THESIS

IMPLEMENTATION EVALUATION OF BEST START FOR BABIES

Submitted by

Lindsay Hornell

Department of Human Development and Family Studies

In partial fulfillment of the requirements

For the Degree of Master of Science

Colorado State University

Fort Collins, Colorado

Spring 2014

Master's Committee:

    Advisor: David MacPhee

    J. Douglas Coatsworth
    Mary Khetani

ABSTRACT

IMPLEMENTATION EVALUATION OF BEST START FOR BABIES

One of the most important aims of evaluation science is to determine whether

interventions have a positive impact on the lives of their participants. A component of program

evaluation is to assess program implementation. This evaluation study assessed the

implementation of the Best Start for Babies/Best Start for Toddlers (BSB/T) program by

examining its fidelity to protocol, dosage, and participant engagement in relation to program

impact. BSB/T strives to promote positive child outcomes by providing parents with support,

education, and community referrals throughout the prenatal to toddler years. Facilitator logbooks

from the first three years of the BSB/T were evaluated. Program impacts were assessed through

use of (1) the Knowledge of Infant Development Inventory (KIDI), which measures a caregiver's

understanding of developmental milestones, strategies for child rearing, and knowledge of basic

child development; and (2) the Self-Perceptions of the Parental Role (SPRR), which assesses

parent self-efficacy and investment in the parental role. Both of these measures were

administered at the beginning and end of the program. By focusing on a program that has not yet

been systematically evaluated, I was able to determine how degree of fidelity, dosage, and

participant engagement contributed to BSB/T program outcomes. Results from this study suggest

that adhering to the curriculum is more important for certain activities, but not for other

activities. Findings on dosage, regarding how time was allocated in classes, suggest that

providing parents with more time to interact with one another was related to better program

outcomes. Finally, greater parental engagement was found to be associated with better program

outcomes – whereas parental resistance was associated with lower program outcomes. The

findings from this study offer current and future facilitators with valuable information on how to

best implement the BSB/T program.

TABLE OF CONTENTS

Introduction

One of the leading objectives of evaluation science is to determine whether interventions have a positive impact on the lives of their participants (Flay et al., 2005). In order to do this, it is essential to evaluate not just the outcomes of an intervention program but the degree to which it was implemented in accordance with a curriculum or blueprint. Evaluators differentiate between implementation evaluations (the means) and impact or summative evaluation (the end).

In a broad sense, implementation is what a program consists of as it is being delivered in a particular setting. Implementation data are gathered through various techniques that include facilitator logbooks, participant self-report accounts, and evaluator observations. This information is used to evaluate the delivery of a program as well as how participants receive a program (Durlak & DuPre, 2008). The purposes of this study are to (1) evaluate the implementation of Best Start for Babies/Toddlers, a recently developed parent education program aimed to promote positive child outcomes through parent support, and (2) to determine whether variations in program implementation are related to variations in the impact of Best Start for Babies/Toddlers. In the sections that follow, I will discuss the importance of implementation as well as particular dimensions of implementation. These dimensions include fidelity to protocol, dosage, and participant engagement. I will also review evidence that links implementation to program impact.

Literature Review

Implementation refers to how a program was delivered to a particular setting (Durlak & DuPre, 2008). Correct implementation has been found to be an important factor in achieving desired program outcomes. According to Durlak and DuPre (2008), if a program is not implemented correctly, negative program results can occur. In their review, Durlak and DuPre examined 59 articles related to assessing the impact of implementation on outcomes. They determined that careful implementation was associated with better program outcomes. In programs where the implementation was free of any problems, participants received much higher benefits from their participation (Durlak & DuPre, 2008). In another meta-analysis of program implementation and impact, Dubois, Holloway, Valentine, and Cooper (2002) found greater effect sizes in youth mentoring programs that monitored their implementation. In some cases, effect sizes were as much as three times greater for programs that monitored implementation versus programs that did not monitor implementation (Dubois et al., 2002). Furthermore, Kalafat, Illback, and Sanders (2007) performed an implementation analysis to assess how implementation relates to program outcomes. Looking specifically at a school-based family support programs, their findings showed a strong positive relation between program implementation, specifically fidelity to the intended program plan, and program-level outcomes achieved by participants (Kalafat et al., 2007).

Aside from implementation playing a major role in program outcomes, it also affects program evaluation. Durlak and DuPre suggested that without carefully collected implementation information, a program evaluation is not complete. It would be difficult to actually assess how the program was conducted and how to interpret the outcome data without such information. As well, replication programs would not have clear guidelines on how to

deliver the program effectively if systematic implementation data were not collected. Summerfelt (2003) asserted that it is almost impossible to assess the effectiveness of a program if the intervention failed to be implemented properly.

Despite its importance to program evaluation, flawless implementation is nearly impossible to achieve. Achieving this is in fact unrealistic. Instead, positive results from programs have been shown with levels of implementation around 60% (Durlak & DuPre, 2008). Thus, if a program does not follow the intended program administration flawlessly, participants may still benefit.

Assessing program implementation requires measuring several dimensions, including fidelity to protocol, dosage, and participant engagement. Without understanding the aspects of a program that contributed to its success, it would be difficult to replicate the program and have the same successful outcomes (Summerfelt, 2003). Therefore, it is crucial to document dosage, fidelity, and participant engagement, which will be described next.

Fidelity to Protocol

Explanation and reasons for evaluation. Fidelity is the extent to which a program is administered as originally intended (Dane & Schneider, 1998). Program fidelity occurs when the program is delivered in a comparable manner to all recipients and when the content delivery remains faithful to the program's theory and goals (Dumas, Lynch, Laughlin, Smith, & Prinz, 2001). Assessing program fidelity is a crucial step in program implementation evaluation. If a program is not delivered as it was intended, evaluation results could be meaningless (Kalafat, 2007). Also, without knowing how well a program adhered to its initial plan of intervention, it would be difficult to replicate the program to achieve the same results (Summerfelt, 2003).

Numerous benefits occur as a result of evaluating program fidelity. By knowing how well a program maintained fidelity to the original administrative intentions, it can help determine a program's overall impact. Dusenbury, Brannigan, Falco, and Hansen (2003) also explained how measuring program fidelity could provide crucial information related to the effects of any program adaptations that occurred. This information could lead to modifications to the program in the future. Additionally, evaluating program fidelity can prevent Type III error from occurring. Type III error, as described by Scanlon, Horst, Nay, Schmidt, and Waller (1977), transpires when a program is not properly implemented, leading to poor program outcomes. Instead of blaming the poor implementation, the program is seen as ineffective and as a result, could be terminated. By evaluating how well service providers maintained fidelity to a program, the likelihood of Type III error occurring decreases.

Fidelity in program evaluation. Evidence-based interventions commonly have a manual to describe the content and program guidelines. For example, in an effort to assess the ease of adhering to program guidelines, Cantu, Hill, and Becker (2010) evaluated the Strengthening Families Program. They found the curriculum guidelines allowed for consistent delivery across populations. Well-developed programs with distinct training protocols make adherence to program content easier for facilitators. The use of manuals, however, does not guarantee consistent and faithful delivery of the program (Forgatch, Patterson, & DeGarmo, 2005). Therefore, program delivery must be evaluated for fidelity.

Despite its importance, fidelity to a program's original intentions often is ignored in program evaluation when moving from controlled to real-world settings. Perepletchikova, Treat, and Katzdin (2007) found that in psychosocial interventions, only 3.5% of program evaluations actually acknowledged fidelity. Ennet et al. (2001) evaluated program fidelity in a real-world

setting of a program they delivered. The program examined was an in-school substance use prevention program. The researchers examined adherence to the curriculum in terms of the delivery of program content and delivery strategies. Results indicated that only one-third of the service providers delivered the entire curriculum according to the recommended schedule and only one-fourth of the service providers adhered to both the curriculum's content and delivery strategies. This indicates that despite the importance of adherence to program fidelity, program staff in real-world settings often do not make it a priority.

Fidelity-adaptation argument. In the field of program evaluation, specifically implementation evaluation, there is a discussion about the degree of fidelity necessary in program delivery. The relation between the degree of program fidelity and the amount of adaptation allowed while maintaining desired program outcomes is a widely debated topic. Numerous researchers believe it is imperative to strive for high fidelity with minor adaptations to population needs, believing that altering a program to fit participant needs can compromise program outcomes (Bauman et al., 1990; Domitrovich et al., 2010; Dumas et al., 2001; Kalfat et al., 2007). Failure to maintain fidelity can prevent outcome evaluation conclusions from being drawn. Without fidelity, it is nearly impossible to determine whether program outcomes are a result of the effectiveness of the intervention as designed, or the added/eliminated aspects of the intervention that result from program adaptation (Dumas et al., 2001). In fact, most researchers consider too much program adaptation as an implementation failure (Durlak & DuPre, 2008). With that said, Cantu et al. (2010) do acknowledge that some degree of adaptability is inevitable.

According to Castro, Barrera, and Martinez (2004), there are two types of adaptations that can occur in program implementation. The first adaptation is a modification of the form of program delivery. This includes making changes to the person delivering the program, the

channel of delivery (e.g., Internet vs. school classroom), and the location of delivery (e.g., school classroom vs. community church). Another form of adaptation is modification of the actual program content. This occurs when program content is altered to fit the needs of the participating population.

Despite a disagreement regarding the acceptable amount of adaptation of program curricula, limited research on this topic exists in the field of community-based parent education programs. However, an extensive amount of research is available regarding program fidelity with various community-based prevention and intervention programs. Several studies that focused on drug prevention programs have found higher fidelity of implementation is associated with better program outcomes (Battistich, Shaps, Watson, & Solomon, 1996; Botvin, Baker, Dusenbury, Tortu, & Botvin, 1990; Robert et al., 1998). Also, Burke, Oats, Ringle, Fichtner, and DelGaudio (2011) found that higher program fidelity to a classroom management program was related to better academic engagement and fewer suspensions for children involved in the intervention. In order to determine if community-based parent education programs are similar to the aforementioned prevention and intervention programs, more research in this field needs to occur. Future studies should focus on linking program fidelity to improved program outcomes in the field of parent education interventions.

To the contrary, some researchers believe that it is not strictly necessary to adhere to a program's curriculum (Castro et al., 2004; Sais et al., 2007). In fact, some degree of adaptation to address the specific needs of the each target population is essential. For instance, Saias et al. (2007) provided an example where program adaptation was necessary. The evaluation occurred through an analysis of data from 1,058 case notes of 105 families who took part in a maternal and child health home visiting program. They found it was nearly impossible to adhere

completely to the intended program curriculum. Instead, two discrepancies typically occurred in the program: Facilitators either omitted certain topics or added topics based on the family's needs. Saias et al. suggested that home visitation programs should discuss the curriculum with the family in order to determine what aspects would be most beneficial for that particular family. With this information, it seems as if home visitation programs should allow for enough flexibility in the curriculum to accommodate each family's individual needs (Saias et al., 2007).

In an in-depth study of program adaptation, Castro et al. (2004) described different program alterations that could benefit families receiving intervention services. The first adaptation incorporates changes based upon a family's cultural needs. These changes extend beyond surface structure changes, such as changing the ethnicity of role models, to changes more deeply rooted in the program's curriculum. Cultural adaptations would require program delivery staff to be culturally competent. Staff would then be aware of strong core values and norms and how to provide services while respecting those values and norms. An additional adaptation to address specific family needs is to change program delivery. Depending on different families' needs, a program may be adapted to different channels and locations of delivery. If a family is unable to attend a program, perhaps meeting online could be one delivery adaptation to consider. Also, the location of the program could be determined based on family needs. Hosting a program in a community-based center rather than a classroom also could serve specific families better (Castro et al., 2004).

Although these results indicate adaptations could potentially improve program outcomes, most interventionists still recommend adherence to program curriculum as much as possible. At this point, there is not enough evidence-based research concerning program implementation and adaptation in the field of prevention science (Kalafat, 2007). It is possible, with future research,

to discover in what type of settings more adaptations might improve program outcomes and in what type of settings high adherence to protocol will result in improved program outcomes. Another option to resolve this debate would be to have program developers specify what aspects of a program could be negotiated based on the participants' needs without taking away from the underlying theory behind the program (Elliot & Mihalic, 2004). Currently, however, with the lack of agreement on how to resolve this debate, it appears as if the majority of researchers suggest less fidelity results in diminished program effects.

Dosage

Dosage refers to the amount of a program that has been delivered to the participants (Dane & Schneider, 1998). Dosage is typically measured by the number of sessions a program intends to administer as well as how many sessions a participant actually attends. According to Korfmacher et al. (2008), dosage can be determined in many ways. First, a program must decide how frequently to meet. For example, a program can meet either once a week or once a month. Then, the length of contact must be determined. A program can meet for one hour per meeting or for three hours per meeting. Finally, the duration of the program must be determined. How long will the program itself last? Will it go for 10 weeks or for 20 weeks? Additionally, Nation et al. (2003) described dosage not only as the quantity, but the quality of contact hours. This refers to the way in which time is spent during the intervention.

Determining the correct amount of dosage. Some researchers believe that it is not always the case that higher dosage equates to better program outcomes. Cross, Gottfredson, Wilson, Rorie, and Connel (2010) evaluated an after-school program in terms of the quality of program implementation. They found that participant satisfaction with a program was not strongly related to the number of times the participant attended the program. Therefore, it should

not be assumed that a program with a higher amount of dosage is better than a program with a lower amount of dosage. In fact, having high attendance at a poorly run program could result in more negative outcomes than positive outcomes. These findings suggest that dosage alone should not be a measure of program quality (Cross et al., 2010).

However, some researchers claim that dosage is an important factor in program quality and program outcomes. McGuigan, Katzev, and Pratt (2003) found that the most important predictor of families remaining in an Oregon Healthy Families program was the number of hours per month of supervision the families received. This shows the actual amount of contact was more important than any other measurable characteristic of a program in determining retention. Additionally, in one study of a parent education program for caregivers who had allegations of maltreatment, Maher, Marcynyszyn, Corwin, and Hodnett (2011) showed that higher dosage was associated with more positive program outcomes that persisted well after participation in the program ended. Researchers found that after 6 months of program participation, caregivers who had attended more sessions were significantly less likely to be reported for child maltreatment. In fact, Maher et al. found that chances of a reported child maltreatment incident occurring within 6 months after participating in the program decreased by 3.8% with each additional educational session attended. Moreover, caregivers who had attended more sessions were also less likely to have been reported for child maltreatment two years after participation. Further research has shown that higher levels of program dosage have been associated with better outcomes for the children of participating parents. Wagner and Clayton (1999) evaluated the Parents as Teachers program and found a positive association between more home visits and greater parent impacts as well as greater child development. However, results were only statistically significant for the greater child development impacts.

With that said, a program would have no effect if dosage levels were at zero, or if a program never met. Some researchers suggest there could be a threshold effect where above a certain amount of dosage, for certain problems, additional dosage may not matter. The threshold level might be higher for more serious or chronic issues such as child abuse, and lower for more minor issues. Bagnato, Suen, and Fevola (2011) attempted to determine a minimum dosage threshold level, specifically for early childhood intervention programs. An early childhood intervention is described as any service program for children birth to 8 years. Children served by such programs typically are at risk for developmental delays due to circumstances such as the effects of poverty or developmental disabilities. Bagnato et al. were able to determine minimum functional dosage using data from 1,350 children in the Heinz Pennsylvania Early Childhood Initiative through a two-step calculation process. Their calculations looked specifically at the minimum expected progress and how many days it took to reach that progress.

Additionally, Wasik, Ramey, Bryant, and Sparling (1990) demonstrated a threshold effect in their study looking at an early intervention program for children at risk of developmental delays. Researchers compared three groups of participants. The first group attended an educational daycare program and received educational home visits. The second group had only the educational home visits, while the third group received no intervention. Researchers found the most improvements in children's' cognitive performance in the first group. However, the additional home visit did not affect parent or child behavior. These findings show the additional dosage (the home visits) had little to no effect on participant outcomes, suggesting a threshold limit exists. This information helps both service providers and government regulators in developing program dosage requirements and regulating that the specific dosage requirements are being fulfilled. It can also help private funders fund programs with adequate dosage

requirements. With that being said, it is important to note these findings are specific to early intervention initiatives and may not be applicable to a wide array of other programs. However, it does show that a threshold level is able to be determined and with further research could be determined for community based parent education programs.

Issues related to dosage. A problem frequently considered in implementation research is how to keep participants attending a program after their enrollment. The most common reason for participant dropout in parent education programs is related to a lack of time or scheduling conflicts (Gross et al., 2001). Nix, Bierman, and McMahon (2009) evaluated a parent education program for parents of children with conduct problems and found that single parents were more likely to have lower levels of attendance than co-parents. Low maternal age and factors that contribute to stress, such as parental depression and low social support, were associated with lower levels of attendance as well.

Currently, more information in the field of implementation evaluation is required about the frequency, duration, and intensity of a program to determine what produces the best program outcomes (Azzi-Lessing, 2011). With home visitation programs, Raikes et al. (2004) found that providing the full number of planned visits to at-risk families resulted in the best program outcomes. However, actually being able to do so proved difficult due to family scheduling and dropout rates, as previously discussed. Therefore, future research should focus on ways to determine ideal dosage levels in order to achieve desired program outcomes, while keeping in mind attrition and family schedules.

Participant Engagement

What is participant engagement? Participant engagement has been described as participants' feelings towards the services they are receiving from a particular program

(Korfmacher et al., 2008). A person's quality of interaction with program staff as well as understanding the program's curriculum will help foster participant engagement.

There are two types of participant engagement. First, there is positive engagement, which can be defined in terms of commitment to the program, a willingness to work with facilitators, showing receptivity for the program, and being satisfied with the services received (Heinicke et al., 2000; Korfmacher et al., 2008). On the contrary, negative engagement occurs when there is dissatisfaction with the program, a lack of commitment, and/or conflict with the program facilitators (Korfmacher et al., 2008). How well a person engages with a program will help determine program outcomes (Korfmacher et al., 2008).

Contributors to participant engagement. For participants to receive optimal benefits from a program, positive engagement must occur. Characteristics of not only the participants but the program itself both affect participant engagement. First, participant attributes can affect engagement. These attributes can include demographic variables such as age, employment, marital status, and income. For example, Roggman, Boyce, Cook, and Cook (2002) found that parents who had to worked long hours were less likely to attend an Early Head Start program for low-income families with infants and toddlers. Participant factors can also include levels of depression, number of close relationships, and motivation to change. In the same study, Roggman et al. found that fathers who were less depressed and less anxious were more involved in the program. Additionally, parents who are highly motivated to improve their parenting abilities are more willing to be involved in programs (Korfmacher et al., 2008). Lastly, Daro, McCurdy, Falconnier, and Stajonovic (2003) were able to determine what characteristics of participants and programs affect engagement. Their retrospective data analysis from 816 families participating in one of 17 Healthy Families of America program sites around the country showed

12

certain participant characteristics lead to longer service duration and a higher number of actual home visits. Their findings showed that older, unemployed mothers had the highest retention rate. This suggests that mothers who were unsure of their future economic security had the highest level of motivation to remain in this particular program. Researchers also found that African Americans and Hispanics had the highest retention rates as well as the greatest number of home visits. However, researchers were unsure of the exact reason for this finding.

Nix et al. expanded on contributors to participant engagement by demonstrating how parents must be willing to engage in program meetings, not just attend. Data from 455 parents who participated in a parental management training intervention showed that active participation contributed to greater gains. Specifically, if parents were willing to participate in discussions, pay attention, stay on topic, and enact role plays, improvements in perceptions of their children, parental warmth, and nonharsh discipline were seen at the end of the program (Nix et al., 2009).

Expanding beyond just individual level factors, community factors can also influence participant engagement. McGuigan et al. found that mothers living in counties with poorer community health were significantly less likely to actively participate in home visitation programs. Their study included 4,057 mothers with first-born infants enrolled in a Healthy Start home visitation program over a 3-year period. Their findings suggested that poor health communities - where low birth weight babies, high infant mortality, high suicide rates, and high alcohol related deaths are more common - have a lower level of engagement for services like the Healthy Start home visitation program. Engagement was defined as being involved in the program for more than 3 months. This could be due to the fact that mothers might be less willing to engage in such services because they do not see these types of programs as effective. On the contrary, healthy communities may see such programs as more beneficial. McGuigan et al. also

discovered that first-time mothers who were either isolated or had low levels of social support were less likely to actively engage in home visitation programs. However, further research needs to be done to determine why these types of mothers have a lower participation level in programs designed to give newborns a healthy start (McGuigan et al., 2003).

Program characteristics also contribute to levels of participant engagement. One aspect of a program that may contribute to participant engagement is the service provider's experience in the field and professional qualifications. Kormacher, Kitzman, and Olds (1999) looked at a nurse home visit program for first-time mothers and infants. They found that families visited by nurses, rather than paraprofessionals, had more contact and a smaller dropout rate. Daro et al. (2003) found that families participating in Healthy Families of America had more contact and a higher retention rate with facilitators who had higher levels of experience in delivering the services, regardless of professional background. In addition, Dara et al. found other service provider characteristics important in participant retention. Their data showed young, African American service providers also had the greatest success in retaining participants. The age factor can be attributed to the fact that younger workers were able to provide a significantly higher number of home visits. Thus, dosage may also vary with facilitator attributes.

Beyond service provider characteristics, other aspects of programs are important in maintaining participant engagement. The number of participants in a group can affect participant engagement. With lower caseloads, programs are able to provide more contact with families, which may result in higher levels of engagement (Daro et al., 2003). Daro et al. also discovered engagement and retention were influenced by whether or not the service providers matched the participants in terms of parenting status as well as race/ethnicity.

One final aspect of a program that has the ability to affect participant engagement is its cultural sensitivity. Research has shown that sensitivity and respect for different cultures and family traditions can increase participant engagement (Kumpfer, Alvarado, Smith, & Bellamy, 2002). Looking specifically at participants who were nonnative speakers, Eisner and Meidert (2011) found that the language barrier was the most significant contributing factor to low engagement in a parent education program. Nonnative speakers were just as likely to enroll in the program as native speakers, but significantly less likely to actually attend classes. These findings suggest that programs need additional cultural adaptations beyond just translation of the curriculum. Such adaptations include adjusting recruitment styles, changing the delivery of the program, and adjusting actual program content as well (Kumpfer et al., 2002). However, a significant amount of research is still necessary in this area to determine how exactly to adapt programs to be more culturally sensitive without compromising fidelity (Kumpfer et al., 2002).

For this study, I evaluated the implementation of a parent education program in Larimer and Weld Counties of Colorado. The program originated as Best Start for Babies (BSB) and expanded its age range to include the Best Start for Toddlers (BST) class as a continuation of BSB. BSB is currently in its third year, while BST is in its second year. BSB/T strives to promote positive child outcomes by providing parents with support throughout the prenatal to toddler years. BSB/T provides parents information about normative infant/toddler development as well as strategies for parent/child attachment and access to community resources.

Research Questions

Previous research on program implementation demonstrates the importance of fidelity, dosage, and participant engagement to successful program outcomes. Although the norm for evaluation has been to focus on actual program outcomes, there has been a relative lack of

attention paid to determine how programs essentially achieve those outcomes. By focusing on a program that has not yet been systematically evaluated, I was able to determine how dosage, degree of fidelity, and participant engagement contribute to BSB/T program outcomes. This information can provide directions on how to facilitate programs for staff and create a more effective program for participating individuals and families. Shonkoff and Phillips (2000) explained that much of human brain development occurs prenatally through the first few years of life. Therefore, it is crucial to inform parents how to best encourage healthy development for their children. Through parent education programs, like BSB/T, parents are able to learn how to properly care for and encourage healthy development in their children's first few years of life.

Through this evaluation I hope to answer the following questions:

1. Is greater fidelity to the BSB/T curriculum related to more benefits from the program?

2. What sorts of adaptations where made and did the adaptations (lower levels of fidelity) contribute to diminished program effects?

3. Does a greater dosage contribute to greater program effects?

4. Does more positive participant engagement contribute to greater program effects?

Method

Development of the BSB Curriculum

The program developer and two other facilitators attended a 4-day training on the Partners for a Healthy Baby curriculum put on by Florida State University, which the BSB/T curriculum is based on. In order to promote fidelity to the curriculum, when new facilitators are hired, the program developer conducts a half-day meeting to review the curriculum. Additionally, the program developer covers each week's materials with the new facilitators to ensure they are prepared for the classes.

Sample

A total of 444 parents from Larimer County (Fort Collins) and Weld County (Fort Lupton and Greeley) participated in the BSB/T program. BSB has had 10 cohorts total whereas BST has enrolled 3 cohorts.

Participants for BSB/T were recruited through flyers distributed at more than 50 community organizations. Locations included libraries, laundry facilities, and mobile home parks. Organizations from around each community also referred pregnant mothers and their partners, or parents with a child under 3 years of age, to the BSB/T program. Examples of these organizations include the Healthy Beginnings Program, which is a prenatal program for future parents in Northern Colorado; the Poudre Valley School District; and the Salud Clinic, which provides primary health care needs to families in Northern Colorado. Participants were primarily mothers (70%) who were high school educated ($M = 12.41$ years) and low socioeconomic status ($M = 20.62$ on the Duncan SEI, where 20 = manual laborer). The remaining 30% of participants were fathers (27.5%), grandmothers (.3%) and stepparents (.5%). Participants in Weld County were significantly more likely to be low SES, younger, cohabitating, and Hispanic. Demographic

information is summarized in Table 1. One unexpected fact is that the age of the target child for BSB in Weld County is higher than the age of the target child for BST in Weld County. This could be explained by the fact that participants in the BSB program who were expecting their first child had a missing value for the child's age and so it was not factored in when calculating the mean. Also, some participants who may have already had a child, but were expecting a second child would have had the first-born child's age entered even though they enrolled because they were expecting their second child.

A composite risk index was calculated to capture the collective disadvantage that some participants may be facing. The risk index is comprised of the following factors: Parent education level, income, marital status, employment status, and whether the mother is currently (coded as 2) or was (coded as 1) a teen mother. The highest score of 2 was given for the lowest category of education and income as well as being single. Participants in BSB/T were similarly represented across risk levels; however, those with the highest risk - primarily adolescent mothers - were fewest in number. An additional indicator of risk is whether the pregnancy was planned; 54% said they did not mean to get pregnant, but 22.2% felt a little or very uncertain about "having a child at this time in your life."

Table 1

Mean (SD) Participant Demographics, by Program and Locale

	Larimer	Weld
	Best Start for Babies	
Education (years)	12.75 (2.05)	12.06 (1.97)
Occupation (Duncan SEI)	23.33 (22.38)	18.23 (19.29)
Age	24.57 (6.26)	22.62 (7.03)
Marital Status- Single	39%	41%
-Married	47%	43%
Ethnicity- White	66%	38%
- Hispanic	30%	59%
Age of target child	1.68 (2.16)	1.64 (2.81)
	Best Start for Toddlers	
Education (years)	13.41 (2.21)	11.30 (1.85)
Occupation (Duncan SEI)	33.52 (20.30)	13.16 (11.27)
Age	27.25 (5.11)	24.47 (6.89)
Marital Status- Single	3.1%	36.8%
- Married	59.4%	57.9%
Ethnicity- White	56.3%	34.2%
- Hispanic	37.5%	65.8%
Age of target child	1.83 (.49)	1.31 (1.13)

Measures of Implementation

Each BSB/T class had ten sessions. After each BSB/T session, facilitators completed a logbook. These logbooks had Likert-type ratings as well open-ended questions about participant engagement and program implementation. Descriptions of specific aspects of the implementation evaluation are below.

Fidelity to protocol. Facilitators used a Likert-type scale to rate fidelity to the curriculum for that week. Each activity that took place during a session was listed in the facilitator logbook. Facilitators then rated the delivery of each activity on a scale from 1 (*needs some work)* to 3 (*very successful).* Thirty BSB/T sessions were selected at random to examine the concordance in fidelity ratings between facilitators. Cohen's kappa statistic was used as measure of agreement: $\kappa = .76, p < .001.$ Ideal implementation of activities would be scored as a 3. The program

developer had clustered all BSB/T activities into the specific domains listed in Table 2. In total, there were 12 domains for BSB activities and 11 domains for BST. Across all 10 sessions, the average rating for fidelity to protocol within each activity domain was computed.

Table 2
Activity Domains for Best Start for Babies and Best Start for Toddlers

BSB	BST
1. Child and Brain Development	1. **Child and Brain Development**
Fetal Development	How Children Learn Language
Bright Beginnings	Bright Beginnings
Learning Games	Learning Games
Monthly Development	Social-Emotional Development in Children
Right to Read	Union Colony Music Presentation
ERCC Literacy Program	Children and Nature Connect
Developmental Screenings	Story Time Rhymes
Songs and Games	Songs and Games
2. Discipline	**2. Discipline**
Consequences and Discipline	The Joy of Misbehaving
	Toddler Tough Spots
3. Healthy Prenatal Care	**2. Discipline (continued)**
Prenatal Nutrition	How to Beat Your Kids (At Their Own Game)
Health Pregnancy Quiz	Erasing the Battle Lines
Prenatal Discomforts	Effective Discipline for Toddlers
	3. Healthy Prenatal Care – not covered in BST
4. Health and Safety	**4. Health and Safety**
Sleep	Toddler Tough Spots
Infant CPR and Choking	Child CPR
What to do When Your Child Gets Sick	Health and Nutrition for Toddlers
Nutrition	Family Health
Food and Fitness Matter	Stop Smoking
Safe Kids Presentation	Toddler Feeding Tips
Healthy Oral Care	**5. Social/Emotional Needs**
Dental Screenings	Happiest Toddler on the Block
Car Seat Safety	SocioEmotional Development
Home Safety Quiz	Parent Focus Groups
5. Social/ Emotional Needs	**6. Parent Self-Efficacy**
Recognizing Baby's Clues	Goal Setting
Happiest Baby on the Block	

Table 2 Continued

6. Parent Self- Efficacy	**7. Stress Management**
Memory Pages	Budgeting
7. Stress Management	Emotional Well-Being
How to Balance Work and Parenting	Parenthood
Finance	Back to School Information
Budgeting	Financial Fitness
Brain Development and Stress	**8. Social Support**
Brain Myth Puzzles	Diaper Rules
Stress and Temperaments	Toddler Olympics
8. Social Support	Making Food Toys
Friend and Resource Matrix	**9. Parent and Child Attachment**
Partner Support	Building Positive Relationships
Pregnancy Support Program	Separation/Stranger Anxiety
9. Parent and Child Attachment	Father, Parenthood, Relationships
Dunstan Baby Language	**10. Identifying and Accessing Quality Childcare**
Temperament Treasure Hunt	Quality Childcare Video
"The First Years Last Forever" video	**11. Identifying and Accessing Community Resources**
Male Role Models	Center for Adult Learning
Fathering	Larimer Workforce
10. Identifying and Accessing Quality Childcare	**12. Program Processes**
Quality Childcare Video	Introductions and Expectations
11. Identifying and Accessing Community Resources	Prize Drawings
Resource Scavenger Hunt	Certificates
Head Start Speaker	Rules
12. Program Processes	
Introductions and Expectations	
Prize Drawings	
Certificates	

Additionally, open-ended comments by the facilitators were content coded (described later). The open-ended comments answered questions about how the session went, if there were unusual circumstances in the session, or if there was any participant feedback. Among the content codes, there were four themes that represented fidelity. These included fidelity, limited time, improvements, and culture; see Table 3 for definitions and examples of each theme. All of these themes addressed adaptations to the curriculum that had occurred, or suggestions for future adaptations that could improve the program. Across all 10 sessions, the number of times each theme was mentioned was totaled.

Dosage. For each session, sign-in sheets were kept to monitor attendance of participants. These sign-in sheets were also used to compute the number of hours of contact. If participants missed a session, a home visit was provided as a substitute for the missed session.

Facilitator logbooks also reported how time was spent in each session. Facilitators indicated the amount of time spent on meals and socializing, individual/group activities, and facilitator lectures.

Participant engagement. Facilitators rated participants' engagement in each session, using a scale from 1 (*passive*) to 5 (*active*), where "passive" indicated that participants were inattentive or disengaged and active indicated they were attentive and engaged. Facilitators also rated how much the parents accepted new ideas, on a scale of 1 (*rejected*) to 5 (*embraced*), and how much the parents got along with each other, from 1 (*conflict*) to 5 (*support*). Finally, facilitators rated how much the parents respected others' perspectives on a scale from 1 (*put downs*) to 5 (*valued*).

The logbooks also included two items related to how well facilitators thought they had prepared and delivered the program curriculum. Facilitators rated (1) how well-prepared they were and (2) how clear and organized their presentation was on a scale of 1 (*strongly disagree*) to 5 (*strongly agree*).

Furthermore, open-ended comments by the facilitators were content coded for participant engagement. Among the content codes, there were four that represented participant engagement. These included participant engagement, resistance, program delivery, and disruptions (see Table 3). These themes addressed situations where participants were engaged or not, and any issues with the program that may have affected participants' ability to engage. Across all 10 sessions, the sum of each theme was totaled and a total percentage was computed.

22

Measures of Program Impact

       Knowledge of infant development. The Knowledge of Infant Development Inventory (KIDI; MacPhee, 1981) measures a caregiver's understanding of developmental milestones, strategies for child rearing, and knowledge of basic child development. The original KIDI is a 58-item scale. In order to shorten the scale for the BSB/T evaluation, items that were not as germane to the program objectives and curriculum were dropped to create a 48-item KIDI. The first 29 items ask parents to agree or disagree with descriptions of child behaviors that are normative based on general child development, as well as factors that could potentially influence both growth and behavior of children (MacPhee, 1981). On the 17 milestone items, which relate to specific norms of motor, social, and cognitive development, respondents selected agree, "younger," or "older." The final two questions were multiple choice items related to management of infant or toddler behavior. A total percent correct score is derived. For a normative sample, alpha reliability was .91, while Cronbach's alpha for the BSB/T sample was .82. Its validity has been established in terms of its relation to formal versus informal experiences with infants and children, its sensitivity to differences in professional training and to intervention, its convergence with other measures of knowledge of child development, and its correlation with various measures of child rearing (e.g., Huang, Caughy, Genevro, & Miller, 2005) and the home environment (MacPhee, 1981).

       Parent self-appraisals. The Self-Perceptions of the Parental Role (SPPR) is a 22-item scale assessing various components of parental self-appraisals (MacPhee, Benson, & Bullock, 1986). Components measured included confidence in one's skills as a parent (i.e., parent self-efficacy), satisfaction with the parental role, and investment with the parental role. Each item includes two opposing statements, such as "Some parents often worry about how they're doing

23

as a parent BUT Other parents feel confident about their parenting abilities." Parents respond by indicating the statement that best represents their feelings, selecting either *sort of true for me* or *really true for me*. The SPPR has been found to have high internal reliability (α = .78-.87) as well as high test-retest reliability (r = .80-.88). The SPPR also has been found to have high convergent and factorial validity, and construct validity in terms of relations to difficult child behavior, punitive child-rearing practices, social support, and sensitivity to intervention (Miller-Heyl, MacPhee, & Fritz, 1998) with low-income and at-risk parents (MacPhee et al., 1986; MacPhee, Fritz, & Miller-Heyl, 1996).

Coding of Facilitator Logbooks

Facilitators' responses to open-ended questions in the logbooks were content coded following procedures described by Weber (1990). After each session, facilitators completed logbooks that rated and described (a) how much time was devoted to meals and socializing, individual/group activities, and facilitator lectures; (b) activities during the session and a rating of how well each activity was delivered; (c) ratings of various forms of participant engagement, (d) ratings of facilitator preparation and delivery of program; and (e) space for open-ended comments about engagement, disruptions, and how the session went overall. The initial coding scheme was used to evaluate implementation of the DARE to be You Program (DTBY; Miller-Heyl, MacPhee, Walker, & Podunovich, 2013). The DTBY coding scheme, in turn, was based on one developed by Hill, Maucione, and Hood (2007) to assess fidelity of implementation of a statewide parenting program. This coding scheme was then used by the author and her adviser to independently code facilitator logbooks for one BSB cohort of 10 sessions. As a result of discussing discrepancies in coding, three themes were added to the original set of codes. These coding themes are listed in Table 3. The same two coders then used the expanded coding scheme

with two more cohorts of BSB. Coding discrepancies were resolved by consensus. Interrater

reliability was determined using 15 randomly selected sessions from all 13 cohorts. Cohen's

kappa statistic was used as measure of agreement: $\kappa = .84$, $p < .001$.

Table 3
Facilitator Logbook Content Coding Themes

	Definition	Example
Fidelity		
Fidelity	Adaptations to the curriculum	Presenting on part of the nutrition activity because parents were familiar with the information
Limited Time	When facilitators run out of time	Only completed ½ an activity
Improvements	Suggestions for how to improve future sessions	Pass out a list of resources instead of in the power point
Culture	Cultural modifications to the curriculum	When a translator becomes confused and cannot translate all of the material
Program Delivery	Facilitator preparation, process evaluation	A parent misses the first week because he/she was not aware of the start date
Engagement		
Engagement	Participants show engagement	Parents showed interest during an activity
Resistance	Participants do not show engagement	Parents seemed bored during an activity
Practice	Facilitators report parents used learned skill outside of the session	Parents tried the 5 S's and it helped soothe their crying baby
Disruption	An external event interrupts a session	A parent has to take his/her baby to the hospital mid-class
(Group) Dynamics	Any positive or negative event that causes conflict or enhances group cohesion	Parent work together to accomplish a task during an activity
Benefits	Benefits parents receive through participation	When a parent says he/she is learning how to manage stress

Description of Typical BSB/T Classes

The first set of analyses focused on describing how the typical BSB/T session was implemented. To this end, descriptive analyses were first conducted for each activity domain. The percentage of the total curriculum devoted to each activity domain (described in Table 2) was computed along with mean ratings for how each domain was delivered by facilitators. These ratings were used to measure fidelity. In addition, mean time devoted in each session to meals and socializing, to individual/group activities, and to facilitator lectures were calculated. These data were used as a measure of dosage. A profile of the typical BSB class, across all 10 sessions, is shown in Figures 1 and 2. Graphed is the percentage of class time devoted to each curricular domain as well as how time in session is apportioned to social time, individual and group activities, and facilitator presentations. A profile of the typical BST class, across all ten sessions, is shown in Appendix A.

Figure 1. Percentages for activity domains across all BSB cohorts. Shown are the minimum, mean, and maximum percentages for activity domains across all 10 BSB cohorts. Percentages for activity domains across all BST cohorts are listed in Appendix A.

Figure 2. Time in minutes (dosage) devoted to meals, activities, and presentations across all BSB cohorts. Shown are the minimum, mean, and maximum percentages, by type of curricular activity. Time spent on meals, activities, and presentations across all BST cohorts is listed in Appendix A.

In order to determine whether the activity profile changed as the program matured, the percentage of the curriculum devoted to each domain as well as the activity success means is in tables by cohort; the cohort-specific profiles are listed in Appendix A. Additionally, the minimum, mean, and maximum percentages of the curriculum devoted to each domain across cohorts and time spent on socializing, activities, and presentations is also displayed in Appendix A. Looking at trends of implementation across the BSB cohorts, the primary emphasis of the program was focused on activities having to do with child and brain development, and health and safety. Percentages per cohort of how much of the total class was devoted to these topics ranged from 10% to 48% for child and brain development, and 17% to 42% for health and safety topics. Looking further at trends in implementation across the BSB cohorts, there were also topics that had very little emphasis. These included activities related to parental self-efficacy and identifying and accessing quality childcare. Percentages per cohort of how much of the total class was devoted to these topics were as little as 0% to 6% for parental self-efficacy and 0% to 4% for identifying and accessing quality childcare. Although there was little variation from

27

cohort to cohort in the amount of time devoted to some of the topics, there was quite a bit of variability in how much attention other topics received. For example, child and brain development had a large amount of emphasis across cohorts, but it ranged from 17% to 42%, which is a pretty large variation. This indicates there may not yet be a standardized approach to delivering the program. Also, if the program blueprint included parent self-efficacy and quality child care, but there was very little time devoted to these topics this may suggest some slippage between the blueprint and actual implementation.

Turning to the three BST cohorts, the primary emphasis of the program was focused on activities having to do with health and safety as well as program processes. Percentages of how much of the total class was devoted to these topics, per cohort, ranged from 15% to 21% for health and safety and 11% to 25% for program processes. There was one activity domain that had no emphasis at all, which was identifying and accessing quality childcare: 0% of the curriculum was devoted to this topic. Other activities that had little emphasis in implementation included parent and child attachment activities (0%-8%) and identifying and accessing community resources activities (0%-11%).

The next set of descriptive analyses focused on how facilitators described the program's implementation as well as participant engagement. These analyses were based on content coding of facilitators' logbook comments, as described in Table 3. For each cohort, percentages of how many times the themes related to program implementation appeared in facilitator logbooks were computed. Frequencies of themes by cohort were calculated with the expectation that themes related to fidelity would decline as the program matured. Percentages of the frequency of content codes can be found in Table 4. Common comments related to fidelity included facilitators noting if they were not able to present on a topic because they ran out of time, or providing a way in

28

which an activity can be improved for future classes. Common comments related to engagement included facilitators noting parents being very engaged by asking presenters questions, or performing tasks during group/individual activities such as practicing CPR on mannequins. Comments related to parents being disengaged (or resistant) included facilitators noting if parents were having their own side conversations, or if parents were using their cellphones during a presentation.

In the first BSB class, 32% of content codes were related to fidelity (i.e., fidelity, limited time, improvements, and culture). By the third cohort, themes related to fidelity began to decrease. However, for cohorts 7, 8, and 9, there was an increase in themes related to fidelity: The percentages resemble the percentages from the first few cohorts. The program developer thought that this "rebound" might be due to the co-facilitators not attending a four-day training. Another explanation could be if these cohorts had a lot of off-task behavior, which might have prompted the facilitators to include more comments about fidelity in their logbooks. By the last BSB cohort, 19% of content codes were related to fidelity. For BST, the first cohort had 31% of content codes related to fidelity, whereas the last cohort had 13% of content codes related to fidelity. This indicates there was a downward trend in themes related to fidelity as the program matured. I did not use inferential statistics to test this hypothesis, though, because the small number of cohorts provided inadequate statistical power.

Table 4
Percentage Total of Themes in Codebooks (BSB)

| | Fidelity Themes | | | | | | Engagement Themes | | | |
Cohort	Fidelity	Limited Time	Improve-ments	Culture	Program Delivery	Practice	Disrupt-ions	Resist-ance	Engage-ment	Group Dynamic	Benefits
1	.08	.01	.18	.05	.14	.00	.01	.10	.27	.13	.04
2	.03	.05	.14	.02	.08	.03	.00	.22	.31	.07	.03

Table 4 Continued

3	.00	.00	.16	.03	.08	.03	.02	.16	.37	.06	.06
4	.03	.08	.02	.02	.09	.02	.03	.23	.35	.12	.03
5	.00	.00	.04	.00	.16	.02	.00	.04	.50	.20	.05
6	.00	.00	.00	.05	.10	.00	.05	.15	.45	.05	.15
7	.09	.00	.11	.04	.16	.00	.00	.13	.33	.07	.07
8	.09	.07	.09	.02	.07	.00	.05	.13	.41	.05	.02
9	.11	.02	.13	.04	.05	.02	.02	.23	.36	.04	.00
10	.06	.05	.08	.00	.08	.03	.00	.09	.42	.17	.03

Table 5
Percentage Total of Themes in Codebooks (BST)

	Fidelity Themes						Engagement Themes				
Cohort	Fidelity	Limited Time	Improvements	Culture	Program Delivery	Practice	Disruption	Resistance	Engagement	Group Dynamic	Benefits
11	.07	.03	.05	.16	.08	.00	.00	.18	.39	.02	.02
12	.03	.03	.16	.06	.01	.01	.01	.33	.29	.04	.01
13	.08	.00	.05	.00	.13	.08	.08	.05	.33	.15	.03

Variations in Implementation in Relation to Program Outcomes

The final set of analyses focused on associations between the implementation variables and changes on the outcome measures. Outcome variables were in the form of gain scores, between pretest and posttest, on the KIDI and SPPR. An average gain score was computed for each cohort. These gain scores were correlated with implementation variables that were calculated for each cohort: average number of times fidelity themes were mentioned in the facilitator logbooks, mean activity success ratings, time devoted to each activity domain, attendance, engagement ratings, and average times engagement themes were mentioned in the facilitator logbooks. It is important to note that with a final sample size of 13 cohorts, even large correlations may not be statistically significant: Statistical power for $r = .30$ is .28 and for $r = .50$, power is .60. Weak statistical power combined with multiple correlations generated by the analyses means that these analyses was exploratory; that is, there was limited ability to make inferences about how variations in program implementation were associated with variations in

program outcomes. Given that this was an exploratory study, a more liberal criterion of $r = .48$, which represents a large effect size (Cohen, 1988), will be noted in these analyses.

The first set of correlational analyses was related to the research questions related to program fidelity. Two measures of program fidelity were included in the analyses. First, gain scores were correlated with the average number of times fidelity themes were mentioned in the facilitator logbooks. Although there were no statistically significant findings from these analyses, seven of the correlations do represent large effect sizes (see Table 6). For the BSB cohorts, there was a strong negative association between the culture theme and gain scores on the KIDI. This shows that the more often cultural modifications to the curriculum were mentioned in facilitator logbooks, the lower the gain scores on the KIDI. For BST cohorts, there were strong positive correlations between fidelity and gain scores on the SPRR, culture and gain scores on the SPRR, and limited time and gain scores on the KIDI. This shows that, for the BST cohorts, more frequent mentions of fidelity and limited time were related to better program outcomes, which contradicts what was expected.

Table 6

Correlation between Fidelity Mentions in Facilitator Logbooks and Gain Scores

| | BSB | | BST | |
	KIDI	SPRR	KIDI	SPRR
Fidelity	.22	.17	-.10	.61
Limited Time	-.10	-.38	.87	.30
Culture	-.60+	-.31	.48	.94
Improvements	.23	-.28	-.49	-.49

*$p < .05$ +$p < .10$

Second, gain scores were correlated with "success" of the activity as a measure of fidelity. In the BSB cohorts, none of the fidelity ratings were reliably related to KIDI gain scores (see Table 7). However, improvements in parental self-efficacy were *inversely* related to "success" in implementing four domains of the curriculum: child and brain development,

31

prenatal care, accessing community resources, and program processes (see Table 7). Turning to

analyses with the BST cohorts, gain scores on the KIDI were related positively to facilitators'

ratings of successful implementation for five of the curriculum domains (see Table 7) – child and

brain development, discipline, social/emotional needs, stress management, and social support

activities – whereas the rating of program process was *inversely* related to KIDI gain scores as

well as to changes in parent self-efficacy. Greater improvements on the SPPR were positively

related to ratings of successful implementation for the same five curriculum domains as were

found for KIDI gain scores (see Table 7), with the exception that the domain of meeting

toddlers' social/emotional needs was related to SPPR gain scores but stress management was not.

Table 7
Correlation between Activity Success Rating and Gain Scores, by Program

Domain	BSB		BST	
	KIDI	SPRR	KIDI	SPRR
Development	-.26	-.49	.95	.48
Discipline	-.40	-.35	.95	.48
Healthy prenatal care	-.32	-.49		
Health/Safety	-.24	-.44	.95	.48
Social/Emotional needs	-.23	.38	-.20	.52
Parent efficacy	-.15	---	---	---
Stress management	-.09	.05	.74	.09
Social support	.10	.15	.98	.87
Attachment	-.12	.22	---	---
Child Care	-.09	-.26	---	---
Community resources	-.44	-.78*	---	---
Program processes	-.23	-.73+	-.75	-1.0*

*$p < .05$ +$p < .10$

The second set of correlational analyses concerned the amount of time devoted to each

activity domain and program outcomes, which were related to research question 3 regarding

dosage. In the BSB cohorts, improvements on the KIDI and the SPRR were related positively

with time spent on health and safety as well as stress management curriculum domains. On the

contrary, the amount of time devoted to the parent and child attachment domains was inversely

related to KIDI and SPRR improvements. The amount of time devoted to discipline was also

inversely related to KIDI gain scores and the amount of time devoted to program processes was

inversely related to SPRR gain scores. For BST, there were several strong and positive

correlations between time spent on activities and gain scores. Time spent on child and brain

development, health and safety, parental self-efficacy, identifying and accessing community

resources, and program process activities was positively associated with program outcomes. In

contrast, time spent on discipline, stress management, social support, parent and child

attachment, and program processes were all associated negatively with program outcomes (see

Table 8). It is important to note that dosage in the health and safety curriculum domain was

positively associated with outcome measures across both programs, whereas the opposite was

true for parent and child attachment.

Table 8
Correlation between Percent Time Devoted to Each Activity and Gain Scores, by Program

| Domain | BSB | | BST | |
	KIDI	SPRR	KIDI	SPRR
Child and Brain Devel.	-.11	-.42	.97	.56
Discipline	-.27	-.66+	-.63	.06
Healthy Prenatal Care	.06	-.32		
Health/Safety	.49	.76*	.99+	.63
Social/Emotional Needs	-.19	-.14	.55	-.15
Parent Efficacy	.34	-.20	.95	.48
Stress Management	.47	.65+	-.63	.06
Social Support	.24	.32	-.02	-.69
Attachment	-.55+	-.49	-.95	-.48
Child Care	.06	-.10	---	---
Community resources	-.38	-.31	.75	1.0*
Program processes	-.47	-.20	-.49	.74

*$p < .05$ +$p < .10$

The third set of correlational analyses also looked at the relation between dosage and gain

scores. In addition to the amount of time devoted to each activity as a measure of dosage, the

number of sessions participants attended for each cohort was an indicator of dosage. Except for

the BST cohort and gain scores on the SPRR, more frequent attendance was negatively related to

program outcomes (see Table 9). These results are somewhat misleading given that (1) participants with zero dosage also did not complete posttests and therefore gain scores could not be computed for them, and (2) 85% of participants attended between 7 and 10 sessions, meaning that the dosage variable had a restricted range. The final measure of dosage was how time was apportioned in each class. For these analyses, there was a statistically significant association between time spent socializing and gain scores on the KIDI for BSB classes. Additionally, there was a strong positive correlation between time spent on individual and group activities and program outcomes for the BST cohorts. The amount of time devoted to didactic facilitator presentations was weakly and generally negatively associated with program benefits.

Table 9
Correlation between Dosage and Gain Scores, by Program

	BSB		BST	
	KIDI	SPRR	KIDI	SPRR
Attendance	-.48	-.67	-.62	.98+
Socializing	.70*	.48	.31	-.42
Activities	.09	.17	.50	.95
Presentations	-.28	-.48	-.44	.28

Note: Socializing refers to the time the parents spent eating meals before class began. Activities refer to individual and group activities that facilitators planned to go along with the presentations.
*$p < .05$ +$p < .10$

Last, correlations were computed to determine whether ratings of participant engagement were related to gain scores, which refers to research question 4. Each session, facilitators rated participant engagement by looking at how well participants participated, accepted new ideas, got along with each other, and respected each other's perspectives. For the BST cohorts, results showed that participation, acceptance of new ideas, and respecting each other's perspectives were all positively related to program outcomes. Ratings of how well participants got along with each other were related to an increase in parental self-efficacy. However, there was a negative correlation between how well prepared the facilitators were and program outcomes (see Table 10).

34

Further analyses were performed to examine the association between gain scores and facilitator logbook counts of themes related to engagement (i.e., engagement, resistance, disruption, and group dynamics). There was found to be a statistically significant association between group dynamics and gain scores on the KIDI for BSB classes, indicating that when group dynamics were overt (either positive or negative), participant gains in knowledge of development were greater (see Table 11). On the contrary, more frequent mentions of group dynamics was negatively correlated with scores on the SPRR for BST cohorts. As expected, for the BST classes, the resistance theme was negatively correlated with program outcomes. The number of times disruptions was mentioned was too infrequent to analyze. Finally, engagement mentions were strongly, positively correlation with program outcomes for BST cohorts but not for the BSB cohorts.

Table 10
Correlations between Gain Scores and Ratings of Engagement and Facilitator Preparation, by Program

| | BSB | | BST | |
Process Rated	KIDI	SPRR	KIDI	SPRR
Participation	-.14	-.09	.97	.88
Acceptance of new ideas	.02	.42	1.0*	.72
Getting along	-.23	-.05	-.23	.50
Respecting perspectives	-.24	.10	.57	.98
Facilitator preparedness	.11	-.17	-.71	-.04
Facilitator organization	.39	.39	.04	-.64

*$p < .05$ ⁺$p < .10$

Table 11
Correlations between Engagement Mentions in Facilitator Logbooks and Gain Scores

| | BSB | | BST | |
	KIDI	SPRR	KIDI	SPRR
Engagement	-.11	-.17	.92	.94
Resistance	-.36	-.01	-.69	-.03
Group Dynamics	.68*	.49	.00	-.67

*$p < .05$ ⁺$p < .10$

Additional exploratory analyses were performed to determine which implementation variables distinguished BSB cohorts with no gain scores on the outcome measures ($n = 3$) from those who had positive gain scores on program measures ($n = 7$). Results of these analyses indicated that there were 10 variables that distinguished the two groups in terms of program implementation. A more liberal criterion of $p < .15$ was selected given the small sample sizes. The implementation variables that discriminated between the two groups include the percent of class time spent on healthy prenatal care, health and safety, and parent and child attachment. For the first two of these implementation variables, more time devoted to health and safety was associated with improvement on the outcome measures whereas time devoted to attachment was related to no gains on the outcome measures. Additionally, the "success" ratings for the health and safety domain, the stress management domain, the accessing quality childcare domain, and the program processes domain distinguished between the two groups: the cohorts with no benefits had higher success ratings. Finally, the group of cohorts that showed improvement, compared to cohorts who did not improve, more often had facilitators mention group dynamics in logbooks, had higher facilitator ratings of how well participants respected each other's perspectives, and had higher facilitator self-ratings of their organization in delivering the curriculum.

Potential moderating variables. The final set of exploratory analyses was conducted to look at potential moderating variables. The first analysis was to determine if high engagement combined with more adaptations to the curriculum resulted in greater program effects. I speculated that participants might benefit the most when facilitators were responsive to parents' requests for program modifications that best met their individual needs. To test for moderation, an interaction term was first computed such that the z score for the composite engagement

36

variable, based on facilitators' ratings over the course of each cohort, was multiplied by the z score for the average fidelity or "success" rating for that cohort. This interaction term was correlated with program outcomes. The results of this analysis (in Table 12) indicate that for BSB, engagement levels combined with more adaptations resulted in strong positive correlations with improvements in parental self-efficacy. However, for BST, high engagement combined with high fidelity, or "success," ratings resulted in strong negative correlations.

Further analyses explored whether participant risk level interacted with engagement and fidelity. The rationale for these analyses was that parents with more sociodemographic risks, perhaps reflecting higher stress levels, might derive more benefits from the program if they were highly engaged than would similarly engaged low-risk parents. First, the composite risk score was multiplied by the composite engagement z score to see how they interacted with program outcomes. For BSB, when risk interacted with engagement, there was a strong negative correlation with improvements in parental self-efficacy. For BST, the opposite was true: When risk interacted with engagement; there were strong positive correlations across program outcomes. Similarly, when risk level was multiplied by fidelity levels, they were strongly, negatively associated with BSB program outcome and positively associated with BST program outcomes. Both sets of findings indicate that the greatest BST program benefits accrued to parents who were high in risk status with high engagement and high facilitator ratings of the curriculum's success – whereas the greatest BSB program benefits accrued to parents who were low in risk status with high engagement and high facilitator ratings of the curriculum's success. However, it should be noted that the correlations involving the composite risk variable were near zero (see Table 12), and the correlations involving the interaction term were similar to the main effects of the engagement and fidelity variables alone (cf. Tables 7 and 10).

Table 12

Correlations between Moderating Variables and Gain Scores, by Program

| | BSB | | BST | |
	KIDI	SPRR	KIDI	SPRR
Eng by Fid	-.10	.70[+]	-.79	-.17
Composite risk score	-.05	-.03	-.12	.09
Risk by Eng	-.20	-.70	.96	.89
Risk by Fid	-.40	-.67[+]	.79	.17

Note. Eng = Engagement; Fid = Fidelity.
*$p < .05$ [+]$p < .10$

Discussion

One critical step in determining whether interventions have a positive impact on the lives of their participants is to evaluate a program's implementation. A great deal of previous research has only focused on evaluating a program's outcomes, ignoring the actual implementation process. However, without evaluating implementation, it is impossible to determine whether the program itself or extraneous factors led to program outcomes (Durlak & DuPre, 2008). Important dimensions of implementation that have the ability to affect program outcomes include fidelity to the curriculum, dosage, and participant engagement (Summerfelt, 2003). Evaluating how these three factors affect program outcomes can provide valuable information to those involved with program development.

By evaluating the first three years of the implementation of the Best Start for Babies/Toddlers program, valuable information was gained as to how the program is typically delivered. For instance, the BSB/T program allocates similar amounts of time for meals/socializing and individual or group activities in each class, with a greater amount of time allotted for educational presentations. Within the educational presentations, BSB/T places more emphasis on topics related to child and brain development, health and safety, and parental social support than on other topics such as accessing quality childcare and parent-child attachment.

Evaluating the implementation of BSB/T also provided useful information on how various facets of implementation are related to program outcomes. One question examined in this evaluation study was whether a higher level of fidelity to the BSB/T curriculum produced greater program outcomes. Assessing fidelity is a critical step in implementation evaluation. If a program is not delivered as intended, it is difficult to determine whether program outcomes were a result of the intervention (Kalafat, 2007). For BSB/T, fidelity was assessed in one way by the

number of times themes related to fidelity were mentioned in facilitator logbook comments. It was found that only one type of fidelity theme (cultural modifications) was related to lower program outcomes for BSB classes. However, for BST classes, more mentions of fidelity themes (fidelity, limited time, and culture) were related to better program outcomes, which is not consistent with the literature.

Another purpose of this evaluation study was to identify program adaptations and how those adaptations contributed to program effects. It is important to gain more insights on the issue of program adaptations given the extant debate about whether adhering to the curriculum or making adaptations based on participants needs' is the best alternative. Although previous research has stressed the importance of adhering to the curriculum, program adaptations are inevitable (Bauman et al.,1990; Cantu et al., 2010; Domitrovich et al., 2010; Dumas et al., 2001; Kalfat et al., 2007). The degree of acceptable adaptations, however, is still a widely debated topic. For BSB/T, there was not a direct question in the facilitator logbooks addressing program adaptations. Therefore, program modifications were measured by using the "success" rating of each activity in the program. A high success rating meant that the facilitator delivered the activity as intended. Results from this evaluation showed that BSB cohorts' increase in parental self-efficacy was *inversely* related to successful delivery of several activity domains (child and brain development, prenatal care, accessing community resources, and program processes), which contradicts what was expected. Discussions with BSB staff did not shed any light on what might explain this contrary finding. However, as expected, the BST cohort's program outcomes were positively related to successful delivery of several activity domains. Thus, the BST findings suggest that minimal adaptations to child and brain development, discipline,

social/emotional needs, stress management, and social support were related to better program outcomes.

Two of the content codes in the facilitators' logbook narratives also concerned adaptations to the BSB/T curriculum, these being "fidelity" and "adaptations." When facilitators mentioned these themes, they typically were in reference to changes that were made during the class and ways to improve the classes in the future. For example, facilitators often ran out of time and were unable to finish presentations, which regarded fidelity. Facilitators would then provide a suggestion for what part of the presentation could be removed, so in the future, facilitators would not run out of time, which regarded adaptations.

The third research question was whether greater dosage contributed to better program outcomes. Although most researchers believe that dosage is one of the most influential dimensions that contribute to program outcomes, a poorly run program will lead to poor results no matter the amount of dosage (Cross et al., 2010). Therefore, for this evaluation, measures of attendance as well as how time was allocated during class were used as measures of dosage. In this evaluation, the amount of time participants attended class was almost always negatively related to program outcomes. The dosage variable was negatively skewed given that very few participants attended fewer than 7 of the 10 sessions. Thus, additional dosage measures used in this study offer more important information regarding how dosage relates to program outcomes. The BSB classes had better program outcomes when more time was allotted for participants to socialize with one another, suggesting perhaps that building social support was important to benefiting from the program. Additionally, in the BST classes, time allotted for didactic facilitator presentations was actually related to negative program outcomes. These findings suggest that greater time for parents to interact with one another, perhaps to apply what they

41

were learning through role plays and discussions, and less time on facilitator presentations could produce better program outcomes. In addition to looking at how time was allocate for each cohort, time spent on each activity curriculum also provide valuable dosage information. There were two activity curriculum domains that were related to better program outcomes for BSB. These included health and safety and stress management. For BST, child and brain development, health and safety, parental self-efficacy, and identifying and accessing community resources were all related to better program outcomes.

The final research question of this evaluation study concerned whether positive participant engagement contributed to greater program effects. Without participants' willingness to work with facilitators and be receptive to the program material, positive program outcomes would be impossible (Heinicke et al., 2000; Korfmacher et al., 2008). True to the literature, BST cohorts showed that positive participant engagement (e.g., participation, accepting new ideas, and respecting others' perspectives) were all positively related to program outcomes (Heinicke et al., 2000; Korfmacher et al., 2008). The resistance theme was negatively associated with program outcomes, which supports what past research has found (Korfmacher at al., 2008).

From these findings, certain implementation variables seem to be more consistently and powerfully related to program benefits, which can help guide facilitators on how to deliver the BSB/T program. First, performing minimal program adaptations to the presentations on child and brain development, discipline, social/emotional needs, stress management, and social support is important. Second, providing parents with time to socialize with one another, whether it is during mealtime or during group activities, is important for BSB/T. This suggestion is consistent with the literature showing that engagement and building support are important elements of effective programs (Heinicke et al., 2000; Korfmacher et al., 2008). Allowing sufficient time for health

42

and safety, stress management, child and brain development, parental self-efficacy, identifying and accessing community resources activities is essential. Finally, creating an environment that will encourage participant engagement is important for program outcomes.

Limitations. Several limitations qualify the strength of conclusions that can be drawn from this implementation evaluation of Best Start for Babies/Toddlers. First, gain scores on the outcome measures were averaged across individuals in each cohort, but implementation was assessed at the group level. In such circumstances, it would have been more appropriate to use multilevel modeling to examine nested effects, but there was insufficient power to conduct such analyses. Typically, it is recommended that multilevel modeling be employed with samples of 25 or more units of analysis (e.g., BSB cohorts). Second, the fidelity rating is currently assessed with a rating of the *success* of activity, not whether the activity was delivered as designed versus adapted. That is, it is not a direct measure of how well the facilitator adhered to the curriculum. Thus, future trials of BSB/T should include questions directly related to fidelity to protocol versus curricular adaptations on the facilitator logbooks. Finally, it is difficult to get a complete picture of participant engagement when relying solely on facilitator ratings. A participant survey asking the same four engagement ratings and two facilitator ratings could be given at the end of each class to better gauge participant engagement. Or, a standardized measure of participant perceptions of the program, such as the Workshop Environment Scale (Moos & Tricket, 1987), could be used.

Conclusions. Overall, this exploratory study was an initial step to evaluate the implementation of the Best Start for Babies and Best Start for Toddlers Program in Northern Colorado. This study helped determine what parts of the curriculum should not be adapted. This study also helped determine how best to allot time during classes based on program outcomes.

43

Finally, consistent with the literature, BSB/T found participant engagement to be related to

program outcomes. With minor adjustments to facilitator logbooks and a greater sample size,

future evaluation could give valuable insight on how best to deliver this program to parents in the

local community.

References

Abbott, R. D., O'Donnell, J., Hawkins, D. J., Hill, K. G., Kosterman, R., & Catalano, R. F. (1998). Changing teaching practices to promote achievement and bonding to school. *American Journal of Orthopsychiatry, 68*, 542-552. doi: 10.1037/h0080363

Azzi-Lessing, L. (2011). Home visitation programs: Critical issues and future directions. *Early Childhood Research Quarterly, 26*, 387-398. doi: 10.1016/j.ecresq.2011.03.005

Bagnato, S. J., Suen, H. K., & Fevola, A. V. (2011). "Dosage" effects on developmental progress during early childhood intervention. *Infants and Young Children, 24*, 117-132. doi: 10.1097/IYC.0b013e3182104896

Battistich, V., Shaps, E., Watson, M., & Solomon, D. (1996). Prevention effects of the child development project: Early findings from an ongoing multisite demonstration trial. *Journal of Adolescent Research, 11*, 12-35. doi: 10.1177/0743554896111003

Bauman, L. J., Stein, R. E. K., & Ireys, H. T. (1991). Reinventing fidelity: The transfer of social technology among settings. *American Journal of Community Psychology, 19*, 619-639.

Botvin, G. J., Baker, E., Dusenbury, L., Tortu, S., & Botvin, E. M. (1990). Preventing adolescent drug abuse through a multimodal cognitive-behavioral approach: Results of a 3-year study. *Journal of Consulting and Clinical Psychology, 58,* 437-446. doi: 10.1037/0022-006X.58.4.437

Burke, R. V., Oats, R. G., Ringle, J. L., Fichtner, L., O., & DelGaudio, M. B. (2011). Implementation of a classroom management program with urban elementary schools in low-income neighborhoods: Does program fidelity affect student behavior and academic outcomes? *Journal for Education for Students Placed at Risk, 16*, 201-218. doi: 10.1080/10824669.2011.585944

Cantu, A., Hill, L., & Becker, L. (2010). Implementation quality of a family-focused preventative intervention in a community-based dissemination. *Journal of Children's Services, 5*, 18-30. doi: 10.5042/jcs.2010.0692

Castro, F. G., Barrera, M., & Martinez, C. R. (2004). The cultural adaptations of prevention interventions: Resolving tensions between fidelity and fit. *Prevention Science, 5*, 41-45. doi: 10.1023/B:PREV.0000013980.12412.cd

Cross, A. B., Gottfredson, D. C., Wilson, D. M., Rorie, M., & Connel, N. (2010). Implementation quality and positive experiences in after-school programs. *American Journal of Community Psychology, 45,* 370-380. doi: 10.1007/s10464-010-9295-z

Dane, A. V., & Schneider, B. H. (1998). Program integrity in primary and early secondary prevention: Are implementation effects out of control. *Clinical Psychology Review, 18*, 23-45. doi: 10.1016/S0272-7358(97)00043-3

Daro, D., McCurdy, K., Falconnier, L., & Stojanovic, D. (2003). Sustaining new parents in home visitation services: Key participant and program factors. *Child Abuse and Neglect, 27*, 1101-1125. doi: 10.1016/j.chiabu.2003.09.007

Domitrovich, C. E., Gest, S. D., Jones, D., Gill, S., Sanford DeRousie, R. M. (2010). Implementation quality: Lessons learned in the context of the head start REDI trial. *Early Childhood Research Quarterly, 25*, 284-298. doi: 10.1016/j.ecresq.2010.04.001

Dubois, D. L., Holloway, B. E., Valentine, J. C., & Cooper, H. (2002). Effectiveness of mentoring programs for youth: A meta-analytic review. *American Journal of Community Psychology, 30*, 157-198. doi: 10.1023/A:1014628810714

Durlak, J. A., & DuPre, E, P. (2008). Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology, 41*, 327-350. doi: 10.1007/s10464-008-9165-0

Dumas, J. E., Lynch, A. M., Laughlin, J. E., Smith, E. P., & Prinz, R. J. (2001). Promoting intervention fidelity: Conceptual issues, methods, and preliminary results from the early alliance prevention trial. *American Journal of Preventative Medicine, 20*, 38-47. doi: 10.1016/S0749-3797(00)00272-5

Dusenbury, L., Brannigan, R., Falco, M., & Hansen, W. B., (2003). A review of research on fidelity of implementation: Implications for drug use prevention in school settings. *Health Education Research, 18*, 237-256. doi: 10.1093/her/18.2.237

Eisner, M., & Meidert, U. (2011). Stages of parental engagement in a universal parent training program. *The Journal of Primary Prevention, 32,* 83-93. doi: 10.1007/s10935-011-0238-8

Elliot, D. S., & Mihalic, S. (2004). Issues in disseminating and replicating effective prevention program. *Prevention Science, 5*, 47-53. doi: 10.1023/B:PREV.0000013981.28071.52

Ennett, S. T., Haws, S., Ringwalt, C. L., Vincus, A. A., Hanley, S., Bowling, J. M., & Rohrbach, L.A. (2011). Evidence-based practice in school substance use prevention: fidelity of implementation under real-world conditions. *Health Education Research, 26*, 361-371. doi: 10.1093/her/cyr013

Flay, B. R., Biglan, A., Boruch, R., Castro, F. G., Gottfredson, D., Kellam, S., Mościcki, E. K.,…Ji, P. (2005). Standards of evidence: Criteria for efficacy, effectiveness and dissemination. *Prevention Science, 6*, 151-175. doi: 10.1007/s11121-005-5553-y

Forgatch, M. S., Patterson, G. R., & DeGarmo, D. S. (2005). Evaluating fidelity: Predictive validity for a measure of competent adherence to the Oregon model of parent management training. *Behavior Therapy, 36*, 3-13. doi: 10.1016/S0005-7894(05)80049-8

Gross, D., Julion, W, & Fogg, L. (2001). What motivates participation and dropout among low-income urban families of color in a prevention intervention. *Family Relations, 50*, 246-254.

Heinicke, C. M., Goorsky, M., Moscov, S., Dudley, K., Gordon, J., Schneider, C. & Guthrie, D., (2000). Relationship-based intervention with at-risk mothers: Facorts affecting variations in outcome. *Infant Mental Health Journal, 21*, 133-155. doi: 10.1002/1097-0355(200007)21:3<133::AID-IMHJ1>3.0.CO;2-P

Hill, L. G., Maucione, K., & Hood, B. K. (2007). A focused approach to assessing program fidelity. *Prevention Science, 8*, 25-34.

Huang, K., Caughy, M., Genevro, J. L., & Miller, T. L. (2005). Maternal knowledge of child development and quality of parenting among White, African-American and Hispanic mothers. *Journal of Applied Developmental Psychology*, *26*, 149-170. doi:10.1016/j.appdev.2004.12.001

Kalafat, J., Illback, R. J., & Sanders Jr, D. (2007). The relationship between implementation fidelity and educational outcomes in a school-based family support program: Development of a model for evaluating multidimensional full-service programs. *Evaluation and Program Planning, 30*, 136-148. doi: 10.1016/j.evalprogplan.2007.01.004

Korfmacher, J., Green, B., Staerkel, F., Peterson, C., Cook, G., Roggman, L.,…Schiffman, R. (2008). Parent involvement in early childhood home visiting. *Child Youth Care Forum, 37*, 171-196. doi: 10.1007/s10566-008-9057-3

Korfmacher, J., Kitzman, H., & Olds, D. (1999). Intervention processes as predictors of outcomes in a preventative home visitation program. *Journal of Community Psychology, 26*, 49-64. doi: 10.1002/(SICI)1520-6629(199801)26:1<49::AID-JCOP5>3.0.CO;2-X

Kumpfer, K. L., Alvarado, R., Smith, P., & Bellamy, N. (2002). Cultural sensitivity and adaptation in family0based prevention interventions. *Prevention Science, 3*, 241-246. doi: 10.1023/A:1019902902119

MacPhee, D. (1981). *Manual for the Knowledge of Infant Development Inventory.* Unpublished manuscript, University of North Carolina.

MacPhee, D., Benson, J.B., & Bullock, D. (1986, April). *Influences on maternal self-perceptions*. Poster presented to the 5th biennial International Conference on Infant Studies, Los Angeles, CA.

MacPhee, D., Fritz, J., & Miller-Heyl, J. (1996). Ethnic variations in personal social networks and parenting. *Child Development, 67*, 3278-3295.

Miller-Heyl, J., MacPhee, D., & Fritz, J. (1998). DARE to be You: A family-support, early prevention program. *Journal of Primary Prevention, 18*, 257-285. doi:10.1023/A:1024602927381

Miller-Heyl, J., MacPhee, D., Walker, A. K., & Podunovich, R. (2013). *Ready families: Effects of a community-based, family-school Bridge Program.* Manuscript submitted for publication.

Maher, E. J., Marcynyszyn, L. A., Corwin, T. W., & Hodnett, R. (2011). Dosage matters: The relationship between participation in the nurturing parenting program for infants, toddlers, and preschoolers and subsequent child maltreatment. *Children and Youth Services Review, 33*, 1426-1434. doi: 10.1016/j.childyouth.2011.04.014

McGuigan, W. M., Katzev, A. R., & Pratt, C. C. (2003). Multi-level determinants of mother's engagement in home visitation services. *Family Relations*, 52, *271-278.*

Moos, R.H., & Trickett, E. .J. (1987). *Classroom environment scale* (2nd ed.) Palo Alto, CA: Consulting Psychologists Press.

Nation, M., Crustco, C., Wandersman, A., Kumpfer, K. L., Seybolt, D., Morrissey-Kane, E., & Davino, K. (2003). What works in prevention: Principles of effective prevention programs. *American Psychologist*, 58, 449-456. doi: 10.1037/0003-066.58.6-7.449

Nix, R. L., Bierman, K. L., & McMahon, R. J. (2009). How attendance and quality of participation affect treatment response to parent management training. *Journal of Consulting and Clinical Psychology, 77*, 429-438. doi: 10.1037/a0015028

Perepletchaikova, F., Treat, T. A., & Kazdin, A. E. (2007). Treatment integrity in psychotherapy research: Analysis of the studies and examination of the associated factors. *Journal of Consulting and Clinical Psychology, 75,* 829-841. doi: 10.1037/0022-006X.75.6.829

Raikes, H., Green, M. L., Atwater, J., Kisker, E., Constantine, J., & Chazan-Cohen. R. (2006). Involvement in Early Head State home visiting services: Demographic predictors and relations to child and parent outcomes. *Early Childhood Research Quarterly, 21*, 2-24. doi: 0.1016/j.ecresq.2006.01.006

Roggman, L.A., Boyce, L. K., Cook, G. A., & Cook, J. (2002). Getting dads involved: Predictors of father involvement in Early Head Start and with their children. *Infant Mental Health Journal, 23*, 62-78. doi: 10.1002/imhj.100004

Scanlon, J. W., Horst, P., Nay, J. W., Schmidt, R. E., & Waller, J. D. (1977). Evaluability assessment: Avoiding type III or type IV errors. In G. R. Gilbert and P.J. Conklin (Eds.), *Evaluation Management: A source book of readings.* Charlottesville: U.S. Civil Service Commission. From the Kutash Article (Description of Fidelity implementation system:

an example from a community based children's mental health program.

Saias, T., Lerner, E., Greacen, T., Simon-Vernier, E., Emer, A., Pintaux, E.,…Revah-Levy, A. (2012). Evaluating fidelity in home visiting programs: A qualitative analysis of 1058 home visit case notes from 105 families. *PLoS ONE* , 7*,* 1-10. doi: 10.1371/journal.pone.0036915

Shonkoff, J., & Phillips, D. (2000). *From neurons to neighborhoods: The science of early childhood development.* Washington, DC: National Academy Press.

Spoth, R., Redmond, C., Hockaday, C., & Shin, C. Y. (1996). Barriers to participation in family skills prevention interventions and their evaluations: a replication and extension. *Family Relations, 45*, 247-254.

Summerfelt, W. T. (2003). Program strength and fidelity in evaluation. *Applied Developmental Science, 7*, 55-61. doi: 10.1207/S1532480XADS0702_2

Wagner, M. M., & Clayton, S. L. (1999). The Parents as Teachers program: results from two demonstrations. *The Future of Children, 9*(1), 91-115.

Wasik, B. H., Ramey, C. T., Bryant, D. M., Sparling, J. J. (1990). A longitudinal study of two early intervention strategies: Project CARE. *Child Development, 61*, 1682-1696. doi: 10.2307/1130831

Weber, R. P. (1990). *Basic content analysis* (2nd ed.). Newbury Park, CA: Sage.

Appendix A
Cohort Specific Profiles

<u>**Cohort One**</u>

Table 13
Percent Time Devoted and Average Ratings (AR) for Each Activity Domain

	Domain											
	1	2	3	4	5	6	7	8	9	10	11	12
Total %	.48	.04	.00	.24	.08	.04	.04	.04	.00	.00	.04	.00
AR	2.20	1.00	---	2.30	3.00	2.00	3.00	3.00	---	---	2.00	---

Note. Ratings were on a scale from 1 (*needs some work*) to 3 (*very successful*)

Table 14
Minutes Distributions in Each Session

Meals/Socializing	Individual/ Group Activities	Presentations
44.06	47.19	83.13

Table 15
Gain Scores

KIDI	SPRR
.54	---

Note. Gain scores on the SPRR are not available for this cohort.

<u>**Cohort Two**</u>

Table 16
Percent Time Devoted and Average Ratings (AR) for Each Activity Domain

	Domain											
	1	2	3	4	5	6	7	8	9	10	11	12
Total %	.10	.03	.18	.28	.08	.03	.10	.08	.05	.03	.03	.03
AR	2.89	2.00	2.79	2.63	2.67	3.00	3.00	2.67	3.00	2.50	2.50	3.00

Note. Ratings were on a scale from 1 (*needs some work*) to 3 (*very successful*)

Table 17
Minutes Distributions in Each Session

Meals/Socializing	Individual/ Group Activities	Presentations
51.25	45.25	75.50

Table 18
Gain Scores

KIDI	SPRR
.55	---

Note. Gain scores on the KIDI are not available for this cohort.

Cohort Three

Table 19
Percent Time Devoted and Average Rating (AR) for Each Activity Domain

	Domain											
	1	2	3	4	5	6	7	8	9	10	11	12
Total %	.18	.03	.18	.18	.09	.03	.12	.03	.03	.00	.06	.09
AR	2.67	3.00	2.83	2.58	3.00	3.00	2.63	3.00	3.00	.00	3.00	2.83

Note. Ratings were on a scale from 1 (*needs some work*) to 3 (*very successful*)

Table 20
Minutes Devoted In Each Session

Meals/Socializing	Individual/ Group Activities	Presentations
32.22	58.89	83.33

Table 21
Gain Scores

KIDI	SPRR
.45	.39

Cohort Four

Table 22
Percent Time Devoted and Average Rating (AR) for Each Activity Domain

	Domain											
	1	2	3	4	5	6	7	8	9	10	11	12
Total %	.11	.00	.06	.42	.06	.00	.11	.08	.06	.03	.03	.06
AR	2.63	---	2.75	2.60	3.00	---	2.75	2.67	3.00	2.00	1.00	2.50

Note. Ratings were on a scale from 1 (*needs some work*) to 3 (*very successful*)

Table 23
Minutes Devoted In Each Session

Meals/Socializing	Individual/ Group Activities	Presentations
54.17	44.72	75.00

Table 24
Gain Scores

KIDI	SPRR
.52	1.14

Table 25
Percent Time Devoted and Average Rating (AR) for Each Activity Domain

						Domain						
	1	2	3	4	5	6	7	8	9	10	11	12
Total %	.14	.06	.06	.36	.06	.03	.11	.08	.03	.03	.00	.06
AR	3.00	2.50	3.00	2.92	3.00	3.00	2.63	3.00	3.00	3.00	---	3.00

Note. Ratings were on a scale from 1 (*needs some work*) to 3 (*very successful*)

Table 26
Minutes Devoted In Each Session

Meals/Socializing	Individual/ Group Activities	Presentations
61.25	34.75	83.25

Table 27
Gain Scores

KIDI	SPRR
.76	.67

Cohort Six

Table 28
Percent Time Devoted and Average Rating (AR) for Each Activity Domain

						Domain						
	1	2	3	4	5	6	7	8	9	10	11	12
Total %	.25	.04	.07	.25	.11	.00	.07	.04	.07	.00	.04	.07
AR	3.00	3.00	3.00	2.86	3.00	---	3.00	3.00	3.00	---	3.00	3.00

Note. Ratings were on a scale from 1 (*needs some work*) to 3 (*very successful*)

Table 29
Minutes Devoted In Each Session

Meals/Socializing	Individual/ Group Activities	Presentations
34.44	38.61	78.89

Table 30
Gain Scores

KIDI	SPRR
-.22	-.08

Cohort Seven

Table 31
Percent Time Devoted and Average Rating (AR) for Each Activity Domain

						Domain						
	1	2	3	4	5	6	7	8	9	10	11	12
Total %	.25	.04	.11	.25	.04	.00	.07	.04	.04	.04	.11	.04
AR	2.50	2.00	2.50	2.86	3.00	---	2.50	3.00	3.00	3.00	2.50	3.00

Note. Ratings were on a scale from 1 (*needs some work*) to 3 (*very successful*)

Table 32
Minutes Devoted In Each Session

Meals/Socializing	Individual/ Group Activities	Presentations
45.71	71.29	78.21

Table 33
Gain Scores

KIDI	SPRR
.41	.54

Cohort Eight

Table 34
Percent Time Devoted and Average Rating (AR) for Each Activity Domain

						Domain						
	1	2	3	4	5	6	7	8	9	10	11	12
Total %	.22	.03	.13	.28	.06	.00	.09	.03	.06	.03	.06	.00
AR	2.86	3.00	3.00	2.78	2.75	---	3.00	3.00	3.00	3.00	2.75	---

Note. Ratings were on a scale from 1 (*needs some work*) to 3 (*very successful*)

Table 35
Minutes Devoted In Each Session

Meals/Socializing	Individual/ Group Activities	Presentations
47.48	44.44	85.56

Table 36
Gain Scores

KIDI	SPRR
.41	-.01

Table 37
Percent Time Devoted and Average Rating (AR) for Each Activity Domain

						Domain						
	1	2	3	4	5	6	7	8	9	10	11	12
Total %	.17	.08	.08	.17	.03	.03	.06	.08	.06	.06	.08	.11
AR	2.83	2.50	3.00	2.83	3.00	3.00	2.25	2.50	3.00	3.00	2.50	3.00

Note. Ratings were on a scale from 1 (*needs some work*) to 3 (*very successful*)

Table 38
Minutes Devoted In Each Session

Meals/Socializing	Individual/ Group Activities	Presentations
37.50	40.83	92.22

Table 39
Gain Scores

KIDI	SPRR
.00	-.22

.

.

Cohort Ten

Table 40
Percent Time Devoted and Average Rating (AR) for Each Activity Domain

Domain

Cohort 11

Table 43
Percent Time Devoted and Average Rating (AR) for Each Activity Domain

						Domain						
	1	2	3	4	5	6	7	8	9	10	11	12
Total	.16	.11	---	.21	.11	.05	.11	.11	.00	.00	.11	.05
AR	3.00	3.00	---	2.88	3.00	3.00	2.75	3.00	---	.00	2.25	2.50

Note. Ratings were on a scale from 1 (*needs some work*) to 3 (*very successful*)

Table 44
Minutes Devoted In Each Session

Meals/Socializing	Individual/ Group Activities	Presentations
36.94	53.61	89.44

Table 45
Gain Scores

KIDI	SPRR
.80	.64

Cohort 12

Table 46

Percent Time Devoted and Average Rating (AR) for Each Activity Domain

	1	2	3	4	5	6	Domain 7	8	9	10	11	12
Total	.08	.15	---	.15	.08	.04	.15	.12	.08	.00	.00	.15
AR	2.75	2.38	---	2.63	3.00	3.00	2.38	2.67	2.50	---	---	2.70

Note. Ratings were on a scale from 1 (*needs some work*) to 3 (*very successful*)

Table 47

Minutes Devoted In Each Session

Meals/Socializing	Individual/ Group Activities	Presentations
36.25	46.00	91.75

Table 48

Gain Scores

KIDI	SPRR
.12	.43

Cohort 13

Table 49

Percent Time Devoted and Average Rating (AR) for Each Activity Domain

	1	2	3	4	5	6	Domain 7	8	9	10	11	12
Total	.15	.05	---	.20	.15	.05	.05	.15	.00	.00	.00	.20
AR	3.00	3.00	---	2.88	2.83	3.00	3.00	2.83	---	---	---	2.70

Note. Ratings were on a scale from 1 (*needs some work*) to 3 (*very successful*)

Table 50

Minutes Devoted In Each Session

Meals/Socializing	Individual/ Group Activities	Presentations
41.75	42.25	83.00

Table 51

Gain Scores

KIDI	SPRR
.58	.43

Figure 3. Percentages for activity domains across all BSB cohorts. Shown are the minimum, mean, and maximum percentages for activity domains across all 10 BSB cohorts.

Figure 4. Time in minutes (dosage) devoted to meals, activities, and presentations across all BSB cohorts. Shown are the minimum, mean, and maximum percentages, by type of curricular activity.

Figure 5. Percentages for activity domains across all BST cohorts. Shown are the minimum, mean, and maximum percentages for activity domains across all 3 BST cohorts.

Figure 6. Time in minutes (dosage) devoted to meals, activities, and presentations across all BST cohorts. Shown are the minimum, mean, and maximum percentages, by type of curricular activity

Table 52
Percentage Total of Themes in Codebooks (BSB)

Cohort	Fidelity	Limited Time	Improvements	Culture	Program Delivery	Resistance	Practice	Disruption	Engagement	Group Dynamics	Benefits
1	.08	.01	.18	.05	.14	.10	.00	.01	.27	.13	.04
2	.03	.05	.14	.02	.08	.22	.03	.00	.31	.07	.03
3	.00	.00	.16	.03	.08	.16	.03	.02	.37	.06	.06
4	.03	.08	.02	.02	.09	.23	.02	.03	.35	.12	.03
5	.00	.00	.04	.00	.16	.04	.02	.00	.50	.20	.05
6	.00	.00	.00	.05	.10	.15	.00	.05	.45	.05	.15
7	.09	.00	.11	.04	.16	.13	.00	.00	.33	.07	.07
8	.09	.07	.09	.02	.07	.13	.00	.05	.41	.05	.02
9	.11	.02	.13	.04	.05	.23	.02	.02	.36	.04	.00
10	.06	.05	.08	.00	.08	.09	.03	.00	.42	.17	.03

Table 53
Percentage Total of Themes in Codebooks (BST)

Cohort	Fidelity	Limited Time	Improvements	Culture	Program Delivery	Resistance	Practice	Disruption	Engagement	Group Dynamics	Benefits
11	.07	.03	.05	.16	.08	.18	.00	.00	.39	.02	.02
12	.03	.03	.16	.06	.01	.33	.01	.01	.29	.04	.01
13	.08	.00	.05	.00	.13	.05	.08	.08	.33	.15	.03

Table 54
Correlation between Fidelity Mentions in Facilitator Logbooks and Gain Scores

	BSB		BST	
	KIDI	SPRR	KIDI	SPRR
Fidelity	.22	.17	-.10	.61
Limited Time	-.10	-.38	.87	.30
Culture	-.60[+]	-.31	.48	.94
Improvements	.23	-.28	-.49	-.49

*$p < .05$ [+]$p < .10$

Table 55

Correlation between Activity Success Rating and Gain Scores, by Program

Domain	BSB		BST	
	KIDI	SPRR	KIDI	SPRR
Development	-.26	-.49	.95	.48
Discipline	-.40	-.35	.95	.48
Healthy prenatal care	-.32	-.49		
Health/Safety	-.24	-.44	.95	.48
Social/Emotional needs	-.23	.38	-.20	.52
Parent efficacy	-.15	---	---	---
Stress management	-.09	.05	.74	.09
Social support	.10	.15	.98	.87
Attachment	-.12	.22	---	---
Child Care	-.09	-.26	---	---
Community resources	-.44	-.78*	---	---
Program processes	-.23	-.73+	-.75	-1.0*

*p < .05 +p < .10

Table 56

Correlation between Percent Time Devoted to Each Activity and Gain Scores, by Program

Domain	BSB		BST	
	KIDI	SPRR	KIDI	SPRR
Child and Brain Devel.	-.11	-.42	.97	.56
Discipline	-.27	-.66+	-.63	.06
Healthy Prenatal Care	.06	-.32		
Health/Safety	.49	.76*	.99+	.63
Social/Emotional Needs	-.19	-.14	.55	-.15
Parent Efficacy	.34	-.20	.95	.48
Stress Management	.47	.65+	-.63	.06
Social Support	.24	.32	-.02	-.69
Attachment	-.55+	-.49	-.95	-.48
Child Care	.06	-.10	---	---
Community resources	-.38	-.31	.75	1.0*
Program processes	-.47	-.20	-.49	.74

*p < .05 +p < .10

Table 57

Correlation between Dosage and Gain Scores, by Program

	BSB		BST	
	KIDI	SPRR	KIDI	SPRR
Attendance	-.48	-.67	-.62	.98+
Socializing	.70*	.48	.31	-.42
Activities	.09	.17	.50	.95
Presentations	-.28	-.48	-.44	.28

Note: Socializing refers to the time the parents spent eating meals before class began. Activities refer to individual and group activities that facilitators planned to go along with the presentations.
*p < .05 +p < .10

Table 58

Correlations between Gain Scores and Ratings of Engagement and Facilitator Preparation, by Program

	BSB		BST	
Process Rated	KIDI	SPRR	KIDI	SPRR
Participation	-.14	-.09	.97	.88
Acceptance of new ideas	.02	.42	1.0*	.72
Getting along	-.23	-.05	-.23	.50
Respecting perspectives	-.24	.10	.57	.98
Facilitator preparedness	.11	-.17	-.71	-.04
Facilitator organization	.39	.39	.04	-.64

*p < .05 +p < .10

Table 59

Correlations between Engagement Mentions in Facilitator Logbooks and Gain Scores

	BSB		BST	
	KIDI	SPRR	KIDI	SPRR
Engagement	-.11	-.17	.92	.94
Resistance	-.36	-.01	-.69	-.03
Group Dynamics	.68*	.49	.00	-.67

*p < .05 +p < .10

Table 60

Correlations between Moderating Variables and Gain Scores, by Program

| | BSB | | BST | |
	KIDI	SPRR	KIDI	SPRR
Eng by Fid	-.10	.70+	-.79	-.17
Composite risk score	-.05	-.03	-.12	.09
Risk by Eng	-.20	-.70	.96	.89
Risk by Fid	-.40	-.67+	.79	.17

Note. Eng = Engagement; Fid = Fidelity.

*p < .05 +p < .10