use propensity score techniques to estimate the effect of attending each additional hour of group sessions. Under this approach, the impact of group session attendance would be evaluated as a continuous measure, rather than in binary categories such as the “any attendance” or “frequent attendance” categories used in this report.

When estimating the effect of an additional hour of BSF group session attendance, one would have to account for the likelihood that attending an additional hour may have different effects for couples with different levels of attendance. For example, TOT impacts may not be observed until a couple has reached a certain number of hours of group sessions or impacts may level off after a certain number of hours. A continuous TOT approach that allows for different effects for different levels of attendance can require a complex statistical model with a number of practical and technical drawbacks.

First, using a statistical model that estimates the impact of an additional hour of group attendance and accounts for the possibility of different effects at different levels of participation would require a very large research sample to have adequate statistical power to measure these different effects precisely. Moreover, using the available statistical power to measure BSF’s effects at many different participation levels limits the statistical power available to answer the basic question, “Did BSF have effects on couples who attended group sessions?”

Second, a model designed to measure impacts at different levels of participation would be based on a set of assumptions about the nature of the relationship between participation and program impacts. For example, the model could be based on the assumption that the effect of an additional hour of group sessions is the same for couples in certain ranges, such as those with less than five hours of attendance, those with six to ten hours of attendance, and so on. Alternatively, the models could allow for different effects for different levels of attendance in other ways, such as including in the statistical model measures of total hours of group sessions attended, the square of hours attended, and perhaps other higher order participation terms. These types of assumptions can be criticized as being arbitrary or reflecting idiosyncratic characteristics of the analysis sample that may not be generalizable to a broader population.

Third, assessing the predictive power of a continuous participation model and determining how the predictive power may influence the TOT estimates is more complicated than it is for the traditional matching and likely attender approaches used in this report. As with these approaches, the continuous TOT approach relies critically on the predictive power of the propensity model. For the traditional and likely attender approaches, one can assess the strength of the analysis by evaluating the differences in baseline characteristics for the BSF and comparison groups and by examining the overlap between couples who were actual attenders and those identified as likely

attenders by the propensity model. There is no straightforward analog to these assessments for the continuous TOT approach.

Finally, the results of the continuous TOT approach are less readily comparable to standard ITT impact estimates. Estimating TOT impacts using a continuous approach would provide an estimate of the effect of an additional hour of group attendance for couples at different levels of attendance, while the TOT impact estimates derived using the traditional matching and likely attender methods represent average effects for couples receiving certain specified amounts of group sessions. These latter estimates are comparable to the main ITT impacts, which represent average effects of the program for all couples who signed up for BSF.

For these reasons, the analysis presented in this report provides estimates of the effects of discrete “doses” of exposure to the program, specifically whether couples attended group sessions at all or attended at least half the sessions. It does not attempt to estimate the effects of each additional hour of group attendance.

### TOT Findings by BSF Curriculum Group

The TOT analysis aims to examine the effect of exposure to group sessions. In order to assess the impact of exposure to a particular BSF curriculum, the analysis team conducted TOT analyses in which BSF programs were grouped according to the curriculum used. These groupings include: (1) programs that used the *Becoming Parents* curriculum (only Oklahoma City); (2) those that used the *Love’s Cradle* curriculum (Houston and San Angelo); and those that used the *Loving Couples, Loving Children* curriculum (Atlanta, Baltimore, Baton Rouge, Florida, and Indiana).

Tables A.11 to A.13 present standard ITT estimates for the three curriculum groupings. Tables A.14 to A.16 present TOT estimates for the effect of attending at least one group session for each of the three curriculum groups. Tables A.17 and A.18 present the TOT estimates for the effect of attending at least half of the group sessions for the two curriculum groups for which analysis can be conducted.<sup>11</sup> The curriculum-level TOT results are broadly similar to the curriculum-level ITT results. Details about these results are as follows:

- ***Becoming Parents (Oklahoma City)***. The pattern of consistent positive effects found in the ITT results is broadly similar to the pattern of effects found in the TOT results for both the “any attendance” and “frequent attendance” analyses (Tables A.11, A.14, and A.17). In some cases, the TOT effects on relationship status are a bit larger than the ITT effects, while the TOT effects on relationship quality are in some cases a bit smaller than the ITT estimates. In addition, because of the lower levels of statistical power in the TOT analysis, fewer of the TOT effects are statistically significant than in the ITT analysis.
- ***Love’s Cradle (Houston and San Angelo)***. In the ITT analysis, the two *Love’s Cradle* programs have no statistically significant effects on the eight primary relationship outcomes (Table A.12). Similarly, the TOT analysis of the effect of attending at least one

---

<sup>11</sup> As noted above, the *Love’s Cradle* curriculum group has inadequate statistical power for the “at least half of group sessions” analysis.

group session yields almost no statistically significant impacts (Table A.15). The one exception is a statistically significant positive effect on living together at follow-up using the “likely attender” analysis. However, this effect is only statistically significant at the 10 percent level and does not remain statistically significant when corrected for multiple comparisons within the relationship status domain.

- ***Loving Couples, Loving Children (Atlanta, Baltimore, Baton Rouge, Florida, and Indiana)***. In the ITT analysis, the five *Loving Couples, Loving Children* programs have one statistically significant impact on a key relationship outcome—a positive impact on the proportion of couples remaining faithful (Table A.13). However, this impact is only statistically significant at the 10 percent level and is not robust to multiple comparison corrections within the relationship quality domain. For the TOT analyses, the impact on fidelity is only statistically significant using the traditional matching method for analysis of the effect of attending at least half of the group sessions (Table A.18). Similar to the ITT results, for all the TOT analyses, the largest estimated effects are on the fidelity measure.

**Table A.11. Impact of BSF on Key Outcomes at 15-Month Follow-up: *Becoming Parents Curriculum* (Oklahoma City)**

Outcome	BSF Group	Control Group	ITT Impact	p-Value	Effect Size
<b>Relationship Status</b>					
Still romantically involved (%)	81.5	76.4	5.1*	0.065	0.187
Living together, married or unmarried (%)	70.2	65.6	4.6	0.122	0.129
Married (%)	24.8	25.1	-0.3	0.923	-0.008
<b>Relationship Quality<sup>a</sup></b>					
Relationship happiness	8.51	8.20	0.31***	0.010	0.205
Support and affection	3.50	3.43	0.06**	0.025	0.157
Use of constructive conflict behaviors	3.33	3.22	0.11***	0.003	0.190
Avoidance of destructive conflict behaviors	2.80	2.71	0.09**	0.035	0.141
Neither reports infidelity (%)	82.0	77.2	4.9*	0.071	0.182
<b>Sample Size</b>	<b>435</b>	<b>442</b>			

Source: BSF 15-month follow-up survey.

Notes: Impacts are adjusted using a pooled regression controlling for the couple’s baseline relationship and demographic characteristics. Impact estimates are calculated based on a weighted average of program-level impacts in which all programs are weighted equally. The difference between BSF and control group means may not equal the estimated impact due to rounding.

\*\*\*/\*\*/\* Statistically significant at the .01/.05/.10 level.

<sup>a</sup> Details on the construction of relationship quality measures are provided in the technical supplement to the 15-month impact report (Wood, Moore, et al. 2010).

ITT = intent to treat.

**Table A.12. Impact of BSF on Key Outcomes at 15-Month Follow-up for All Couples: Love's Cradle Curriculum (Houston and San Angelo)**

Outcome	BSF Group	Control Group	ITT Impact	p-Value	Effect Size
<b>Relationship Status</b>					
Still romantically involved (%)	83.2	83.3	-0.1	0.963	-0.006
Living together, married or unmarried (%)	76.3	72.5	3.8	0.270	0.120
Married (%)	19.5	22.8	-3.3	0.226	-0.119
<b>Relationship Quality<sup>a</sup></b>					
Relationship happiness	8.70	8.58	0.11	0.246	0.075
Support and affection	3.49	3.47	0.01	0.635	0.036
Use of constructive conflict behaviors	3.32	3.31	0.02	0.683	0.030
Avoidance of destructive conflict behaviors	2.83	2.83	0.00	0.926	-0.007
Neither reports infidelity (%)	81.7	84.9	-3.3	0.271	-0.143
<b>Sample Size</b>	<b>330</b>	<b>316</b>			

Source: BSF 15-month follow-up survey.

Notes: Impacts are adjusted using a pooled regression controlling for the couple's baseline relationship and demographic characteristics. Impact estimates are calculated based on a weighted average of program-level impacts in which all programs are weighted equally. The difference between BSF and control group means may not equal the estimated impact due to rounding.

\*\*\*/\*\*/\* Statistically significant at the .01/.05/.10 level.

<sup>a</sup> Details on the construction of relationship quality measures are provided in the technical supplement to the 15-month impact report (Wood, Moore, et al. 2010).

ITT = intent to treat.

**Table A.13. Impact of BSF on Key Outcomes at 15-Month Follow-up: *Loving Couples, Loving Children Curriculum* (Atlanta, Baltimore, Baton Rouge, Florida Counties, and Indiana Counties)**

Outcome	BSF Group	Control Group	ITT Impact	p-Value	Effect Size
<b>Relationship Status</b>					
Still romantically involved (%)	72.1	74.6	-2.6	0.120	-0.079
Living together, married or unmarried (%)	54.3	55.8	-1.6	0.392	-0.038
Married (%)	13.7	14.5	-0.8	0.489	-0.038
<b>Relationship Quality<sup>a</sup></b>					
Relationship happiness	8.25	8.26	-0.01	0.896	-0.006
Support and affection	3.43	3.44	0.00	0.813	-0.010
Use of constructive conflict behaviors	3.22	3.21	0.01	0.541	0.023
Avoidance of destructive conflict behaviors	2.72	2.73	-0.01	0.744	-0.012
Neither reports infidelity (%)	70.6	67.5	3.1*	0.078	0.088
<b>Sample Size</b>	<b>1,452</b>	<b>1,449</b>			

Source: BSF 15-month follow-up survey.

Notes: Impacts are adjusted using a pooled regression controlling for the couple's baseline relationship and demographic characteristics. Impact estimates are calculated based on a weighted average of program-level impacts in which all programs are weighted equally. The difference between BSF and control group means may not equal the estimated impact due to rounding.

\*\*\*/\*\*/\* Statistically significant at the .01/.05/.10 level.

<sup>a</sup> Details on the construction of relationship quality measures are provided in the technical supplement to the 15-month impact report (Wood, Moore, et al. 2010).

ITT = intent to treat.

**Table A.14. Estimated Impact of BSF on Key Outcomes at 15-Month Follow-up for Couples Who Attended at Least One Group Session: *Becoming Parents Curriculum* (Oklahoma City)**

Outcome	Using Traditional Matching Method					Using “Likely Attenders” in Both Research Groups				
	BSF Group	Comparison Group	TOT Impact	p-Value	Effect Size	BSF Group	Comparison Group	TOT Impact	p-Value	Effect Size
<b>Relationship Status</b>										
Still romantically involved (%)	83.1	76.2	6.9	0.131	0.261	82.2	75.6	6.6**	0.036	0.240
Living together, married or unmarried (%)	74.3	67.0	7.2	0.168	0.212	72.0	66.0	6.0*	0.068	0.172
Married (%)	27.7	26.3	1.4	0.785	0.044	28.3	26.5	1.8	0.553	0.055
<b>Relationship Quality<sup>a</sup></b>										
Relationship happiness	8.46	8.12	0.34*	0.061	0.220	8.44	8.26	0.18	0.125	0.128
Support and affection	3.50	3.42	0.08*	0.081	0.190	3.50	3.48	0.01	0.641	0.036
Use of constructive conflict behaviors	3.33	3.21	0.13	0.116	0.224	3.30	3.24	0.06	0.117	0.113
Avoidance of destructive conflict behaviors	2.78	2.67	0.11	0.103	0.171	2.76	2.71	0.05	0.292	0.079
Neither reports infidelity (%)	83.9	76.2	7.8*	0.080	0.297	82.8	78.5	4.3	0.159	0.168
<b>Sample Size</b>	<b>336</b>	<b>202</b>				<b>336</b>	<b>354</b>			

Sources: BSF 15-month follow-up survey and BSF management information system records.

Notes: Impacts are adjusted using a pooled regression controlling for the couple’s baseline relationship and demographic characteristics. Impact estimates are calculated based on a weighted average of program-level impacts in which all programs are weighted equally. The difference between BSF and comparison group means may not equal the estimated impact due to rounding.

\*\*\*/\*\*/\* Statistically significant at the .01/.05/.10 level.

<sup>a</sup> Details on the construction of relationship quality measures are provided in the technical supplement to the 15-month impact report (Wood, Moore, et al. 2010).

TOT = treatment on the treated.



**Table A.15. Estimated Impact of BSF on Key Outcomes at 15-Month Follow-up for Couples Who Attended at Least One Group Session: *Love's Cradle* Curriculum (Houston and San Angelo)**

Outcome	Using Traditional Matching Method					Using "Likely Attenders" in Both Research Groups				
	BSF Group	Comparison Group	TOT Impact	p-Value	Effect Size	BSF Group	Comparison Group	TOT Impact	p-Value	Effect Size
<b>Relationship Status</b>										
Still romantically involved (%)	86.4	83.7	2.8	0.538	0.132	87.4	84.6	2.8	0.399	0.140
Living together, married or unmarried (%)	81.1	74.7	6.4	0.309	0.226	82.3	74.7	7.6*	0.069	0.275
Married (%)	21.8	24.6	-2.8	0.539	-0.095	22.8	25.4	-2.6	0.490	-0.085
<b>Relationship Quality<sup>a</sup></b>										
Relationship happiness	8.72	8.59	0.14	0.358	0.090	8.73	8.67	0.06	0.590	0.044
Support and affection	3.53	3.47	0.06	0.300	0.155	3.51	3.49	0.02	0.664	0.042
Use of constructive conflict behaviors	3.35	3.30	0.05	0.478	0.095	3.36	3.33	0.03	0.510	0.060
Avoidance of destructive conflict behaviors	2.86	2.80	0.06	0.450	0.084	2.87	2.85	0.01	0.794	0.023
Neither reports infidelity (%)	86.0	88.1	-2.2	0.661	-0.116	86.0	86.8	-0.8	0.838	-0.040
<b>Sample Size</b>	<b>222</b>	<b>137</b>				<b>222</b>	<b>200</b>			

Sources: BSF 15-month follow-up survey and BSF management information system records.

Notes: Impacts are adjusted using a pooled regression controlling for the couple's baseline relationship and demographic characteristics. Impact estimates are calculated based on a weighted average of program-level impacts in which all programs are weighted equally. The difference between BSF and comparison group means may not equal the estimated impact due to rounding.

\*\*\*/\*\*/\* Statistically significant at the .01/.05/.10 level.

<sup>a</sup> Details on the construction of relationship quality measures are provided in the technical supplement to the 15-month impact report (Wood, Moore, et al. 2010).

TOT = treatment on the treated.

**Table A.16. Estimated Impact of BSF on Key Outcomes at 15-Month Follow-up for Couples Who Attended at Least One Group Session: *Loving Couples, Loving Children* Curriculum (Atlanta, Baltimore, Baton Rouge, Florida Counties, and Indiana Counties)**

Outcome	Using Traditional Matching Method					Using “Likely Attenders” in Both Research Groups				
	BSF Group	Comparison Group	TOT Impact	p-Value	Effect Size	BSF Group	Comparison Group	TOT Impact	p-Value	Effect Size
<b>Relationship Status</b>										
Still romantically involved (%)	73.4	74.5	-1.2	0.699	-0.037	73.4	75.3	-1.8	0.420	-0.058
Living together, married or unmarried (%)	56.5	55.3	1.3	0.739	0.031	57.5	58.7	-1.2	0.634	-0.030
Married (%)	15.1	15.3	-0.2	0.933	-0.009	16.9	17.7	-0.8	0.643	-0.034
<b>Relationship Quality<sup>a</sup></b>										
Relationship happiness	8.17	8.19	-0.02	0.883	-0.012	8.21	8.21	0.00	0.976	-0.002
Support and affection	3.43	3.42	0.01	0.854	0.017	3.42	3.44	-0.01	0.605	-0.031
Use of constructive conflict behaviors	3.21	3.19	0.02	0.768	0.028	3.21	3.21	0.00	0.973	0.002
Avoidance of destructive conflict behaviors	2.68	2.72	-0.04	0.417	-0.059	2.68	2.71	-0.02	0.517	-0.036
Neither reports infidelity (%)	72.5	67.2	5.3	0.139	0.154	71.9	68.7	3.2	0.186	0.093
<b>Sample Size</b>	<b>718</b>	<b>490</b>				<b>718</b>	<b>699</b>			

Sources: BSF 15-month follow-up survey and BSF management information system records.

Notes: Impacts are adjusted using a pooled regression controlling for the couple’s baseline relationship and demographic characteristics. Impact estimates are calculated based on a weighted average of program-level impacts in which all programs are weighted equally. The difference between BSF and comparison group means may not equal the estimated impact due to rounding.

\*\*\*/\*\*/\* Statistically significant at the .01/.05/.10 level.

<sup>a</sup> Details on the construction of relationship quality measures are provided in the technical supplement to the 15-month impact report (Wood, Moore, et al. 2010).

TOT = treatment on the treated.

**Table A.17. Estimated Impact of BSF on Key Outcomes at 15-Month Follow-up for Couples Who Attended at Least Half of Group Sessions: *Becoming Parents Curriculum* (Oklahoma City)**

Outcome	Using Traditional Matching Method					Using “Likely Attenders” in Both Research Groups				
	BSF Group	Comparison Group	TOT Impact	p-Value	Effect Size	BSF Group	Comparison Group	TOT Impact	p-Value	Effect Size
<b>Relationship Status</b>										
Still romantically involved (%)	83.9	76.8	7.0	0.171	0.271	79.9	74.0	5.9*	0.091	0.203
Living together, married or unmarried (%)	75.9	65.6	10.3**	0.032	0.303	72.2	66.0	6.2*	0.075	0.176
Married (%)	30.4	27.1	3.3	0.545	0.097	28.1	28.2	-0.2	0.962	-0.005
<b>Relationship Quality<sup>a</sup></b>										
Relationship happiness	8.51	8.21	0.31	0.112	0.227	8.43	8.24	0.20	0.149	0.138
Support and affection	3.49	3.47	0.03	0.576	0.064	3.49	3.49	0.00	0.929	-0.008
Use of constructive conflict behaviors	3.33	3.23	0.09	0.176	0.156	3.29	3.24	0.05	0.255	0.096
Avoidance of destructive conflict behaviors	2.78	2.72	0.06	0.526	0.093	2.74	2.70	0.04	0.477	0.058
Neither reports infidelity (%)	85.2	79.1	6.1	0.170	0.254	83.6	79.0	4.6	0.163	0.182
<b>Sample Size</b>	<b>283</b>	<b>183</b>				<b>283</b>	<b>299</b>			

Sources: BSF 15-month follow-up survey and BSF management information system records.

Notes: Impacts are adjusted using a pooled regression controlling for the couple’s baseline relationship and demographic characteristics. Impact estimates are calculated based on a weighted average of program-level impacts in which all programs are weighted equally. The difference between BSF and comparison group means may not equal the estimated impact due to rounding.

\*\*\*/\*\*/\* Statistically significant at the .01/.05/.10 level.

<sup>a</sup> Details on the construction of relationship quality measures are provided in the technical supplement to the 15-month impact report (Wood, Moore, et al. 2010).

TOT = treatment on the treated.

**Table A.18. Estimated Impact of BSF on Key Outcomes at 15-Month Follow-up for Couples Who Attended at Least Half of Group Sessions: *Loving Couples, Loving Children* Curriculum (Atlanta, Baltimore, Baton Rouge, Florida Counties, and Indiana Counties)**

Outcome	Using Traditional Matching Method					Using “Likely Attenders” in Both Research Groups				
	BSF Group	Comparison Group	TOT Impact	p-Value	Effect Size	BSF Group	Comparison Group	TOT Impact	p-Value	Effect Size
<b>Relationship Status</b>										
Still romantically involved (%)	81.5	78.7	2.7	0.511	0.103	78.7	79.1	-0.4	0.927	-0.013
Living together, married or unmarried (%)	64.6	57.1	7.5	0.207	0.191	65.5	60.9	4.6	0.247	0.120
Married (%)	17.3	16.5	0.8	0.842	0.032	19.6	19.7	-0.1	0.984	-0.002
<b>Relationship Quality<sup>a</sup></b>										
Relationship happiness	8.35	8.22	0.13	0.465	0.095	8.37	8.33	0.03	0.792	0.024
Support and affection	3.47	3.43	0.04	0.424	0.092	3.48	3.48	0.00	0.928	-0.008
Use of constructive conflict behaviors	3.27	3.21	0.06	0.421	0.104	3.26	3.24	0.01	0.782	0.024
Avoidance of destructive conflict behaviors	2.69	2.69	0.00	0.976	0.003	2.76	2.75	0.01	0.833	0.017
Neither reports infidelity (%)	79.8	69.3	10.5**	0.023	0.338	80.5	75.7	4.8	0.228	0.170
<b>Sample Size</b>	<b>327</b>	<b>250</b>				<b>327</b>	<b>320</b>			

Sources: BSF 15-month follow-up survey and BSF management information system records.

Notes: Impacts are adjusted using a pooled regression controlling for the couple’s baseline relationship and demographic characteristics. Impact estimates are calculated based on a weighted average of program-level impacts in which all programs are weighted equally. The difference between BSF and comparison group means may not equal the estimated impact due to rounding.

\*\*\*/\*\*/\* Statistically significant at the .01/.05/.10 level.

<sup>a</sup> Details on the construction of relationship quality measures are provided in the technical supplement to the 15-month impact report (Wood, Moore, et al. 2010).

TOT = treatment on the treated.

**APPENDIX B**  
**SUPPLEMENTAL TABLES**



**Table B.1. Estimated Impact of BSF on Key Outcomes at 15-Month Follow-up for Couples Who Attended at Least One Group Session, Averaged Across All Programs**

Outcome	Using Traditional Matching Method					Using “Likely Attenders” in Both Research Groups				
	BSF Group	Comparison Group	TOT Impact	p-Value	Effect Size	BSF Group	Comparison Group	TOT Impact	p-Value	Effect Size
<b>Relationship Status</b>										
Still romantically involved (%)	77.8	77.2	0.6	0.785	0.022	77.8	77.8	-0.0	0.990	-0.001
Living together, married or unmarried (%)	64.8	61.7	3.1	0.279	0.081	65.4	63.7	1.7	0.361	0.046
Married (%)	18.4	19.0	-0.7	0.729	-0.027	19.6	20.8	-1.2	0.450	-0.043
<b>Relationship Quality<sup>a</sup></b>										
Relationship happiness	8.34	8.29	0.05	0.579	0.035	8.37	8.33	0.03	0.627	0.024
Support and affection	3.46	3.43	0.03	0.401	0.062	3.46	3.46	-0.00	0.927	-0.004
Use of constructive conflict behaviors	3.26	3.23	0.03	0.389	0.058	3.26	3.24	0.01	0.581	0.025
Avoidance of destructive conflict behaviors	2.73	2.74	-0.01	0.893	-0.008	2.74	2.75	-0.01	0.769	-0.013
Neither reports infidelity (%)	77.3	73.6	3.7	0.129	0.120	76.8	74.5	2.3	0.201	0.075
<b>Sample Size</b>	<b>1,276</b>	<b>823</b>				<b>1,276</b>	<b>1252</b>			

Sources: BSF 15-month follow-up survey and BSF management information system records.

Notes: Impacts are adjusted using a pooled regression controlling for the couple’s baseline relationship and demographic characteristics. Impact estimates are calculated based on a weighted average of program-level impacts in which all programs are weighted equally. The difference between BSF and comparison group means may not equal the estimated impact due to rounding.

\*\*\*/\*\*/\* Statistically significant at the .01/.05/.10 level.

<sup>a</sup> Details on the construction of relationship quality measures are provided in the technical supplement to the 15-month impact report (Wood, Moore, et al. 2010).

TOT = treatment on the treated.

**Table B.2. Estimated Impact of BSF on Key Outcomes at 15-Month Follow-up for Couples Who Attended at Least Half of Group Sessions, Averaged Across All Programs**

Outcome	Using Traditional Matching Method					Using “Likely Attenders” in Both Research Groups				
	BSF Group	Comparison Group	TOT Impact	p-Value	Effect Size	BSF Group	Comparison Group	TOT Impact	p-Value	Effect Size
<b>Relationship Status</b>										
Still romantically involved (%)	83.7	79.6	4.1	0.191	0.168	82.2	80.9	1.3	0.656	0.052
Living together, married or unmarried (%)	71.5	61.7	9.8**	0.045	0.267	71.9	64.8	7.1**	0.017	0.199
Married (%)	21.6	20.1	1.5	0.643	0.054	22.4	22.5	0.0	0.991	-0.001
<b>Relationship Quality<sup>a</sup></b>										
Relationship happiness	8.45	8.26	0.18	0.257	0.136	8.43	8.35	0.09	0.387	0.059
Support and affection	3.49	3.44	0.04	0.266	0.107	3.48	3.47	0.01	0.795	0.018
Use of constructive conflict behaviors	3.31	3.24	0.07	0.223	0.125	3.29	3.25	0.04	0.334	0.066
Avoidance of destructive conflict behaviors	2.74	2.71	0.03	0.606	0.045	2.76	2.75	0.02	0.673	0.027
Neither reports infidelity (%)	82.3	75.1	7.3*	0.072	0.264	83.0	79.6	3.5	0.208	0.138
<b>Sample Size</b>	<b>694</b>	<b>493</b>				<b>694</b>	<b>707</b>			

Sources: BSF 15-month follow-up survey and BSF management information system records.

Notes: Impacts are adjusted using a pooled regression controlling for the couple’s baseline relationship and demographic characteristics. Impact estimates are calculated based on a weighted average of program-level impacts in which all programs are weighted equally. The difference between BSF and comparison group means may not equal the estimated impact due to rounding.

\*\*\*/\*\*/\* Statistically significant at the .01/.05/.10 level.

<sup>a</sup> Details on the construction of relationship quality measures are provided in the technical supplement to the 15-month impact report (Wood, Moore, et al. 2010).

TOT = treatment on the treated.





---

**MATHEMATICA**  
Policy Research, Inc.

---

[www.mathematica-mpr.com](http://www.mathematica-mpr.com)

Improving public well-being by conducting high-quality, objective research and surveys  
Princeton, NJ ■ Ann Arbor, MI ■ Cambridge, MA ■ Chicago, IL ■ Oakland, CA ■ Washington, DC

Mathematica® is a registered trademark of Mathematica Policy Research