

Assessing Treatment Fidelity: Lessons Learned from a Multi-Site National Study

Eva Gabrielle Mamak

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill
in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the
School of Education (School Psychology)

Chapel Hill
2007

Approved by:

Barbara H. Wasik, Ph.D.

Donna Bryant, Ph.D.

Henry Frierson, Ph.D.

Samuel Song, Ph.D.

Joseph Sparling, Ph.D.

Steven Knotek, Ph.D.

© 2007

Eva Gabrielle Mamak

ALL RIGHTS RESERVED

ABSTRACT

Eva Gabrielle Mamak

Assessing Treatment Fidelity: Lessons Learned from a Multi-Site National Study
(Under the Direction of Barbara H. Wasik, Ph.D.)

Research in the field of early intervention indicates that providing educational services early in a child's life can produce meaningful changes in outcomes for children. The Partners for Literacy curriculum was provided through the Classroom Literacy Interventions and Outcomes study of the federal Even Start program. The Partners for Literacy curriculum was implemented for two years, and implementation was measured over four time points during this study. The purpose of this dissertation was to determine which factors related to positive implementation of this curriculum and to examine changes in curriculum implementation over time. Results of this study indicated that none of the seven measured factors was significant in predicting a model of curriculum implementation in the first year of the study. However, in the second year, a model including the number of teacher feedback forms that were returned was successful in classifying over 80% of classrooms through logistic regression as high implementation or moderate/low implementation, $-2 \text{ Log Likelihood} = 47.143, \chi^2 (2) = 24.250, p < .001$. Results of repeated measures analysis through a SAS proc mixed procedure also indicated that implementation of this curriculum improved significantly over two years, and while director attendance at training was related to increases in implementation, it did

not mediate this effect. Time was the only significant indicator in this model, $t(176) = 10.89, p < .001$. These results suggest that interventions should be in place for a significant period of time prior to measuring their effects, as implementation develops slowly over time.

ACKNOWLEDGEMENTS

This dissertation could never have been completed without the help and support of many people. I gratefully acknowledge my committee members, Dr. Donna Bryant, Dr. Henry Frierson, Dr. Samuel Song, and Dr. Joseph Sparling, who not only guided me through the dissertation process, but who contributed to my graduate experience in meaningful ways. Their expertise in different areas helped my dissertation become well rounded and comprehensive. I would like to thank Dr. Simeonsson for helping to ensure that my dissertation could be completed in a timely manner, and for his help in editing final drafts of this manuscript. I want to thank Dr. Barbara Wasik, my advisor, for all of her expertise, guidance, and support throughout graduate school and my dissertation. I tremendously appreciate the time and effort she has shared with me in guiding me through this dissertation, and all aspects of graduate school.

I would also like to thank the program consultants, Beth Jones, Terry Grimes, Terry McCandies, Gail DelGreco, and Sally Edwards for their commitment to training teachers and directors and consistently working to collect valid and meaningful data. I am grateful for the efforts of Jeff Lyons, our project coordinator, and the rest of the graduate assistants for the Partners for Literacy project, for their efforts to clean and track down meaningful, reliable, and complete data. I would like to thank the teachers, directors, parents and children participating in the Classroom Literacy Interventions and Outcomes study as well.

I also am incredibly grateful to all of my friends and family who were beside me throughout graduate school and the dissertation process. They have provided me with the encouragement, support and motivation that made it possible to complete this degree.

TABLE OF CONTENTS

Chapter	Page
I. INTRODUCTION	1
A. Why Study Treatment Fidelity?.....	3
B. Barriers and Facilitators to Treatment Fidelity	7
C. Models of Treatment Implementation	9
D. Implementation in Early Intervention Research	15
1. Abecedarian Project	16
2. Project CARE.....	18
3. High/Scope Perry Preschool Program	20
4. Infant Health and Development Program	20
5. Early Head Start	26
E. Family Literacy	29
F. Classroom Literacy Interventions and Outcomes Study.....	34
G. Partners for Literacy	36
H. Purpose.....	44
I. Research Questions	44

J.	Hypotheses.....	45
II.	METHOD	47
A.	Participants.....	47
B.	Instruments.....	49
C.	Procedure	54
III.	RESULTS	58
IV.	DISCUSSION	62
V.	TABLES	80
VI.	FIGURES.....	93
VI.	APPENDIX A: TEACHER FEEDBACK FORM AND MANUAL.....	97
VII.	APPENDIX B: DIRECTOR/MENTOR-TEACHER FEEDBACK FORM AND MANUAL.....	103
VIII.	APPENDIX C: PROGRAM CONSULTANT CODING SYSTEM – RATINGS OF SITE IMPLEMENTATION AND CLASSROOM QUALITY	109
IX.	REFERENCES	110

LIST OF TABLES

Table 1	Role of Staff	80
Table 2	Ethnic and Gender Distribution of Staff	81
Table 3	Descriptive Data – Implementation Factors	82
Table 4	Descriptive Data – Categorical Implementation Factors	83
Table 5	Descriptive Data – Site Ratings	84
Table 6	Bivariate Correlations – Year 1	85
Table 7	Bivariate Correlations – Year 2	87
Table 8	Logistic Regression – Year 1	89
Table 9	Logistic Regression – Year 2	90
Table 10	Mean Initial Ratings of Attrition Group and Non-Attrition Group	91
Table 11	Model of Time and Director Training on Rating.....	92

LIST OF FIGURES

Figure 1	Research Design for the Classroom Literacy Interventions and Outcomes Study	94
Figure 2	Partners for Literacy Site Locations	95
Figure 3	Boxplot of Mean Ratings at Four Time Points	96

CHAPTER 1

INTRODUCTION

Early childhood intervention is based on the assumption that providing meaningful educational experiences early in life can produce significant positive academic and social outcomes for young children. Using a variety of measures, research has shown that specific intervention procedures have been able to achieve positive and significant child outcomes using a variety of measures. We know considerably less, however, about treatment fidelity or the procedures and processes of interventions and how these contribute to the success or failure of interventions.

Treatment procedures are those variables introduced to bring about outcome changes. When outcomes are not significant, the treatment is typically assumed to have been ineffective. Yet, without information on whether the treatment was actually implemented and to what degree, we can't say with certainty that a specific treatment was not effective. Furthermore, when a treatment is effective in bringing about significant outcomes, we don't know which factors contributed to this change, nor do we know what intensity of the treatment was associated with different outcome levels.

Increasingly, as the field of early intervention has become more sophisticated, questions about fidelity to an intervention or treatment are being raised (Carroll et al., 2000; Domitrovich & Greenberg, 2000; Goss, Noltemeyer, & Devore, 2007; St.Pierre, Swartz, Murray, & Deck, 1996). Increased numbers of components in an intervention lead to increased complexity of that intervention, and the complexity can cause difficulty

in delivering the intervention in a uniform way. Increased attention to treatment implementation, especially in regards to multi-site studies, has resulted in compelling questions about whether a specific intervention has been implemented as planned.

Many early childhood classroom interventions can be divided into several components in order to obtain a more detailed analysis of the intervention procedures. These components include the intervention or curriculum materials used in teaching (such as books or learning activities), specific teaching techniques to be used in the classroom (such as scaffolding children's learning), professional development (including the participation of staff in workshops and ongoing consultation) and assessment and monitoring of the intervention. Studies examining the effectiveness of a curriculum need to address all these components to gain an understanding of treatment fidelity.

The goal of the current study is to examine treatment fidelity in large scale intervention by using data from the implementation of the Partners for Literacy intervention. The Partners for Literacy Curriculum is an early childhood and parenting curriculum with numerous components, including a multi-component early childhood curriculum, parenting education, and methods for integrating early childhood education and parenting education. Common strategies and materials are used across program components. In a recent, multi-site experimental study using the Partners for Literacy Curriculum, the developers collected extensive treatment fidelity data to draw conclusions about how well it had been implemented. This study will examine models of the factors contributing to treatment fidelity for the Partners for Literacy curriculum, correlate implementation factors with program variables that could influence

implementation, and describe the levels of treatment fidelity achieved over the first two years of the intervention research.

Why Study Treatment Fidelity?

In general, program implementation is generally defined as the extent to which an intervention is actually carried out (Paulson, Post, Herinckx, & Risser, 2002; Lichstein, Riedel, & Grieve, 1994). *Treatment fidelity* refers to the extent to which a treatment is executed as intended. While similar, treatment fidelity and implementation are distinct concepts. Treatment implementation generally includes the process by which an intervention is carried out, while fidelity is primarily focused on the degree to which the intervention was carried out, with less focus on process. Moncher and Prinz (1991) were the first to formally define treatment fidelity as “confirmation that the manipulation of the independent variable occurred as planned” (p. 247). This definition underscores the necessity of treatment fidelity to confirm internal validity. Treatment fidelity specifies the extent to which internal validity is achieved within a study, that is, the extent to which we can be certain that the results of an experiment can be attributed to the manipulation of the independent variable, rather than to another confounding variable.

In order to attribute outcomes to changes in a specific treatment variable, one must document that the treatment variable was successfully manipulated in the intended manner. Treatment implementation includes both treatment adherence and competence. Santacroce, Maccarelli and Grey (2004) defined intervention *adherence* as “...the degree to which prescribed elements of an intervention are delivered and proscribed elements avoided...,” a specific aspect of treatment implementation. Intervention *competence* has been defined as the degree to which interventionists are trained and knowledgeable about

the use of the treatment or intervention under study, and the extent to which interventionists are prepared to implement the treatment (Santacroce et al., 2004). Santacroce and colleagues noted that the assumption that interventions have been delivered in a consistent manner is imperative to conclude that findings are valid, yet determining adherence to the intervention protocol and competence in delivering interventions is not the standard in research. For example, Gresham and colleagues (2000) reviewed intervention articles in several journals, and found that only 49% described integrity, and only 18.5% measured treatment integrity.

Further justification for the necessity of treatment fidelity data is its importance in the inference of external validity. Moncher and Prinz (1991) noted that the “...verification of fidelity is necessary to ensure that fair, powerful, and valid comparisons of replicable treatments can be made” (p. 247), and that the results of a study can be generalized to suitable populations. Without knowing the extent to which a treatment was carried out with one population, it would be difficult to determine the extent to which that treatment might be extended or generalized to other populations.

Even within a study, the effectiveness of different interventions cannot be compared without first examining and comparing implementation of each treatment. Ignoring treatment fidelity can generate false data in a trial that compares the efficacy of two or more interventions (Leventhal & Friedman, 2004). For example, in a study comparing two different treatments, the experimental group for treatment 1 might be delivered with high integrity, while the experimental group for treatment 2 might be delivered with low integrity. If outcomes in treatment group 1 are more positive, researchers may believe that this intervention is more effective, when in fact the differences in outcome might be

explained best through differences in delivery of the intervention. A further independent benefit to the study of treatment fidelity is the potential to improve statistical power. Statistical power is improved through an increase in treatment effect size and reduction in the number of subjects required to find effects, thereby decreasing the costs and improving efficacy of intervention research programs which choose to maximize treatment fidelity (Bellg et al., 2004).

Treatment fidelity research is helpful in that it can be used for process evaluation, the ongoing evaluation of a treatment that contributes to its development over time. By monitoring the fidelity of a treatment while it is ongoing, program developers can track progress in delivering the treatment, and provide feedback to implementers. Monitoring treatment delivery, and the extent to which the treatment is delivered as intended, can serve to minimize treatment drift over time (Paulson et al., 2002). However, fidelity is rarely measured over a series of time points, as would be required to determine treatment fidelity across the implementation period.

Monitoring fidelity over time can aid program developers in understanding the strengths and weaknesses of designed programs by examining specific variables, allowing them to better understand why measured outcomes occurred in the manner observed (Zins, Elias, Greenberg, & Pruett, 2000). Furthermore, by evaluating fidelity, one can determine the degree to which an intervention differs from a control condition, by determining actual differences between the experimental group and control group. Assessing treatment validity provides the possibility of determining the variation in fidelity among different treatment sites and the degree to which treatment fidelity is associated with desired outcomes (Carroll et al., 2000). Treatment fidelity can also be

examined to determine the minimum level of the treatment that might be beneficial, as well as the determination of the optimal level of implementation, and to verify that all parts of a treatment are necessary in order to achieve positive outcomes.

When reviewing literature regarding the effectiveness of particular interventions, information on what distinguishes the intervention group from the control group, in real and practical terms, contributes to accurate interpretation of outcomes. As Peterson and McConnell (1993) point out, regardless of the potential impact of an intervention, an intervention that is not used is not effective. A recent emphasis on evidence based practices within the field of education has led to the identification of a variety of practices that are supported by research. However, there is little guidance in professional development literature regarding *how* these practices might be implemented in the classroom once they are identified (Winton, 2006). The increased focus on evidence-based practice has led to a focus on treatment implementation within professional development literature. For example, assessing teacher fidelity to the curricula is now recognized as an essential variable in drawing conclusions on the effectiveness of a curriculum. As Goss, Noltemeyer and Devore (2007) note, “if the way in which an intervention is implemented is not monitored or its components defined, it is difficult to build an inventory of evidence-based interventions that can be shared and replicated” (p. 35).

Given the importance of studying fidelity to understanding treatment effects, one would expect this type of research to be common in a variety of fields of study. Treatment fidelity research is common in the fields of medicine (i.e., Bellg et al., 2004; Lichstein et al., 1994; Santacroce et al., 2004), psychotherapy (i.e. Carroll et al., 2000;

Moncher & Prinz, 1991; Waltz, Addis, Koerner, & Jacobson, 1993) and mental health prevention (Dane & Schneider, 1998; Domitrovich & Greenberg, 2000; Nelson, Amio, Prilleltensky, & Nickels, 2000; Zins et al., 2000). Generally, treatment fidelity studies in medicine, psychotherapy, and mental health prevention fall into two broad categories. One category addresses barriers and facilitators to treatment implementation, sometimes measuring the relative impact of these factors on outcomes. This category of studies generally addresses single site interventions (e.g., Meyer, Nicholson, Danish, Fries, & Polk, 2000; Noell et al., 2005; Waltz et al., 1993; Domitrovich & Greenberg, 2000). A second category of interest includes studies that propose a model of treatment implementation and report whether components of that model are used, sometimes using case studies (e.g., Kitson, Harvey & McCormack, 1998). Both categories will be examined in further depth.

Barriers and Facilitators to Treatment Fidelity

Many studies attempt to measure some absolute level of fidelity to treatment, or some proportion of participants who implemented a given program (i.e., Meyer, Nicholson, Danish, Fries, & Polk, 2000; Noell et al., 2005; Waltz et al., 1993; Domitrovich & Greenberg, 2000). This goal is achieved through a variety of techniques, including the use of fidelity scales and observation of interventionists. One area of study has been the examination of factors that promote treatment fidelity and the barriers that prevent an intervention from being used effectively in practice. These studies have largely been conducted in order to promote stronger fidelity to treatment which can lead to more effective intervention. In their study of the use of respite care in rural Pennsylvania, Petchers, Biegel and Snyder (1991) examined perceived incentives and

barriers to implementation after they found low fidelity to treatment for this type of service. Findings from this study were used by program developers who modified the intervention before its wide-spread delivery, resulting in an overall perceived improvement in the program by recipients.

Studies that relate to barriers and facilitators to treatment fidelity are also found in educational research. Orrill and Anthony (2003) examined the barriers to treatment fidelity found in classrooms attempting to use the standards set forth by the National Council of Teachers of Mathematics. Their study was conducted through qualitative survey and participant observation. This study found that implementation was challenged due to prior beliefs about teaching and learning and lack of teacher commitment to the new program. Teachers had to find a new way to assess learning and knowledge and had to find new strategies to manage behavior in a group format for work. A lack of knowledge about the new curriculum created difficulty in implementation, and teacher discomfort arose from a lack of experience with the materials. These barriers were related to a common teacher attitude that they had not been trained adequately in the curriculum they were expected to implement. The fit between teacher beliefs about teaching mathematics and the techniques employed by the curriculum were important for fidelity to this program. Another barrier in this study was teacher concern that the program was not meeting the needs of lower-achieving students. The authors suggested considering three factors when thinking about implementation of a new curriculum or program: (1) influences both external to the curriculum and internal to the teacher; (2) who has control over various barriers to implementation that were identified; and (3) barriers that might underlie teacher's complaints about a curriculum.

Hutinger and Johanson (2000) also conducted a study of implementation in a program in early childhood special education. The intervention included training and staff development for teachers and other staff in the field of technology. Findings from this study showed that teachers tended to adopt new practices in their classroom when there was a well-trained, on-site team to assist with technology. As well, teachers were more likely to adopt these new practices after they observed positive outcomes for children involved with the intervention, and when they had the opportunity to observe other teachers successfully using the intended practices. One factor that enhanced treatment fidelity was a flexible program, and a good fit between the program and the goals and culture of the school in which it was used. As well, teacher acceptance was important in maintaining implementation of the intervention. Finally, a recommendation from this study is that interventions might be most successfully implemented if they are initially carried out on a small scale, with only those teachers who agree to participate in the program, and expand the program after it has been more widely observed by other teachers (Hutinger & Johanson, 2000).

In summary, the examination of barriers and facilitators to treatment fidelity provide a context in which programs can be modified, as a type of process evaluation, in order to maximize the likelihood of participation in the actual intervention, as well as identifying ways to support treatment fidelity in practice.

Models of Treatment Implementation

To conduct a comprehensive study of implementation, one needs to describe a conceptual model for implementation, including the factors considered necessary in order to determine the level of implementation. Many studies conduct a review of research to

determine what identified factors in an implementation model are commonly used, and what factors are overlooked. One study that used this methodology was Moncher and Prinz's study (1991) of treatment fidelity in psychological intervention research. This study intended to evaluate outcome studies from 1980-1988 to determine researcher attention to the issue of implementation. In their review of 359 research articles, Moncher and Prinz reviewed articles for (1) the presence of training for interventionists, (2) procedures used to promote fidelity, (3) assessment of the aspects of treatment which were verified, (4) what methods were used in assessing fidelity, and (5) whether assessment of fidelity was included in the interpretation of results. The authors found that, in general, significant increases occurred over time in the proportion of studies that examined implementation and the proportion that used supervision to promote implementation, but that the majority of studies still tended to ignore these issues. Overall, only one of every eight studies combined the use of treatment manuals, supervision of treatment practices, and assessment of adherence to protocol in their studies (Moncher & Prinz, 1991).

Another study of this type by Witt and Elliott (1985) included a conceptual model for understanding and evaluating the likelihood of implementation. This theoretical model included four elements of the intervention that the authors believed would relate to the likelihood that the intervention would be carried out as anticipated, thereby achieving desired outcomes. These elements included *intervention acceptability*, *use*, *integrity*, and *effectiveness*. These elements were described as sequential, such that an intervention must be acceptable before it is used, and it must be used at some level before it can be used with integrity. However, these elements are also reciprocal. The likelihood that the

intervention will be considered acceptable will increase if an interventionist actually uses the intervention.

However, all these elements are also dependent on external factors, including the rationale used to describe the treatment, as well as the theoretical orientation used in describing it (Witt & Elliott, 1985). Lichstein and colleagues (1994) included the same four elements in their model of implementation, but added two facets, treatment receipt and treatment enactment, to their model. Treatment receipt is ensuring that treatment has been received, which includes data that confirms that the participant understands and demonstrates knowledge of treatment skills, and demonstrates the ability to use treatment skills in any setting. Treatment enactment is defined as the assessment and confirmation that the participant applies the skills learned in the intervention to their daily life. Witt and Elliott's (1985) model involves determining the intervention acceptability, use, integrity, receipt, enactment, and effectiveness, all reciprocally involved with one another. One advantage of this model is the comprehensiveness with which it considers the outcomes of treatment; however, it fails to describe specific methodology necessary to measure the multiple domains of the model, and what types of measurement might be desirable in order to assess each aspect of the model.

Carroll and her colleagues (2000) specified another model for evaluating what they called adherence and competence in a specific psychotherapy protocol. They called their model the Technology Model of implementation, which was used in medical research by Santacroce et al. (2004). The Technology Model specifies that in order to effectively examine implementation, one must first "rigorously evaluate" behavioral treatments through the specification of dose, or number of contacts, and the active and

inactive “ingredients” of the intervention. This model includes identifying the common and unique elements of the intervention being studied. One must specify the conditions under which the intervention should be administered, and assess whether the treatment was adequately delivered to all participants. These objectives may be achieved through four requirements, which include: (1) specification of treatment in a manual, (2) training interventionists in the uniform delivery of the treatment and supervision to ensure uniform training occurs, (3) monitoring treatment delivery, and (4) including intervention fidelity as an independent variable in outcome analysis (Carroll et al., 2000).

The creation of a manual allows for explication and standardization of the intervention’s elements, including the theory, goals and strategies used in the intervention. It provides an objective means for the comparison of interventions and replication of the study, and allows for competent transfer of the intervention from research to practice. Manuals also help minimize variability in outcomes for an intervention that might result from implementation effects (Santacroce et al., 2004).

Training and supervision of interventionists is an important component of this model, as it allows the opportunity for program developers to mold, refine, and expand the skills of interventionists. Monitoring the implementation of interventions tends to be expensive and time consuming; however, it is essential in the final component of the technology model of implementation – relating implementation to intervention outcomes.

Findings from the study by Carroll and her colleagues (2000) found that therapists tended to over-report their own level of adherence relative to that reported by independent raters. Thus, independent measurement of implementation becomes necessary, suggesting that a conservative approach to verifying implementation of an

intervention might be best accomplished through the use of independent observations conducted by raters who are educated in the elements of the intervention. Self-ratings of implementation of a therapeutic protocol have not demonstrated adequate reliability; however, self-rating is often used successfully as a tool in training and supervision (Santacroce et al., 2004). The focus of the study by Carroll and her colleagues (2000) was the development of a specific observation rating system for a variety of psychotherapy studies, with a goal of assessing therapist implementation of a protocol and the frequency of this behavior, as well as the quality of therapy provided.

In a multi-site treatment fidelity workgroup, as a part of the National Institutes of Health Behavior Change Consortium, many of the strategies mentioned above were combined into a set of “best practices” in the field of treatment fidelity research (Bellg et al., 2004). One of the practices involved ensuring that the treatment dose is the same within conditions. In the field of educational interventions, this treatment dose might include similar numbers of contacts with sites or participants, equal durations of treatment for all conditions of the study and equal informational content in each intervention tested. These contacts should be monitored by independent observers occasionally, and feedback should be provided based on these observations. This group also recommends that interventionists must plan for implementation setbacks, which they specify might include procedures such as training spare intervention providers and tracking treatment provider attrition. Ensuring that treatment providers have been properly trained to provide an intervention is essential in implementation, and this skill requires monitoring interventionist skills and knowledge both during training and while providing the intervention. Standardized training procedures are necessary, including use

of the same trainers to conduct training, the measurement of skill acquisition, and planned procedures to prevent intervention drift over time. Recommended procedures during standardized training include the use of standardized training materials, role-playing, the observation of intervention procedures, and evaluation of implementation of a protocol within the training itself. Finally, training materials to ensure practice with materials and similar delivery of training across sessions is useful (Bellg et al., 2004).

The study of implementation practices has developed considerably over the past 30 years, beginning in the late 1970s with little mention of the idea of implementation, to definition and examination of implementation in the early 1990s, and, more recently, recommendation of best practices. Over this course of implementation research, there has been little investigation into patterns of change in implementation.

One development in this field of study is a recent backlash against treatment fidelity as it was originally conceived. An assumption of fidelity research has traditionally held the idea that greater fidelity to procedures outlined in an intervention protocol is ideal. An alternative to this view is that intervention needs to be individualized in order to maximize positive outcomes, and that interventions must be responsive in order to be most effective (Bierman et al., 2002). For instance, Leventhal and Friedman (2004) suggest that strict adherence to protocol may at times be counterproductive, given that there are natural variations among intervention providers and their clients, and unique interactions between the two. They suggest that rigid implementation of therapeutic protocols tends to limit generalizability, and that rigid implementation also ignores the contextual differences such as features of the setting of the intervention and features of the recipient of the treatment. Thus, the alternative put forth by Leventhal and Friedman

(2004) is that research studies should determine the active component of an intervention, and adapt it closely to the theoretically relevant construct that was used in the development of that component. The measures that researchers use to assess fidelity to treatment should also be closely tuned to that active component of an intervention. In this view of fidelity, measurement and study of fidelity to treatment is still important, but individual adaptation of the treatment is more acceptable, and fidelity can be defined in more broad, goal-oriented terms.

Implementation in Early Intervention Research

Despite the developing literature on implementation that is found in the fields of medicine, prevention and psychotherapy, not as much published research occurs in the fields of education or early intervention. The focus of this dissertation is to examine the implementation of the Partners for Literacy curriculum, developed for and used in the national experimental study of the Even Start program. Currently, there is limited research addressing implementation of early childhood curriculums, and there is little information regarding factors that might be related to implementation.

Partners for Literacy is a research based curriculum. Previous early intervention programs have created a strong rationale for the Partners for Literacy curriculum. The Carolina Abecedarian Project, Project CARE, the Infant Health and Development Project, and the Perry Preschool Project all suggest that efforts to impact children's learning early in life benefits at-risk children in both the short and long term. Few measures of treatment fidelity were included in these studies, with the exception of the Infant Health and Development Program. Despite large-scale, longitudinal study of the effects of early intervention and high-quality early education on child outcomes (i.e.,

NICHD Early Child Care Research Network, 2006; Halpern, 2001; Layzer, & St.Pierre, 1997), it is still unclear to what extent early intervention programs were implemented as planned and through what processes or mechanisms positive change was achieved.

Abecedarian Project

One study of the effects of intensive early intervention is the Carolina Abecedarian Project. This experimental study involved random assignment of 111 children in two phases who were screened to be at-risk for school failure through a combination of demographic factors (Campbell & Ramey, 1994). The two phases of this set of studies included an intensive preschool intervention or control group assignment, as well as the random assignment of half of each of these groups to a school-age intervention program or control group assignment. The preschool intervention group received high-quality, intensive preschool care for 5 years, using a curriculum aimed at enhancing cognition, language, social development, and perceptual-motor skills (the LearningGames Curriculum). The follow up data collected on children 4-7 years after intervention suggested that preschool intervention produced positive changes in intellectual development and academic achievement (Campbell & Ramey, 1994). In the Campbell and Ramey study (1994), few indicators for participation or program integrity were explicitly reviewed. In this single-site study conducted in close proximity to where the researchers based their work, adherence to intervention protocol was monitored by the curriculum developers. However, data regarding child participation in the program is largely absent from publications on this study.

Parkinson (1975), in her study of the Abecedarian project, did include what she called “implementation data”, defined as the successful or unsuccessful completion of a

curriculum activity as observed by independent raters, and the latency and duration of this activity. In this study, implementation was defined as the completion of a task; the term implementation did not include the actual frequency of use of these activities over a day in the classroom. Parkinson (1975) did, however, investigate early enrollment in the study, defined as the time of enrollment in the intervention, finding it to be positively related to performance on the Mental Development Index of the Bayley Scales of Infant Development. Attendance was also included in models of motor and cognitive performance, and a positive correlation was found for the relationship between attendance and success with motor items, and between age of enrollment and success with motor items. Attendance for children included in this study ranged from 104 to 222 days in one year, with a mean of 180.6 days of attendance. The amount of treatment received, measured as enrollment age and days of attendance per year, was positively related to cognitive outcomes. However, data were not included in later studies related to the Abecedarian project.

A follow-up study suggested that cognitive effects as well as positive effects on reading and math achievement were maintained into adulthood, but that both the experimental and control groups had shown some decline in cognitive and achievement scores over time (Campbell, Pungello, Miller-Johnson, Burchinal, & Ramey, 2001). Changes in reading and math achievement were mediated by intellectual effects of early intervention. In addition to the effects of this early intervention program on cognitive abilities and academic achievement, children in the experimental group showed less risk of school drop-out, as well as an increased tendency to attend a four year college

(Campbell et al., 2001). This follow up study included little data concerning program participation, implementation, or treatment integrity.

A cost-benefit analysis of the Abecedarian Project suggested that there were additional, indirect effects of the program. Masse and Barnett's (1994) study of the benefits of this program included a lower rate of smoking for those young adults assigned to the experimental group, which could be an indirect effect of cognitive and achievement gains. Additionally, mothers of children in the experimental group tended to have higher-paying jobs and attain a higher level of education when their children were 5 years of age than mothers of children in the control group. Masse and Barnett's study (1994) also found that children in the experimental group tended to make use of the Aid to Families with Dependent children less often than children assigned to the control group. In summary, positive effects on cognition, reading, and mathematics, as well as a variety of indirect effects, were longitudinal outcomes of the Abecedarian Project, and these changes were maintained through adulthood. Additional support for the validity of these findings would be possible if intervention fidelity data and participation data were included in outcome analysis.

Project CARE

Project CARE also examined the impact of early intervention on low-income, low-literacy families in a randomized, experimental study. This study expanded the Abecedarian Project, in a three-group design. The three groups were as follows: (1) an educational day-care program and family education group, (2) a family education alone group, and (3) experimental control group. A total of 65 families were enrolled in the study based on risk status according to a survey, then randomly assigned to one of the

three groups (Wasik, Ramey, Bryant, & Sparling, 1990). Like the Abecedarian Project, the main curriculum used in the child development center was the *LearningGames* curriculum. The family education component was delivered through home visits and monthly parent meetings, using a supportive approach to family education, thorough modeling and information, as well as the implementation of a program to help parents learn problem solving and coping skills. Findings from this study indicated that the group of participants in the educational day-care plus family education condition of the study showed a significantly higher IQ score at 12, 18, and 24 months of age, but the family education component of the study alone did not produce similar changes. As well, family education did not appear to change the home environment or parent or child behaviors, as measured by the HOME inventory and parent interview (Wasik et al., 1990).

A study which included data from 161 African-American children, using the data from both the Abecedarian Project as well as Project CARE, also looked at predictors in the cognitive changes found in these two studies. This analysis found that children in the early childhood education groups of the interventions were rated as more task oriented in infancy, with higher IQ scores at 10 measurement points between 12 and 90 months of age (Burchinal, Campbell, Bryant, Wasik, & Ramey, 1997). In addition, children with higher IQ scores tended to show greater gains in IQ over time. Maternal marriage status and education were both significant predictors of child IQ scores. The score on a measure of the home environment was positively related to cognitive scores. As well, the effect of the childcare intervention was mediated in part by the intervention's effect on the child's responsiveness to their environment (Burchinal et al., 1997). Despite positive

findings on child outcomes from Project CARE, there was no reported data on participation by families or factors that might influence the amount of treatment that families received, as well as the degree to which the intervention was delivered as intended by curriculum developers.

High/Scope Perry Preschool Program

The High/Scope Perry Preschool Program is another study supporting the effects of high quality, intensive early childhood education on children born to families with low socio-economic status. Schweinhart (2003) reviewed the study in a cost-benefit analysis of the program. The Perry Preschool Program included 123 participants who were exclusively African-American and low-income. This program provided for the intensive preschool education of those children assigned to an experimental group, which included two years of instruction at 3 and 4 years of age as well as lengthy home visits from the child's preschool teacher. Similar to the data gained in the Abecedarian project, attrition for this sample was extremely low, at 5% (Schweinhart, 2003) and longitudinal follow-up suggested positive outcomes in educational performance and economic benefits as well as a high return on taxpayer investment. Of interest, an additional outcome was the prevention of crime after participation in the experimental group. However, as noted in the Abecedarian project and in Project CARE, little data on child and parent attendance or participation are available, and the degree to which the curriculum was implemented as intended. Documentation of participation, adherence to intervention or curriculum protocol, and interventionist implementation of the programs would increase the external validity of all three of these early intervention studies.

Infant Health and Development Program

One study of early intervention which included research concerning implementation as well as rates of participation is the Infant Health and Development Program (IHDP, Ramey et al., 1992). This program was focused on improving the cognitive, health and behavioral outcomes of 377 low birth weight infants at eight sites, through a three year intervention. IHDP included home visits, parent support groups, a center-based early childhood program, and pediatric follow-up (Ramey et al., 1992). Participation in the intervention group of the program was positively related to IQ scores and fewer behavior problems. Positive gains in math, receptive language, and risk behaviors were maintained through 18 years of age (McCormick et al., 2006). Ramey et al. (1992) describe a “Family Participation Index” that was calculated for use in data analyses. The Family Participation Index combines the number of home visits, attendance at parent support group meetings, and days of child attendance at the center based program into a single participation score, which related positively to child IQ scores at age 3. This relationship was found to be significant even after controlling for a variety of demographic variables. As well, participation data was similar at all eight sites, indicating consistent family involvement with the intervention.

A later study by Sparling and Lewis (1993) examined a variety of factors that influenced the degree to which the intervention was delivered to families as intended. Data were collected through a twice-yearly survey to teachers and home visitors, asking staff to list up to four factors that had positively or negatively impacted the delivery of either the parent or child curricula as intended by the developers. One of the factors most often mentioned was “parents as individuals,” comments related to parental personal problems, communication with parents, or parental confidence. Comments in this

category were positive slightly more often than they were negative. The parental response to the program was a factor mentioned generally in a positive context, including comments regarding a parent's opinion of the curriculum and parental involvement. The child as an individual was often mentioned as a positive factor influencing participation. This category included the comments about child achievement and progress, as well as child enjoyment of home visits. Management of home visits was initially often listed as a negative factor, though comments shifted in a positive direction later in the study. Comments in this category included scheduling problems and communication between the parent and home visitor. One of the categories of comments most often found to be positive was the management of the child development center, including parent communication with the center and attendance at the center. Comments about parent-child relationships were positive about three-quarters of the time, including comments about the parent's desire for a good life for the child. Comments about the family living situation were most often negative, while comments about other family member's influence on participation were about half positive and half negative.

Overall, a regression of initial status (control versus intervention group), participation in the program, and curriculum variables accounted for about 49% of variance in child outcomes. Total negative implementation factors added about 1% to this model, while adding total positive implementation factors into this model did not add significantly to explanation of variance (Sparling & Lewis, 1993). Thus, some factors influencing program implementation were examined in IHDP, with significant influences on participation including factors unique to the parent and child, the parent-child relationship, and the management of home visits and the child development center.

Consideration of the negative factors influencing implementation added significantly to a regression model. Measurement of the ‘dosage’ of intervention provided to families was accomplished through the use of a Family Participation Index, which was significantly related to child outcome.

Sparling and Lewis (1991) also studied the relationship between one measure of implementation and child cognitive outcome in IHDP. The Early Partners curriculum was designated as the first part of the intervention, delivered by a home visitor in the first months after the child left the hospital. Early Partners was aimed at supporting parent-child interactions in the areas of reading cues from the baby, understanding sleep/wake states, calming the child, awareness of levels of stimulation, interaction and communication with the baby, knowledge of muscle tone and eye-hand coordination, and developing independent handling and manipulation. The second part of the intervention, the Partners for Learning Curriculum, was delivered by a home visitor as well as in a child development center, from 12 to 36 months of age. The activities in this curriculum involved game-like learning, using two week cycles for each game. The curriculum also included a component to develop adult skills in supporting the child’s learning, which were hierarchical in nature and progressively more complex. In order to measure implementation of this complex intervention, four variables were considered: the number of activities introduced in the child development center, the number of activities introduced during home visits, the activity episodes per day at the child development center, and the number of activities per home visit. Regression was carried out using the last two variables listed: the activities per day at the child development center and at home visits. While the initial demographic variables and overall participation index

accounted for a large proportion of the variability in 36 month-old Stanford-Binet IQ scores (39%), the number of activity episodes per day at the child development center accounted for an additional 2% of the variance, and the activities per home visit accounted for an additional 6% of variance, both significant at the $p = 0.01$ level.

Liaw, Meisels and Brooks-Gunn (1995) built on Sparling and Lewis' work (1991) by considering the data on implementation in an alternative fashion. These authors used three different variables to measure implementation and relate it to cognitive (36 month-old scores on the Stanford-Binet IQ test) outcomes. These variables were exposure (the number of contacts in the home and at the child development center), rate (the number of activities presented per visit, or per child development center day), and active experience (a combination of the parent's interest in the activities in the home, as well as the child's mastery of intervention tasks taught at the center). Thus, the study by Liaw, Meisels and Brooks Gunn (1995) added to the Sparling and Lewis (1991) study by including three measures of the degree to which participants actually experienced the intervention, and their participation in the intervention – a concept that is rarely considered in implementation research. By including these three variables, the study considered the degree to which participants actually received the intervention, which might differ from the degree to which teachers and parent educators attempted to deliver the intervention. In the regression analysis by Liaw and colleagues (1995), though about 53% of variance was accounted for by demographic and background variables known to relate to child IQ, active experience with the curriculum accounted for about 10% of the variance in IQ scores. Three percent of variance was accounted for by the exposure to the curriculum that the child experienced, and an additional 3% of variance was explained by the rate at

which the child was exposed to the curriculum. Thus, the degree to which parents and children actively participated in the curriculum added significantly to a model accounting for variability in IQ scores. In both the Liaw and colleagues (1995) study and the study by Sparling and Lewis (1991), implementation was considered in its relation to child outcomes, revealing specific intervention variables were significantly related to the degree to which children benefited from this intervention.

The IHDP research study contained many of the aspects that the Technology Model for intervention (Carroll et al., 2000) specifies in the documentation of treatment fidelity. First, a manual specifying treatment was included as a key component in this multi-site study. The manual was used as a part of training, which was uniform across all interventionists and aimed at training interventionists to deliver the program in a uniform fashion at all sites (Ramey et al., 1992). Supervision of interventionists was completed by program developers and was aimed at ongoing training and feedback (McCormick, 2006). Monitoring of treatment delivery was achieved through several means (B. H. Wasik, personal communication, April 2, 2006). Five senior staff conducted site visits two to three times per year to monitor implementation and provide constructive feedback to all staff based on observations of classrooms, home visits, and staff meetings. Implementation data from sites was reported weekly for early childhood classrooms and after each home visit. This implementation data was monitored weekly by IHDP research staff, with follow-up to sites as needed. At each site, an educational director was taught a problem solving model to be used with difficulties in implementation. The educational director met with site staff weekly and with IHDP implementation staff 3-4 times per year. Thus, several methods were used for monitoring ongoing

implementation. IHDP included the components of treatment fidelity as outlined in the Technology model, including the use of a treatment manual, uniform training of interventionists and supervision of training, monitoring of treatment delivery, and including implementation in outcome analyses.

Early Head Start

The Early Head Start program has also investigated aspects of treatment fidelity in early childhood intervention. The Early Head Start program is focused on the provision of individualized child development, parent education, and family services using a plan of care and referral to community services. The purpose of the National Early Head Start Research and Evaluation project is to assess the quality of child care services provided through this program and measure implementation for the research sites included in the project. Paulsell, Kisker, Love, and Raikes (2002) examined implementation processes in the program over a four year span of program funding. They examined implementation through the development of detailed rating scales and a process to rate the implementation of programs; ratings were carried out by trained members of the national evaluation site visit team. Additional data were collected through staff, parent, and community interviews, case file review, staff surveys, and observations of home visits and center-based intervention. Implementation was measured on a five point scale for all components, from minimal implementation (rating of 1) to enhanced implementation (rating of 5).

Findings from this study indicated that 6 of 17 programs were considered fully implemented during the first year of funding, while 12 of these programs had achieved full implementation by the fourth year of implementation. Factors that were associated

with incomplete implementation included an early leadership change at the program, high rates of staff turnover, difficulty responding to feedback from Head Start Bureau monitors, and difficulties with community partnerships (Paulsell et al., 2002). These data led to strategies that were effective in improving implementation ratings for programs included changing curricula and assessment tools for children to increase focus on child development, creating child care centers, efforts to increase the quality of existing child care centers, creation of systems to track services, the formation and strengthening of partnerships with community agencies, creating new staff positions, and improving staff training and education (Paulsell et al., 2002).

In addition to early implementation findings, the same research group developed a Manualized Assessment of Progress system (MAP), with the goal of monitoring “the efficacy and fidelity of program implementation” (Dickstein, Seifer, Eguia, Kuersten-Hogan, & Magee, 2002, p. 233). This system involves computer tracking with weekly ratings by teachers and Early Head Start staff. A preliminary study of progress towards Early Head Start goals suggested that 100% of children tracked received developmental readiness goals, but only 67% had goals related to health, nutrition, and mental health. Parent goals were more varied, but about 57% of parents chose self-sufficiency goals, and 77% chose goals related to the promotion of child development. Reflection on the MAP system suggested it enhanced cooperation among Early Head Start staff, challenged staff to think flexibly in achieving goals, promoted family-school partnerships, and required staff to reflect regularly on progress towards individualized child and family goals (Dickstein et al., 2002). These combined studies suggest that there are numerous

methods for assessing implementation of programs, with some seen as positive in promoting advancement of the programs themselves.

Love and colleagues (2005), in a more recent consideration of implementation and its relation to outcomes in the Early Head Start program, found a relationship between site implementation of the program and child cognitive and language outcomes. This study examined interview, standardized assessment, and observational data on 3001 families randomly assigned to participation in the Early Head Start program. Love and colleagues (2005) found that, overall, Early Head Start produced positive outcomes on children's cognitive and language development, as well as decreases in aggressive behavior and increases in sustained attention. They separated programs into three sets – programs that implemented Early Head Start at a high level soon after receiving funding, programs that implemented the program at a high level later, and programs that did not implement the intervention at a high level throughout the course of the study. All three groups showed positive outcomes in the areas listed above, but programs that implemented the program at a high level, either early or late in the study, produced greater improvements in child outcomes than did programs that did not implement the program at a high level. The results of this study provide further evidence that the Early Head Start program, when implemented in accordance with federal guidelines, is responsible for positive impacts on child outcomes.

Early Head Start, through the national research and evaluation program, has incorporated many of the components of the Technology Model of implementation in their studies of program effects. Early Head Start has a manual, which takes the form of a start-up planning technical assistance report from the Early Head Start National

Resource Center (1999). This report provides step-by-step assistance in starting an Early Head Start program, including how to set up the program itself, and monitoring strategies for evaluation and improvement. Interventionists do not receive uniform training under this program, but do have access to a uniform set of materials that are easy to access on web pages. Supervision of interventionists is generally carried out locally, usually by interventionists themselves, which is inconsistent with the objectives of the Technology Model, calling for supervision of interventionists by independent observers (Carroll et al., 2000). However, *monitoring* of treatment delivery is thorough and carried out objectively through the MAP system (Dickstein et al., 2002). Implementation has been incorporated into studies of outcomes over time, through the Manualized Assessment of Progress system and the study by Love and colleagues (2005). The study by Dickstein and colleagues (2002) considered progress towards individualized goals, as assessed in the MAP system, as positive outcomes of the Early Head Start program. Thus, many aspects of the Technology Model of treatment fidelity are present within the Early Head Start program.

Family Literacy

Family literacy programs build on the assumption of other early literacy programs - that children's learning can be meaningfully and significantly impacted by intensive, high quality education in the first years of life. In addition, these programs include an adult literacy component, to help parents improve their literacy skills and advance their education and employment. Also known as intergenerational literacy programs, family literacy programs incorporate a wide range of programs that include both adult and child literacy services (see Wasik, Dobbins, & Herrmann, 2001; Wasik & Herrmann, 2004).

The federal Even Start Family Literacy Program began in 1989 as a federally-funded intergenerational literacy program. The “Even Start program has three related goals: to help parents improve their literacy or basic educational skills; to help parents become full partners in educating their children; and to assist children in reaching their full potential as learners” (St.Pierre et al., 2000, p. 11). To participate in Even Start, parents must be eligible for adult basic education, with low levels of education and/or low levels of English proficiency. Families must be low-income and must have a child between the ages of birth and 8 years. Several components are required of any Even Start program, which may be provided through federal funds or through collaborating agencies and external funding. These include an Early Childhood Education component, an Adult Education component, a Parenting Education component, and time for Interactive Literacy Activities. Several evaluations of the Even Start program have been conducted in order to determine the status of programs that receive Even Start funding and to generate recommendations for improving outcomes of children who participate in these programs (St.Pierre, Swartz, Murray, and Deck, 1996; Tao, Gamse, & Tarr, 1998; St.Pierre et al., 2003; Ricciuti et al., 2004).

St.Pierre and colleagues (2000) conducted a study of local independent Even Start evaluations conducted as a requirement for funding of Even Start programs. They obtained data on four state evaluations and 118 local evaluations. More than 75% of these evaluations focused on project outcomes and were used by local projects to obtain financial or political support. Notably, the systematic use of data was a rare observation in these evaluations, and program evaluations tended to rely on anecdotal information. These investigators found very little data on the quality of implementation of the Even

Start program. Of those studies that chose to address implementation, 62% reported participation information, which was often defined as the average number of months that families participated in local Even Start programs, using self report as the measurement approach. In summary, findings from this review of federal Even Start programs indicated that programs often used evaluation as a forum for advocacy, but that evaluations often lacked conclusions supported by data and there was little information about program implementation in evaluations. Despite positive outcomes that were reported in many evaluations, there is little data to support conclusions that these outcomes were the result of local Even Start program efforts, as little data was collected to measure positive outcomes, and programs did not compare child outcomes with a suitable control group. The report also indicated that there was little information about the quality of implementation in Even Start programs.

Three national evaluations of the Even Start program have been conducted since the program was established. The evaluation by St.Pierre and colleagues (1996) found several barriers to implementation of the Even Start program. These barriers included difficulties with participant recruitment, poor participant retention, poor attendance, low motivation of families, problems of communication and coordination with cooperating agencies, lack of transport for participants, unexpectedly severe social services needs, a lack of quality child care, financial problems, staffing problems, and problems with facilities and space. It also found that while Even Start programs were funded for 4 years at a time, few families participated for more than two years in the program.

Of the families included in the first national evaluation study, those that enrolled in the Even Start program were more likely to participate in adult education, parenting

education, and early childhood education than other low-income parents who did not enroll in the Even Start program. Significant variability was found among projects in all facets of Even Start, including planning and implementation, the characteristics of families served, the amount and duration of program participation, and the outcomes for participating children and adults. In general, cognitive outcomes for children were initially positive, though gains tended to diminish once children were enrolled in school.

The Second National Even Start Evaluation (Tao et al., 1998) aimed to address the populations served by the program, spending and implementation, and developmental gains of participants. Findings related to participation indicated that the majority of families were significantly below federal poverty levels, and the majority of parents had not obtained a high school diploma or a GED before enrollment in Even Start. A significant proportion of participants had limited English proficiency. The second evaluation found an increase from the first evaluation in the proportion of teen parents served by the program, and the proportion of Hispanic/Latino families served by the program. Funding had increased on a per-family basis from the first evaluation to the second. Implementation was addressed by considering participant involvement as a key area for improvement, and participant retention was a similar challenge for programs. Gains were found over time for participants on literacy measures. This study, however, was not conducted as a randomized experiment, thus the gains observed could not be attributed to the Even Start program itself. Overall results of this evaluation were consistent with the first national evaluation, but did show some important shifts in participant demographics and highlighted the need for a randomized, experimental study to examine outcomes of participants in the Even Start program.

The Third National Even Start Evaluation (St.Pierre et al., 2003) collected descriptive data about participants using Even Start, participation data, and outcome data related to child and parent literacy and cognition. The population served by the Even Start program was determined to have substantially lower household incomes than those participating in the Head Start program, and parents participating in this program were much less likely than parents in Head Start programs to have earned a high school diploma. Participation data revealed that families tend to participate in only a small amount of services offered by the Even Start program, but the extent to which parents and children participated in this program was positively related to benefits in child outcomes. Some of the positive outcomes included gains on literacy measures. When participation rates were not considered in analysis, children and parents in the Even Start program made gains that were equal to those found in the control group; however, about one-third of children in the control group participated in other early childhood education programs. One of the major implications of the results cited in this study was that “families did not participate long enough and did not get enough instruction to make the kinds of changes that are needed” (St.Pierre et al., 2003). Participation in the program needed to increase in order to produce meaningful changes in outcomes for enrolled parents and children.

Follow-up findings, collected one year after post-test on the third evaluation of the Even Start program, indicated similar results to those found in the third evaluation (Ricciuti et al., 2004). Namely, some child literacy gains were found, but these did not exceed gains made by children in the control group. The opportunity for an additional year of participation in the program did not result in positive gains for children in the Even Start program when compared to the control group. Similar to the third national

evaluation of the even start program, participation rates in Even Start were low, while about one-third of families in the control group again participated in other types of early childhood education (Ricciuti et al., 2004).

The findings from the third evaluation and follow-up to this evaluation were important in describing the state of the Even Start program at that time, and to highlight areas for potential improvement. One area for improvement noted in both these studies was a lack of language acquisition and reasoning instructional content in classrooms, as well as concern with the quality of classroom instructional practices (St.Pierre et al., 2003). As well, in the third evaluation and follow up study, no significant child literacy outcomes were found when compared with the control group. This disappointing finding could be due to a lack of distinctiveness between the experimental and control groups, in terms of the early childhood education services received, given that one-third of control group members participated in other early childhood education programs. Areas for improvement as found in the third National Even Start study and follow-up were considered when further evaluation of the Even Start program began in 2003, through the Classroom Literacy Outcomes and Interventions study (CLIO).

Classroom Literacy Interventions and Outcomes Study (CLIO)

The Classroom Literacy Interventions and Outcomes (CLIO) study is a research study commissioned by the United States Department of Education to compare the current state of the Even Start program with two experimental sets of curricula for early childhood education and parenting education. This study is being carried out by Westat. Criteria for selection of experimental curricula in this study included evidence of effectiveness of the curriculum, applicability of the curriculum to the population served

by Even Start, a focus on early literacy, the development of integrated parenting programs, and cost. Thus, both interventions had a previous, positive research basis before selection for this study (Westat & Abt Associates, 2004).

Both sets of curricula were designed to be tested in the early childhood classroom and in the parenting component, or in the early childhood classroom alone. The CLIO study also contains a control group of sites that received neither curriculum for the duration of the study. Thus, there are five groups of sites participating in the study: two groups of sites where the early childhood curriculum from each program is being implemented, two groups of sites where the early childhood curriculum as well as the parenting education curriculum are implemented, and one control group (see Figure 1). The original recruitment goal was for each group to include 30 sites, for a total of 150 sites participating in the CLIO study.

One curriculum chosen for use in the CLIO study was developed at the University of North Carolina's (UNC) Frank Porter Graham Child Development Institute (FPG). This program, titled "Partners for Literacy," (PfL) builds on three earlier studies: the Abecedarian project, Project CARE, and IHDP, as well as a descriptive study of family literacy programs (Wasik, Herrmann, Dobbins, & Roberts, 2000; Harbin, Herrmann, Wasik, Dobbins, & Lam, 2004). The other curriculum, from the CIRCLE group at the University of Texas – Houston Health Sciences Center, includes the "Let's Begin with the Letter People" early childhood curriculum and "Play and Learn Strategies" for parenting education. Both these projects had a year to complete extended curriculum development and pilot testing, during which Westat collected baseline data at all participating sites. In the spring of 2004, sites were randomly assigned to an intervention

or control condition, and trainings were conducted for all intervention programs at three different locations across the country. Implementation of the curricula began in the fall of 2004, and continued through the spring of 2006. A second set of trainings were conducted in the summer of 2005 for all participating staff (Westat & Abt Associates, 2004).

The sites participating in the CLIO study were recruited based on several inclusion criteria. First, the chosen Even Start sites had to include a minimum of 5 three- and four-year-old children in one classroom, or 8 of the same aged children in two classrooms at the site. Sites were required to provide at least 12 hours of center-based early childhood instruction per week, and serve a majority of families that spoke English or Spanish. Finally, sites had to be willing to participate in the study and be willing to be assigned to one of the five intervention conditions (Westat & Abt Associates, 2004). Based on the above recruitment criteria, 120 sites agreed to participate. Sites come from 33 states, in every region of the United States (Westat & Abt Associates, 2004).

Partners for Literacy

Partners for Literacy is one of the two experimental curricula chosen for evaluation in the CLIO study. This curriculum, designed by Barbara H. Wasik and Joseph Sparling, is based on the idea that children learn best through intensive, individualized instruction, especially instruction that is presented in a fun, game-like format. Several strategies are included as part of the Partners for Literacy curriculum. These include:

- The 3N strategy,
- Interactive Book Reading, and
- Extended Teaching and Enriched Caregiving.

In addition, several curriculum materials were included in the Partners for Literacy curriculum. These include:

- LiteracyGames,
- LearningGames,
- Manuals for the use of Interactive Book Reading, LiteracyGames, Enriched Caregiving, Problem Solving, Little Conversation Books, and Literacy Rich Classrooms
- Little Conversation Books, and
- Curriculum embedded assessment, through the Record of Mastery and Interactive Book Reading checklist.

The above curriculum strategies and curriculum materials will be described in further depth.

The 3N strategy stands for three steps – Notice, Nudge, and Narrate. These three steps are used repeatedly and recursively throughout different activities, to scaffold a child’s learning. The caregiver first “notices” what the child is doing, by watching the child carefully and using some phrase to tell the child that they notice what the child is doing. The caregiver then “nudges” the child – scaffolding the child to move one step beyond what they could do without assistance. Finally, the caregiver “narrates” the child’s actions, providing a verbal description of what the child has done (Wasik & Sparling, 2005). There are several intents for the 3N strategy. First, by using oral language throughout the process, it encourages the child’s oral language and vocabulary development. Second, by closely attending to the child and individualizing instruction, it provides focused, intensive instruction to the child (Wasik & Sparling, 2005). The 3N

strategy builds on a child's knowledge and skills through individualized, high intensity direct instruction.

Interactive book reading is a set of strategies aimed at engaging the child in reading activities, providing context for reading and building on emergent literacy skills in the preschool classroom. Interactive Book Reading includes the 3S strategy, Expanded Book Reading and the use of Wh Questions (Wasik & Sparling, 2005). Interactive Book Reading is intended to be used with all books within the classroom, including those provided by the Partners for Literacy Project. Teachers were instructed to use Interactive Book Reading strategies with children in groups of one to two children, at least once per day.

The 3S strategy, part of the interactive book reading strategies, uses individualized, intensive instruction in reading to help build on emergent literacy skills. In this strategy, the caregiver begins by using "see", where the caregiver points to elements in the book and encourages the child to direct their attention to a part of the book. Next, the caregiver asks the child to "show" them something, such as pointing to a part of a picture. The third part of this strategy is "say", where the caregiver encourages the child to give a verbal response to some prompt (Wasik & Sparling, 2005). The three steps in this process are recursive much like those in the 3N strategy. The caregiver might start off using the "see" part of the strategy, and might move to "show" soon thereafter. The caregiver can move from level to level depending on the child's skill level. The caregiver can use this strategy to individualize the difficulty of the activity to the child's skill level.

The Wh Questions, another component of the interactive book reading strategies, refer to asking children direct questions about the book. Who, What, Where, When, and

Why are used to ask questions about the book in a variety of levels of difficulty. For example, the “who” and “where” questions tend to be easiest, while the “when” and “what” tend to be slightly more difficult. “Why” tends to be the most difficult type of question, building on inference (Wasik & Sparling, 2005). Wh Questions are used during book reading to promote verbal interaction between the caregiver and child.

Expanded Book Reading refers to a set of activities that expand the reading experience beyond just reading the words written on the page of a book. Expanded Book Reading includes the use of activities to connect the child with the book in a meaningful and interactive way. Introduction refers to the teacher describing the book before it is read, including a discussion of concepts of print. Prediction involves discussing what might happen in the story with the child before it is read. Identification involves a child making personal connections with the story. Retelling asks the child to describe the beginning, middle, and end of the story after it is read. Discussion involves talking about the story after it has been read. The expanded book reading strategies, Wh Questions, and 3S strategies are all components of interactive book reading strategies, a major set of curriculum strategies in the Partners for Literacy curriculum.

Another curriculum strategy in the Partners for Literacy Curriculum is Extended Teaching (for teachers) and Enriched Caregiving (for parents). This strategy calls for teaching throughout the day. Teachers can foster the development of language, vocabulary, and cognitive skills through expanded conversation throughout the day, including snack time, transitions, and transportation (Wasik & Sparling, 2005). Parents can foster the development of vocabulary and language through expanded conversation with their children during meal times, bath time, and bed time.

The final Partners for Literacy strategy is Problem Solving. Problem solving refers to a set of steps designed to help teachers, children, and parents work methodically through a problem. For parents and teachers, these steps include identifying the problem, identifying the goal, generating solutions, evaluating consequences of solutions, deciding on a solution, executing the solution, and evaluating the solution. These steps are simplified for children's use in the classroom.

One of the major curriculum materials for Partners for Literacy is LiteracyGames. LiteracyGames is a set of 100 activities, 50 for three-year-old children and 50 for four-year-old children. Each LiteracyGame includes 3 cycles of increasing difficulty (Wasik & Sparling, 2005). LiteracyGames work on a variety of early learning skills, including oral language, vocabulary, concepts of print, phonological awareness, and alphabet knowledge, with a particular focus on early literacy skills. Sites with a full day of early childhood instruction time were instructed to use LiteracyGames twice daily with each child in the classroom, in groups of one or two children with the teacher or teacher's aide. Half-day programs were instructed to use at least one LiteracyGame per day, with each child. LiteracyGames were designed to last for about 5 minutes each. These games are based on the LearningGames Curriculum, a set of activities aimed at promoting the early development of children ages birth to five years. One aspect of the Parenting Education component of Pfl is LearningGames, a set of 50 games for each year of age, for birth through five years of age. These games have a longstanding research basis in early childhood education settings (i.e., Campbell & Ramey, 1994; Burchinal et al., 1997; Campbell et al., 2001).

Teachers were asked to make their classrooms “literacy rich.” They received training during week-long summer workshops and were provided with materials to facilitate improvement in literacy content. A checklist was completed with program consultants to ensure that components of a literacy rich classroom were included in the classroom. Many of the materials required for this component were provided by the Partners for Literacy project. These materials include classroom labels for centers and materials found in the classroom, labeled in English and Spanish, alphabet letters, and posters of the main strategies for the PfL curriculum. Additionally, teachers received a large number of “conversation stands,” transparent plastic frames in which teachers were asked to place pictures and items to encourage conversation, for placement on tables in the learning centers and on tables used for meals and snacks. Creation of a writing center, including a variety of materials available for use in emergent writing activities, was included as part of the literacy rich classroom in the second year of the intervention.

Another curriculum material is the set of Little Conversation Books. Little Conversation Books are small, four page books printed on a single sheet of paper. These little conversation books build on a variety of early school skills. The Nursery Rhyme Little Conversation books use both well known and less-known nursery rhymes to build on children’s knowledge of rhymes. The ABC stories use repetitive text to help children learn the letters of the alphabet. Sequencing stories include a beginning, middle, and an end in each story to help children learn to re-tell stories. The words-in-words Little Conversation Books use compound words. Problem Solving stories use simple social problems to help children think about problem solving and generate their own solution to problems (Wasik & Sparling, 2005).

The final set of curriculum materials are assessment materials, including the record of mastery, interactive book reading checklist, and feedback forms. In order to fully implement the Partners for Literacy Curriculum, teachers were instructed to complete the Record of Mastery for each child in the intervention, which measured child progress in the LiteracyGames. To ensure daily interactive book reading, teachers were instructed to complete a checklist for interactive book reading. As well, teachers were asked to complete a Teacher Feedback Form for each classroom, every two to four weeks. The director at the site, or the teacher's supervisor or mentor, was asked to complete a Director/Mentor - Teacher Feedback Form for each teacher and teacher's aide within the classroom, every two weeks.

The Partners for Literacy curriculum was supported through intensive training, on-site consultation, and staff support in North Carolina. The project personnel included the principal investigator, Barbara Wasik, co-principal investigator, Joe Sparling, project coordinator, project assistant, five program consultants, and four graduate assistants, as well as numerous contract and FPG Institute staff. The focus of the work of the program consultants was site support and training. Program consultants provided on-site training to all sites, visiting each participating site two to three times per year for observation, intensive training, and support in the implementation of the Partners for Literacy curriculum as described above.

The Technology Model of Implementation, as noted earlier, requires four elements of implementation: (1) the specification of treatment in a manual, (2) training of interventionists in the uniform delivery of the treatment and supervision to ensure

uniform training occurs, (3) monitoring of treatment delivery, and (4) inclusion of intervention fidelity as an independent variable in outcome analysis (Carroll et al., 2000).

The Partners for Literacy curriculum in the CLIO study includes all of the above requirements. The Partners for Literacy curriculum includes an extensive set of manuals that outline specific elements of the intervention (Wasik & Sparling, 2005). Training of interventionists took place at seven large-scale trainings over a period of two years using the same procedures for all interventionists. Though the program developers trained the interventionists during workshops, ongoing monitoring was provided by both program consultants and project directors. Monitoring of treatment fidelity also took place through teacher and director completion of Feedback Forms, and Program Consultant ratings of classroom implementation. Ratings of intervention fidelity will be used by Westat as an independent variable in outcome analysis. Thus, procedures in the Partners for Literacy study fit well with requirements for implementation as outlined in the Technology Model (Carroll et al., 2000).

In addition to the good fit between the Partners for Literacy study procedures and the Technology Model of Implementation, the opportunity exists to examine some of the links between various aspects of this model. A common assumption in implementation research has been that training of interventionists leads to greater implementation of that intervention. However, the link between training and implementation has little research support to this point. The Partners for Literacy project lends itself to an intensive analysis of the factors that influence implementation, given the data that Partners for Literacy collected on various implementation factors. This data was collected from a variety of sources, including (1) training documentation, (2) records of program

consultant site visits, (3) collection of implementation documentation from sites, and (4) observations by Partners for Literacy staff.

Purpose

The purpose of this study was to investigate the factors that contribute to program implementation for the Partners for Literacy Curriculum. Several factors were specifically addressed. The first factors, teacher and director attendance at training, are defined as the number of week-long training sessions that a teacher or director attended over the two year intervention. The next factor, the number of on-site consultations, is defined as the number of times the teacher was visited by a program consultant during the two year intervention. Teacher and director compliance with record keeping is the next set of factors, defined as the number of feedback forms that the director and teacher returned over the two year period. Teacher turnover, another implementation factor, is defined as changes in the lead teacher employed in the classroom over a two year period. The number of children, the final implementation factor, is defined as the number of children enrolled in the CLIO study at the beginning of each intervention year. Implementation ratings were made by program consultants, addressing the quality of Partners for Literacy Curriculum implementation independent of overall classroom quality, which was outlined through meetings with all Partners for Literacy staff.

This investigation focused on the early childhood education component of the intervention. In addition to examining the factors that predict implementation, this study examined changes in implementation over time for a new curriculum in the early intervention setting.

Research Questions

The main foci of this dissertation were (a) to determine the factors that predict quality program implementation in a Partners for Literacy early childhood education classroom and (b) to examine how implementation ratings changed over time in Partners for Literacy early childhood education classrooms. First, correlations among program implementation factors were analyzed.

Purpose:

To examine the correlations among the factors used to define program implementation.

Hypotheses

It was hypothesized that:

- i. Teacher attendance at training would be positively correlated with teacher compliance with record keeping.
- ii. Teacher attendance at training would be negatively correlated with teacher turnover.
- iii. Director attendance at training would be positively correlated with director participation in record keeping.
- iv. Teacher participation in record-keeping would be negatively correlated with teacher turnover.

Next, seven program implementation factors were analyzed for their ability to predict program implementation ratings at the end of two years.

Purpose:

To examine how well seven program implementation factors predicted the fidelity of program implementation in the Partners for Literacy Curriculum measured at the end of two years.

Hypothesis:

It was hypothesized that teacher and director attendance at large-group trainings would be positively correlated with ratings of program implementation conducted by intensively trained program consultants. The number of site visits was hypothesized to relate positively to program implementation ratings. It was hypothesized that teacher and director participation in record keeping would be positively associated with ratings of program implementation. It was hypothesized that the number of children enrolled per classroom would not relate significantly to the degree to which a curriculum is implemented.

To build on the first research question and hypotheses, it was expected that teacher and director attendance at workshop trainings, the number of on-site training sessions, and participation in record keeping on the part of the teacher and director would be the factors which best predicted ratings of program implementation in a classroom. Finally, the patterns of change of implementation ratings was examined.

Purpose:

To examine how ratings of implementation quality changed over a period of two years in classrooms using the Partners for Literacy Curriculum.

Hypothesis:

It was expected that implementation would occur slowly, that by the mid-year point of the first year of implementation, only about 30 percent of sites would have implemented the curriculum to a rating of 3 on a 5 point scale. It was hypothesized that a rating of at least three on a five point scale would rise to about 50 percent of classrooms by the end of the first year, and would rise significantly following a second year of training, to 70 percent.

CHAPTER 2

METHOD

Participants

This study has two groups of participants, the Partners for Literacy program consultants and the program staff at the study sites throughout the country. Consultants provided data regarding ratings of implementation, while site staff involved in implementation provided data regarding other aspects of program implementation.

Program Consultants: The five program consultants were full time staff who provided summer training and ongoing program consultation for all program sites. Each consultant was assigned 8 to 10 of the program sites to provide onsite consultation three times per year. All consultants were female, varying in age from 29 to 55 at the time of data collection. One consultant held a bachelor's degree, three held master's degrees, and one held a Ph.D. Four of the five program consultants had teaching experience ranging from 2 to 20 years. Four consultants were Caucasian and one was African-American. All were native English speakers, and one also spoke Spanish.

Site Staff: Participants were staff at intervention sites recruited for the Classroom Literacy Outcomes and Interventions (CLIO). CLIO is a national experimental study of the federal Even Start Family Literacy program designed to determine if high quality intervention programs will result in positive child and parent outcomes when compared with programs that continue with existing curricula.

Sites were located in 23 states, in all regions of the United States (see figure 2). The number of sites participating in the study varied from initial random assignment to June, 2006. Of the 47 sites originally assigned to the Pfl condition of the study, two sites withdrew prior to the initial training, and are not included in analyses for this study as no data were collected. One site had no 3 or 4 year old children enrolled during either of the years of this study, no data were collected from the site and they were not included in data analysis for this study. Two additional sites withdrew in year two. Data from these sites was collected in year one of the project, but not in year two. Three sites lost funding in the second year of the study, so their data are available for the first year but not the second. Data for year one from these five sites are included in this study. Finally, one site was recruited by Westat during late spring of 2005 and began implementing the Partners for Literacy curriculum in September, 2005, as a replacement for one program that had lost funding. Data from this site will not be included in analyses because their training and participation varied significantly from the 44 sites trained and participating in year one.

There were a total of 277 staff involved in data collection in this analysis. Of these staff, 52 were the director or lead teacher at their site, 96 were Early Childhood Education teachers, and 98 were Teacher Assistants or Teacher Aides. Thirty-four Parent Educators also participated in this intervention but are not included in this study. Staff at the sites varied considerably in education level, age, gender, ethnicity, and languages spoken. See Tables 1-2 for data on race/ethnicity and gender distribution for the participants. Staff at sites were included in analysis if they were teachers at a site when

any of the consultant ratings of implementation occurred, at the mid-year or end of year visits to sites.

Instruments

Several different instruments were used in data collection. These instruments were created for data reporting purposes and have not been evaluated for reliability.

Instruments included: (1) Program consultant site and staff ratings of implementation status (2) the Teacher Feedback Form and (3) the Director/Mentor Teacher Feedback form.

Other documentation which provided data for this study included: (4) attendance records from seven training workshops, (5) travel expense reports, (6) program consultant written summaries of on-site trainings, (7) records of contact with sites by the project coordinator, through informal phone calls and emails, as well as a yearly survey on important site information.

The first factors, teacher and director attendance at training, are defined as the number of week-long training sessions that a teacher or director attended over the two year intervention. Training documentation provided the data for these factors. The next factor, the number of on-site consultations, is defined as the number of times the teacher was visited by a program consultant during the two year intervention. Data for the number of on-site consultations were obtained through travel expense reports and program consultant summaries of visits. Teacher and director compliance with record keeping is the next set of factors, defined as the number of feedback forms that the director and teacher returned over the two year period, which was obtained through data management logs of implementation data that was sent to the Partners for Literacy

offices. Teacher turnover, another implementation factor, is defined as changes in the lead teacher employed in the classroom over a two year period. Data for this factor were obtained through contact with sites by the program coordinator, as well as through site-visit reports completed by program consultants. The number of children, the final implementation factor, is defined as the number of children enrolled in the CLIO study at the time of implementation rating. These data were collected through a survey sent to sites by the project coordinator, and confirmation through observation in the classroom and enrollment data from Westat. It is important to note that this number varied during the year as children were enrolled and dropped out of Even Start programs. The number of children per classroom was based on the initial enrollment at the beginning of each year. Implementation ratings were made by program consultants, addressing the quality of Partners for Literacy Curriculum implementation independent of overall classroom quality, which was defined through meetings with all Partners for Literacy staff.

(1) Program Consultant Site and Staff Ratings.

Program consultants visited each of the participating sites consistently during both years. Sites were to be visited between two to three times per year. Site visits generally had one day per Early Childhood Classroom and one day per parent education classroom, with additional time provided if further staff training was deemed necessary. At mid-year and at the end of the school year, consultants completed ratings on the level of implementation of the Partners for Literacy curriculum, with separate ratings for the ECE and PE components. Program consultants rated program implementation based on close on-site observation at participating sites. Observations included the completion of consultation forms in the first year of implementation, including the content in the

Director/Mentor Teacher Feedback Form. Also, program consultants completed a checklist of elements of a literacy-rich Partners for Literacy Classroom in the first year of implementation, to provide further evidence on which to base their ratings. Consultants were trained to use a rating system for implementation by the principal investigator, who developed the ratings system. At each rating time point, criteria for each rating were reviewed with program consultants.

Ratings were provided for each classroom at each site and for the site overall. These ratings were based on a five-point system, with 1 identified as a very low level of implementation, and 5 indicating very strong implementation. Ratings were made based on a written set of criteria assigned for each rating, 1 through 5. These ratings were obtained in January 2005, June 2005, January 2006, and June 2006. A copy of the ratings protocol is included as Appendix C. Where clarification was necessary, the principal investigator discussed the criteria and characteristics of the sites. These ratings were provided to Westat as the implementation measure to be factored into outcome studies.

Westat completed site visits to all Partners for Literacy sites with independent observers. These observers completed an implementation rating that was based on criteria set by the Partners for Literacy Project, to act as an independent rating of implementation status. Though raw data for sites is not available from Westat, an analysis plan for the CLIO study indicated that ratings made by program consultants in the first year of implementation were correlated with independent ratings made by Westat observers in the classrooms in the first year, $r = 0.414$, $p = 0.006$ (Westat & Abt Associates, 2005). Correlation with independent raters of implementation provides evidence that ratings made by Partners for Literacy consultants have some external

validity. Partners for Literacy ratings of ECE classroom implementation are used in the current study as the implementation status.

(2) Teacher Feedback Form.

The Teacher Feedback Form and the Director/Mentor Teacher Feedback Form were created through an iterative process with Partners for Literacy staff, data management consultants, and the pilot implementation sites. The Teacher Feedback form is a one-page, paper and pencil instrument designed for completion by the classroom teacher. A copy of this instrument and the manual for completion of the form is included as Appendix A. The Teacher Feedback Form was designed through examination of the curriculum, and examination of components and processes that were unique to the intervention. As well, other measures of teacher behavior were reviewed for content, format, and design of the form. The Teacher Feedback Form is designed to record daily information on the LiteracyGames component of the Partners for Literacy curriculum. Information on the number of times a LiteracyGame is played each week is noted. The instrument also includes information on child attendance, as well as teacher ratings of several early literacy skills, including oral language and letter knowledge. Teachers also had the opportunity to provide comments regarding their implementation of the program.

(3) Director/Mentor - Teacher Feedback form.

The Director/Mentor - Teacher Feedback form is a two-page, paper and pencil instrument designed for completion by the site director or teacher mentor. A copy of this instrument and the manual for the completion of the instrument is included in Appendix B. The Director/Mentor - Teacher Feedback Form was designed through examination of the curriculum components and processes. This form is to be completed once every two

to four weeks. It includes a 20 minute observation of the teacher using LiteracyGames with two children, as well as a 20 minute observation of Interactive Book Reading while the teacher reads to children. The instrument includes questions regarding the completion of LiteracyGames and Interactive Book Reading, including how well component strategies (3N strategy and 3S strategy) were used. It also includes ratings on other strategies in Interactive Book Reading, including Expanded Book Reading procedures, and the use of “Wh” questions. Finally, one page of the Director/Mentor Teacher form was dedicated to questions about a specific LiteracyGame used that day, including achievement of key learning objectives for that game. Directors also had the opportunity to provide comments regarding the teacher’s implementation of the program.

(4) Attendance records from workshop trainings

Information about teacher and director attendance at one of the seven training workshops are available through spreadsheets created and maintained by the graduate assistants present at each training. Participants “checked in” to each training in the morning and afternoon, and documentation was kept for attendance for each morning and afternoon session.

(5) Travel Expense Reports

Information about the timing and occurrence of program consultant visits to sites was obtained through review of central university expense and reimbursement reports filed for each consultant during the duration of the intervention.

(6) Records of on-site consultation visits

Information about the number and dates of on-site training were confirmed through the travel records of the program consultants. Each consultant kept detailed notes about

each site visit including the time of the visit, duration, staff present, current state of curriculum implementation and information shared with the staff in an attempt to coach the staff and improve implementation. From these visits, and from ongoing close communication with the staff by the project consultants and the project coordinator, it is also possible to determine the staff turnover that occurred in all sites, as well as the number of children enrolled in the classroom that are part of the CLIO study. For some classrooms, this number differed from the total number of children in the classroom due to age restrictions on CLIO participation – participation was limited to three and four year old children.

(7) Records of Site Contact Maintained by the Project Coordinator

Data on staff turnover and class enrollment were confirmed through review of several databases maintained by the project coordinator as records of contact with all participating sites. This contact took the form of email surveys at the beginning of each year, and phone calls and emails throughout the year.

Procedure

The analysis includes data collected over two years, from June 2004-June 2006 - the implementation period for the Partners for Literacy Project in the CLIO study.

Implementation ratings were made by program consultants for each classroom within participating sites, as well as an overall site rating, twice per year, once in January and once in June.

All participating teachers were asked to complete monthly feedback forms as part of the curriculum. All participating site directors were asked to complete bi-weekly feedback forms as part of the curriculum. At one site, the director was asked to complete

this form less frequently for each classroom (once per month), as six classrooms were located at her site and observation on a bi-weekly basis of 12 teaching staff was not possible. Feedback forms were sent to the FPG Child Development Institute by fax or mail following completion. Research assistants and the project assistant cleaned data according to a manualized set of procedures. Following cleaning, these measures were sent to the data management team who entered data into a SAS database, checked the data, and sent data back to the research assistants for clarification in the case of incomplete or illogical data. A second wave of data cleaning and entry occurred in September-December 2006. Data previously excluded from analysis due to low levels of participant consent (<80% participants consented) were examined thoroughly for later consent that was provided. Data held by the program assistant for further clarification was entered during this second wave of data entry. From the data sets maintained by the data management team, the number of received teacher feedback forms and the number of received director/mentor teacher forms were determined for each classroom following the second wave of data entry, in January 2007.

Information on the training of staff members was collected throughout the seven training sessions. During the large training sessions, each staff member registered for the training with PflL staff, and signed in before each morning and afternoon session, so that documentation for professional development credits could be collected. Attendance data were collected for each individual. Research assistants checked the attendance information during the training sessions to assure that all present had signed in, helping ensure the accuracy of the attendance data.

Records of on-site training, travel expense reports, and project coordinator records of site contact were reviewed for data regarding the number of on-site training visits, the timing of these visits, teacher turnover, director turnover, and the number of children per classroom.

Data from all the above sources were entered into a database in SPSS.

Data Analysis

All analyses were carried out using SPSS 13.0 for Windows and SAS 8.0. Two discrete sets of analyses and follow up procedures were conducted. Missing data were considered and imputed when deemed reasonable through consultation with a statistician. Analysis of attrition sites was conducted to examine systematic attrition based on initial ratings which might bias results. One site was excluded from data analysis following visual inspection, which revealed that the site had completed paperwork in an invalid manner. This site contained two classrooms, and the data from both classrooms was found to be invalid. In addition, an ANOVA of ratings by program consultant was conducted to rule out differences in ratings among the program consultants.

First, in order to address the question of correlation among program implementation factors, simple bivariate correlation was conducted between each pair of factors. Factors included teacher attendance at training, director attendance at training, the number of on-site training sessions, teacher compliance with record keeping, director compliance with record keeping, teacher turnover, and the number of children per classroom.

To address the second specific research question, regarding prediction of implementation ratings based on the seven factors listed above, the final implementation

rating was used as an outcome measure for logistic regression. This regression model initially included all seven factors listed above. Forwards and backwards stepwise selection, including all variables initially, then entering significant variables in the final model, was used to determine the model that best fit the data and theoretical background (Kachigan, 1991, p. 185). Data pre-analysis included confirmation that the data met assumptions regarding collinearity and singularity, normality, linearity, homoscedasticity, and independence of variables (Tabachnick & Fidell, 1996, p. 134-138). Based on pre-analysis of data, two classrooms were eliminated from analysis due to exceedingly high values of the Mahalanobis distance (greater than the value of chi-square for the appropriate degrees of freedom).

We addressed the research question regarding changes in ratings of implementation quality change over a period of two years in classrooms using the Partners for Literacy Curriculum. Four time points were available for analysis, the mid-year and year end data for each year of implementation. First, descriptive data were generated using SPSS to determine qualitative changes in implementation over time. A repeated measures analysis using SAS version 8.0, proc mixed procedure was conducted to determine changes in ratings over time, using a general linear models procedure. Director attendance was included in this analysis as a factor based on literature suggesting that leadership and supervisory support of an intervention may mediate implementation effects (Harachi et al, 1999). Values of mean ratings were plotted in SPSS to permit for visual inspection of these results.

CHAPTER 3

RESULTS

Descriptive data indicated that the outcome variable (consultant rating) was approximately normally distributed, with appropriate variability at each time point (see table 5).

Simple bivariate correlations were calculated to determine the relationship among program implementation factors, including teacher and director attendance at training, teacher and director compliance with record keeping, number of on-site training sessions, teacher turnover, and the number of children per classroom. Calculations were based on separate year one and year two data, see tables 6 and 7. From the first to the second year, the number of children per classroom was moderately correlated, $r = .643, p < .001$. Directors who were trained in year 1 were more likely to be trained in year two, $r = .320, p = .016$. Finally, directors who sent in more feedback forms in year 1 were more likely to send in a larger number of feedback forms in year two, $r = .340, p = .006$. These relationships were non-significant for teacher data.

In year one, the number of consultant site visits was moderately correlated with the number of teacher feedback forms completed, $r = .481, p < .001$. This relationship did not hold for the second year. In both the first and second years, classrooms with more children were more likely to be visited a greater number of times (year 1: $r = .510, p < .001$; year 2: $r = .510, p < .001$). Classrooms with a larger number of children were more likely to return a large number of feedback forms in year 1 ($r = .327, p = .004$) but

were less likely to return a large number of feedback forms in year 2 ($r = -.335, p = .012$). In both the first and second year, classrooms in which the director was trained were more likely to have their teachers return a larger number of teacher feedback forms (year 1: $r = .416, p < .001$; year 2: $r = .280, p = .039$). In the second year alone, teachers who attended training were more likely to return a larger number of teacher feedback forms ($r = .356, p = .008$). Also, there was a relationship between the number of forms that teachers returned and the number of forms directors returned for the second year exclusively ($r = .575, p < .001$). All other correlations were found to be non-significant.

Forward logistic regression was conducted to determine which independent variables (teacher turnover, teacher training, director training, number of children in classroom, number of site visits, number of teacher feedback forms returned, number of director/mentor teacher feedback forms returned) were predictors of classroom status as implemented or not implemented (implementation rating of 1, 2 or 3 compared to implementation rating of 4 or 5). Data screening led to the elimination of two outliers. Logistic regression was performed on data from year one and from year two separately. Data from year one indicate that no measured predictors contributed to a model predicting implementation status, see table 8 for further information. Regression results from the second year of implementation indicate the overall model of one predictor (number of teacher feedback forms returned) was statistically reliable in distinguishing between classrooms that implemented the program well versus those classrooms that did not implement the curriculum well ($-2 \text{ Log Likelihood} = 47.143; \chi^2(2) = 24.250, p < .001$). This model correctly classified 80.8% of cases. Regression coefficients are presented in Table 9. Wald statistics indicated that the number of teacher feedback forms

significantly predicts implementation status in year 2 and odds ratios for these variables indicate significant change in the likelihood of strong implementation status.

To answer the question of change over time, an analysis of site attrition first had to be conducted in order to confirm that site attrition was not systematic. To determine whether sites that left the Partners for Literacy Study had significantly lower ratings at the initial assessment than the remaining sites, a t-test comparing initial means of the two groups was conducted. Please see table 10 for the means of these two groups. A t-test on these two groups was found to be non-significant, $t(72) = -.19, p = .85$, and there is no evidence that lower performing groups had greater attrition.

Another potential source of bias was the ratings by different program consultants. An ANOVA comparing ratings of program consultants was determined to be non-significant, $F(4) = 1.395, p = 0.236$. Thus, program consultants gave similar ratings when collapsing across all time points. An analysis of time by consultant effects on ratings was not possible, as one consultant was not present during the final time point, which would have skewed data. Ratings for this consultant's site were instead made by one of the four remaining program consultants, who all visited the sites prior to rating.

Descriptive analysis of the data indicated that at rating time one, 58 classrooms has low levels of implementation (rating of 1 or 2), 16 had moderate levels of implementation (rating of 3), and none had high levels of implementation (rating of 4 or 5). See table 5 for more detailed data. At rating time two, 33 classrooms had low levels of implementation, 21 had moderate levels of implementation, and 18 had high levels of implementation. At rating time 3, 21 sites had low levels of implementation, 12 had moderate levels of implementation, and 18 had high levels of implementation. At rating

time four, 14 classrooms had low levels of implementation, 11 had moderate levels of implementation, and 31 had high levels of implementation. See figure 3 for visual representation of the descriptive data.

Repeated measures analysis using general linear models was used to examine whether the relationship between the level and change in the consultant site rating was associated with whether the director attended training, based on results of logistic regression. SAS version 8.0, proc mixed program was used to run this analysis. The database was restructured using the SPSS 13.0 restructure function, to create a SAS database with separate cases for each time point. Based on results of logistic regression, director and teacher attendance at training was included in the model to determine if it moderated changes in ratings. This model indicated that director training was significantly associated with better ratings, $b = .02, p = .01$. See table 11 for the complete model. This model indicates that ratings significantly improved across the four time points after adjusting for director training, $b = .52, p < .001$. While having the director attend training was associated with higher ratings overall, the director's attendance at training did not moderate the change in rating.

CHAPTER 4

DISCUSSION

There were three separate sets of analyses carried out in this study: first, the relationship between various factors thought to influence implementation; second, modeling of the influence of these factors on implementation ratings; and third, the change in implementation ratings over time. The results of the first analysis suggest that factors thought to be related to implementation were at best weakly correlated with one another, and appear to represent distinct factors. In this study, teacher attendance at training was found to be weakly positively correlated with teacher compliance with record keeping, for the second year of implementation alone. Teacher attendance at training was not found to significantly relate to teacher turnover. While director attendance did not relate to their participation in record keeping, it did relate positively to teacher participation in record keeping, in both the first and second year. Teacher participation in record keeping was not related to teacher turnover in either year. In addition to the hypothesized correlations, several additional factors were found to be related. Unexpectedly, the number of children per classroom was found to relate to the number of consultant visits in both years of implementation, and also to the number of teacher feedback forms completed, though the direction of this relationship changed between year one and two. Finally, the number of teacher feedback forms completed was moderately positively related to the number of director/mentor teacher feedback forms completed, in the second year of implementation.

The results of the second analysis suggest that modeling of factors related to implementation was possible in the second year of implementation alone. Based on the unexpected pattern of correlation found in these variables, the data fit a model of implementation only in the second year of implementation, and included only one significant predictor variable. Logistic regression found that the year one data did not fit any significant model. In the second year, the data fit a significant model including the number of teacher feedback forms returned, and based on this model over 80 % of classrooms could be correctly categorized. This finding fits well with conceptualization of implementation, as classrooms were asked to do several things, one of which was to return feedback forms. Consultants had limited information about the amount of data that classrooms had completed, so this factor was unlikely to bias ratings. Specifically, there were significant delays between the completion of data, cleaning, entering, and receiving data reports from data management.

It was unexpected to find little relationship between the training of teachers and directors with implementation quality. Several factors should be considered when interpreting this result. First, significant effort was put into training all teachers and directors, and in fact an additional training was conducted in September 2004 to address this need. Thus, there were few staff that were not trained, which led to little variability in training status. Second, even if one staff member at a site was not trained, all sites had several staff and it is likely that the majority of staff at each site were trained. There might have been diffusion of training across staff members at a site. Finally, the comprehensive nature of training manuals might have decreased the importance of in-person training. Also, with low levels of initial implementation, the results might have

been statistically nonsignificant despite an effect of training on implementation, but this difference was not visible due to restriction in the range of implementation quality initially.

The finding of a significant model for year two alone is not surprising given the context of the current intervention. It took a significant amount of time to train teachers in the current curriculum. Based on ongoing process evaluation, and feedback from pilot studies of the curriculum, some elements of the curriculum were not in place until the second year. These elements were not considered in first year ratings of implementation, which prevents biased results due to curriculum-level factors. Given the initially low levels of implementation, it is also reasonable to expect that few factors would relate to implementation. That is, it is difficult to measure factors relating to implementation if few classrooms are implementing the program.

The third set of analyses for this study was the change of implementation ratings over time. Repeated measures analysis indicated that ratings of implementation of the curriculum improved significantly over time. Director attendance at training was related to this pattern of findings, but did not moderate change over time. Teacher attendance at training was not related to improved ratings over time. This result supports hypotheses, in that the change of implementation clearly increased, and did so consistently across all four time points.

The change in implementation ratings was statistically significant, and in addition these findings are clearly significant on a practical level. The first rating of implementation occurred after 3-4 months of implementation, in the winter of the first year of CLIO. At this time, all participating classrooms had implementation ratings of 1,

2, or 3, indicating low to moderate implementation of the curriculum. After two years of implementation, this pattern had changed significantly. At the fourth rating time, 31 classrooms had implementation ratings of 4 or 5, indicating strong implementation. This represented 55.3% of participating classrooms. In this intervention, classrooms required about two years time, including 4-6 onsite training sessions and two large-scale trainings in order implement the curriculum at a high level of proficiency, according to ratings made by program consultants. Even after this considerable intervention, only about 55% were implementing the curriculum at a high level of proficiency, and an additional 20% of classrooms were implementing curriculum at a moderate level. This study suggests that researchers should consider the timing of child outcome measures, as complex interventions require significant time for implementation to develop. In addition, after implementation of the curriculum is strong, the curriculum will require implementation for some time before the impact of the intervention can be assessed. Further, the level of implementation should be considered in analysis of outcomes, such that interactions between exposure time and implementation level should be factored into outcome analyses in order to consider the degree to which an intervention produces change.

In the context of the CLIO study, it is likely that sites were variable in terms of the amount of time during which they successfully implemented the curriculum prior to evaluation of the curriculum. If strong implementation is considered to be ratings of 4 or 5, as has been consistent in this study, then 18 sites achieved strong implementation between fall 2004 and spring 2005, allowing children and families in these classrooms at least one full year of exposure to the curriculum prior to measuring outcomes. An additional 13 classrooms achieved a rating of 4 or 5 between Fall 2005 and Spring 2006,

allowing children and families in these classrooms a few months of exposure, at best, to the curriculum prior to measurement. The time at which classrooms began implementing the curriculum at a strong level (and thereby, the estimated exposure time to the curriculum for families) will be important to consider when measuring outcomes of the curriculum itself.

The variability in implementation time and quality was unexpected given the significant efforts to support implementation at all sites. Goss, Noltemeyer and Devore (2007) report that the provision of performance feedback and modeling both serve to improve treatment fidelity. In this intervention, program consultants provided consistent feedback to teachers and directors on their implementation through site visits. Program consultants were trained to use a variety of coaching strategies, including modeling, to assist in the ongoing training of teachers. Goss, Noltemeyer, and Devore (2007) also report that rehearsal and feedback is useful in promoting treatment integrity. Rehearsal and feedback were incorporated in large scale national trainings with teachers each year. Bellg and colleagues (2004) also presented a set of “best practices” for implementation data. First, they recommend that all groups in a study receive approximately equal contact with the intervention. In this case, all sites received between four and six site visits, and received two large scale national trainings. All sites received the same materials supporting the intervention. Bellg and colleagues (2004) also recommend that the duration of treatment for all groups is equal. In this study, all participating sites were recruited at the same time, had one year of baseline assessment, and then received two years of intervention. The authors also recommend that all sites receive proper training to deliver the intervention. Contact with sites and delivery of training was monitored by

both outside observers (Westat observed each site once per year), as well as program consultants from within the intervention. The group recommended that interventionists should plan for implementation setbacks, and this was done through the provision of an additional large-scale training in Chapel Hill, NC to train those teachers unable to attend one of three planned trainings in the first year. As well, teacher attrition was monitored by program consultants and additional on-site training was provided when necessary, such as in the case of teacher attrition. Ensuring that interventionists were properly trained is another recommendation set forth by Bellg and colleagues (2004), which was achieved through standardized training procedures in the Partners for Literacy project, the use of the same trainers for trainings, and repeated training to minimize intervention drift over time. Clearly, the Partners for Literacy intervention adhered to many of the “best practices” denoted in the field of treatment integrity research, but implementation quality was slow to develop.

The results of the current study add to the literature base in the field. To this point, there is little research on the nature of change in implementation over time. In a study by Harachi, Abbott, Catalano, Haggerty and Fleming (1999), the implementation of several elements of a multi-component school-based prevention program were examined at four time points over the course of two years. This study used continuous coding in observation of the classroom in twenty 90-second intervals to examine frequencies of target behaviors in the classroom. This classroom observation system suggested that teachers adopted some of the practices promoted over the 18 months of the intervention, and that they adopted more of the target practices than did control group teachers. This

study indicated intent to study the incorporation of target practices over time, as they did not examine this variable.

The study by Harachi and colleagues (1999) raises important issues in the field of intervention and program evaluation. They note that the study of implementation can “elaborate on the mechanisms through which changes in the outcomes operate” (p. 711). Investigation of implementation represents a change from the historical pattern of comparing baseline to post-intervention, or control group to experimental group. In the past, differences in the experimental group were considered for their difference from the control group, the effects of which were largely left in the “*black box*” of intervention processes. Investigation into the black box of intervention can assist in determining *what* causes a change in outcomes, and not just that these changes exist. Information about what causes changes is important for future research, in formulating interventions and prevention programs that effectively promote positive outcomes.

In the Partners for Literacy study, there were few factors that related to positive implementation. The single predictor of strong implementation status, the return of teacher feedback forms for each classroom, does add to the literature in terms of factors that relate to positive implementation. Teachers that are willing to report on their own implementation of a curriculum are more likely to be positively implementing that curriculum.

There is also little research concerning factors that might be associated with implementation of educational curricula. One study addressed this issue was found in the field of substance abuse prevention. A study by Rohrbach, Graham and Hansen (1993) found that implementation of the substance abuse prevention program was highly

variable in their study and “surprisingly low” (p. 249). They found that 21% of teachers implemented all their lessons in year one of the study, and 56% implemented some of the lessons. In the second year of their study, only 23% of teachers implemented any of the lessons in their intervention (Rohrbach et al, 1993). It is notable that in this study, teachers received only one training session prior to the first year of intervention. Also, they found that strong implementation was positively related to less teacher experience, increased active participation in training, increased enthusiasm about the program, stronger teacher self-efficacy, and more experience with teaching methods consistent with the intervention. These factors differed from implementation factors included in the present study. Finally, the study by Rohrbach and her colleagues (1993) found that the degree to which a program is delivered as intended is positively associated with positive and immediate program outcomes. Further study of the Partners for Literacy curriculum, and the relationship of implementation of this curriculum with child outcomes will be necessary in order to determine if implementation relates to positive program outcomes.

One theoretical model of implementation discussed previously is the Technology Model of Implementation. The Technology Model of Implementation, as noted earlier, requires four elements of implementation: (1) the specification of treatment in a manual, (2) training of interventionists in the uniform delivery of the treatment and supervision to ensure uniform training occurs, (3) monitoring of treatment delivery, and (4) inclusion of intervention fidelity as an independent variable in outcome analysis (Carroll et al., 2000). The Partners for Literacy curriculum in the CLIO study included all of the above requirements. The Partners for Literacy curriculum included an extensive set of manuals that outline specific elements of the intervention (Wasik & Sparling, 2005). Training of

interventionists took place at seven large-scale trainings over a period of two years using the same procedures for all interventionists. The program developers trained the interventionists during workshops, and ongoing monitoring was provided by both program consultants and project directors. Monitoring of treatment fidelity also took place through teacher and director completion of Feedback Forms, and Program Consultant ratings of classroom implementation. Ratings of intervention fidelity will be used by Westat as an independent variable in outcome analysis. Thus, procedures in the Partners for Literacy study fit well with requirements for implementation as outlined in the Technology Model (Carroll et al., 2000).

In addition to the good fit between the Partners for Literacy study procedures and the Technology Model of Implementation, it is possible to examine some of the links between various aspects of the Technology Model. A common assumption in implementation research has been that training of interventionists leads to greater implementation of that intervention. In this study, we found that director attendance at training was significantly and positively correlated with teacher compliance with record keeping, but that only the teacher's compliance with record keeping contributed significantly to a model of ratings. This might have been due to the non-significant relationship between director attendance at training and director's own compliance with record keeping. It may be the case that training provided directors at sites with a good knowledge of the expectations that they should foster in teacher behavior, without informing directors of their own specified roles in the curriculum, or that directors might be unwilling to report on their own implementation of the curriculum, and might consider teacher implementation of the curriculum more important.

An alternative model of implementation was put forth by Kitson, Harvey, and McCormack (1998) which asserts that implementation is a function of the interplay between (a) the level and nature of research evidence, (b) the context in which the evidence is to be implemented, and (c) the method in which the process is facilitated. Kitson and colleagues suggest that currently, especially in the field of health care, practice is assumed to be directly affected by research (evidence) through synthesis and distillation by interested parties. Evidence includes the degree to which research is credible, rigorous, and systematic, which determines the extent to which the research is likely to be implemented. In other words, the publication of a credible study is thought to be sufficient for implementation of novel practices. However, research is only one of three equally important factors in the implementation process. In addition to research, the context and environment in which the research shall be implemented is crucial. The organizational culture of the environment in which implementation is to occur is important to consider, as contexts in which professional roles are clear, leadership is strong, and performance is consistently monitored are more likely to have strong implementation of a given intervention. Finally, facilitation involves interpersonal methods to support the implementation of a program, through coaching, encouragement, and demonstration of technical competence. Kitson and colleagues (1998) indicate that evidence, context, and facilitation all need to be clearly understood in order for implementation to occur. With regards to the Partners for Literacy study, clearly there is strong evidence to support the curriculum, as there is a strong research basis for component activities of the curriculum (Wasik, Herrmann, Dobbins, & Roberts, 2000; Harbin, Herrmann, Wasik, Dobbins, & Lam, 2004) which served as a partial determinant

for its inclusion in the CLIO study. The research basis for Partners for Literacy was quantitative as well as qualitative, rigorous, systematic, and there is a strong consensus within the field that the results are credible. The context of the intervention was clearly considered, through lessons learned from previous research and literature review, studies at two pilot sites, and continuous process evaluation throughout the CLIO study. Specifically, Partners for Literacy was a teacher- and family-centered curriculum with an emphasis on continued education and training in the program, training teachers both at large-scale trainings as well as in-context, at on-site trainings within the classroom, based on observation of that classroom. Finally, facilitation involved mentoring by program consultants, through phone and on-site contact. Sites had consistent support from program consultants who received training in “coaching,” as well as basic interpersonal change skills such as respect, empathy, and authenticity. All program consultants had sufficient education and experience to provide some degree of “authority” for staff, as well as specialized training in the Partners for Literacy curriculum. Due to random assignment and the structure of the CLIO study, there was no flexibility in the quality of the context of sites in regards to their willingness to accept change and readiness for change. However, overall the study had strong evidence, context, and facilitation for change. According to the study by Kitson and colleagues (1998), this should have increased the likelihood that implementation would be strong. Despite these efforts, implementation of the Partners for Literacy Curriculum developed slowly.

Limitations

There are a number of factors that limit the interpretation of the results of this study. First, despite the large number of children and parents involved in this study, the

number of classrooms and teachers were too limited to permit examination of the data through factor analysis or other data reduction methods. In addition, the number of classrooms per site was relatively low, with low variability, which prevented the use of nested data modeling. In addition, the lack of findings in modeling through logistic regression may have to do with limited sensitivity of measures both in terms of what was measured and the level at which it was measured. Specifically, the outcome measurement of implementation rating was made on a 1-5 scale, which might have been insensitive to subtle influences by factors included in this study. As well, by measuring implementation on the classroom level, the sample size might have been too low to detect differences on a classroom level alone.

There are also considerations regarding the outcome variable chosen for this study, consultant rating of implementation. A shortcoming of most rating scale techniques is that to some degree, rater opinion influences results. Program consultants had large amounts of exposure to staff at implementation sites, allowing for personal relationships to build. While in some cases this may have beneficial effects on implementation (through social pressure), it also has the potential to differentially influence some sites more than others. While this was minimized through careful delineation of a rating codebook, there is a possibility that opinion influenced ratings to some minimal degree. This is a shortcoming of most rating scale techniques, one that can be mitigated by interrater reliability data. Unfortunately, no interrater reliability data are available in the Partners for Literacy data, and the expense of travel to program sites prevented multiple ratings of implementation from occurring. However, comparisons of ratings across program consultants revealed no significant differences in the ratings given

by any consultant. If this factor was to influence ratings, it likely did so equally across the different raters, each with classrooms of varying quality.

In this study, there was considerable attrition of classrooms following the first year of study. Analysis of implementation in both the attrition sites and sites who remained in the study did not reveal differences in the degree to which classrooms were implementing the curriculum initially. However, there are a variety of factors which may have affected attrition that the current study did not address. Generalizability of this study may be limited, as the group of classrooms which remained in the study might have shared some set of unknown characteristics making them more likely to continue with this research study, though this is not likely related to motivation to implement a curriculum or effort in implementation, as ratings of implementation were similar between the two groups.

Areas for Future Study

The results of the current study indicate the need for future work in a number of areas. First, this study used a single measure of implementation status, consultant rating. With the release of the data set for the CLIO project in the future, it will be possible to obtain additional sources of information regarding the quality of implementation at sites. Future studies should investigate the degree to which Westat ratings of implementation, and director ratings of implementation, are consistent with consultant ratings. Preliminary data discussed earlier suggested that consultant ratings of implementation were the most conservative estimate of implementation, but these results should be elaborated through detailed study. Furthermore, there are several other sources of data that may quantitatively inform an estimate of implementation, such as the responses that

directors and teachers provided on measures of implementation (teacher and director/mentor – teacher feedback forms). The content of these documents has not yet been analyzed.

There are several factors that may influence implementation, which were not included in the present study. Results from logistic regression in this study found that many of the included factors were non-significant in predicting implementation rating. Examination of additional factors may serve to assist in maximizing program implementation in future large-scale studies of this type. First, there was no measure of teacher acceptability for this curriculum, which likely had a strong influence on teacher willingness to implement the curriculum as designed. Demographic variables of teachers and directors might have also been significant in implementation of the program (e.g., education level and teacher experience). Teacher perceptions of a number of factors might have influenced their implementation of the curriculum as well, such as their perceptions of the adequacy of training, difficulty to implement the program, and the degree to which the curriculum incorporated practices that the teachers perceived as similar to what they were already doing in the classroom. All these factors might be further studied through further contact with teachers, through a questionnaire, interview, or focus group.

Parent and child factors might have also influenced the success with which a teacher was able to implement a curriculum. For example, teachers and directors might have introduced activities to parents and children, but depending on the parent or child's reaction to these activities, the teacher might have become differentially likely to present the curriculum materials again. While teachers might have presented curriculum

materials to parents and children, the curriculum was not fully “implemented” unless parents and children participated in these activities. If parents and children were not fully participating in the study, one cannot ensure treatment receipt. Parent and child perceptions of the curriculum activities would be useful to consider for future studies.

Higher-level administrative factors also had the potential to influence the degree to which teachers and directors implemented the Partners for Literacy curriculum. For example, sites in states where the State director of Even Start supported the curriculum may have been encouraged to implement the curriculum more strongly. In addition, funding was a significant influence on many sites. Many sites uncertain financial status during the course of the study, and four classrooms lost funding during the first year of the study, adding to classroom attrition. In addition, seven classrooms lost funding in the spring of 2006 or early summer, and of those sites where information was available, an additional nine classrooms were operating under a different funding structure (e.g., were not “Even Start” classrooms any more, and used Title 1 funding or mental health funding). Eleven classrooms had their funding cut significantly, leading to a reduction in services. Additional classrooms also were unsure of their status. Only 17 classrooms could confirm that they would be operational in the 2006-2007 year. Clearly, many sites in this study were influenced by financial uncertainty. Teachers and directors at these sites had little job security, and families had little security for depending on the services offered. This situation likely influenced implementation in a negative way. Teachers and directors who had a lack of job security are certainly less likely to be motivated to implement a curriculum. There is also little motivation to learn to use a new curriculum if one is uncertain if the curriculum can be used in subsequent years. In future studies,

funding agencies should consider eliminating this worry, by ensuring that funding will be in place for participating sites for the duration of the study.

Finally, it will be extremely important to consider implementation data when interpreting child outcomes from the Partners for Literacy curriculum. Several possibilities exist for possible child outcomes. First, in the absence of significant positive child outcomes overall, a future study could consider differential outcomes based on the level of implementation of classrooms. In this way, researchers could determine the extent to which a lack of child outcomes was due to failures in implementation versus failure of the curriculum to bring about positive change. Second, in the face of positive child outcomes, implementation should be factored into analyses of change to determine the extent to which positive child outcomes are related to intensive exposure to the curriculum itself, and to increase estimates of effect size of the intervention.

Considering implementation level of classrooms in child outcome analysis may serve to increase effect size for the study. Through the inclusion of implementation, some “noise” in the data (e.g., classrooms in the experimental condition who did not implement the intervention, thereby reducing differences between the control and experimental groups) will be eliminated. This may provide a more accurate estimate of the effects of the Pfl curriculum itself, as opposed to just the effects of inclusion in the experimental condition of the CLIO study.

Current Findings

The current findings have several implications for the nature of research design and professional development. First, it was difficult to determine factors that served to increase implementation. This may be due to low initial implementation, or perhaps due

to the set of factors that were chosen for measurement in this study. Clearly, additional work needs to focus on determining strategies to maximize implementation in interventions.

Second, it is clear that curricula take significant time and training to be implemented effectively. In this study, classroom teachers and site directors received a detailed curriculum manual, two 3-4 day long large scale trainings, and 4-6 on-site training sessions over the course of two years. At the first measure of implementation, after at least one large scale training and one on-site training, only about 20% of classrooms showed a moderate level of implementation of the curriculum, and none showed strong implementation. After two years, when all training had been conducted, this figure increased significantly to about 55% of classrooms that had strong implementation, and an additional 20% that had moderate implementation. The implementation of a curriculum is complex and takes an extended length of time to complete.

Research conducted on curricula, therefore, must take this training and implementation time into account when determining effects of a curriculum on child achievement and behavior. In this study, measurement of child outcomes after 4 months would significantly underestimate the effects of a curriculum on child achievement, as the curriculum had not been implemented as intended. In order to fairly assess the effects of a curriculum on child outcomes, one must first verify that the curriculum was implemented strongly, and then determine the length of time in which children received that highly implemented curriculum. Once both these factors are considered, the actual effects of the curriculum itself can be measured through comparison to a matched control

sample. It is only through the careful study of implementation of interventions that the effects of these interventions can truly be determined.

Table 1

Role of Staff

Staff Position	Frequency	Percent
Director	52	18.57%
Teacher	96	34.29%
Teacher's Assistant	98	35.00%
Parent Educator	34	12.14%

Table 2

Ethnic and Gender Distribution of Staff

Ethnicity	Gender	Frequency	Percent of Sample
African-American	Male	1	0.35%
	Female	34	12.14%
American Indian/Alaska Native	Male	0	0%
	Female	2	0.71%
Asian/Pacific Islander	Male	0	0%
	Female	5	1.79%
Biracial/Multiracial	Male	0	0%
	Female	1	0.35%
Caucasian	Male	7	2.50%
	Female	162	57.86%
Hispanic	Male	5	1.79%
	Female	57	20.36%
Unknown	Male	1	0.35%
	Female	5	1.79%

Table 3.

Descriptive Data – Implementation Factors

Variable	N	Mean	Std Dev.
Number of site visits	76	4.14	1.00
Number of Enrolled Children – Year 1	76	7.80	4.75
Number of Enrolled Children – Year 2	56	7.89	4.61
Teacher Feedback Forms returned – Year 1	76	5.38	3.29
Teacher Feedback Forms returned – Year 2	59	6.05	3.22
Director/Mentor Teacher Feedback Forms returned - Year 1	76	5.13	5.85
Director/Mentor Teacher Feedback Forms returned - Year 2	64	5.05	5.06

Table 4.

Descriptive Data – Categorical Implementation Factors

	N	% Trained	% untrained
Teacher Trained in Year 1	76	85.3%	14.7%
Teacher Trained in Year 2	56	82.1%	17.9%
Director Trained Year 1	75	80.7%	9.3%
Director Trained Year 2	56	96.4%	3.6%
	N	% Teacher Changed	% Teacher did not change
Teacher Turnover	55	25.5%	74.5%

Table 5.

Descriptive Data – Site Ratings

Rating Time	N	Mean	Std Deviation	Frequency of Ratings
Fall 2004				1: N = 41
				2: N = 17
	74	1.66	.816	3: N = 16
				4: N = 0
				5: N = 0
Spring 2005				1: N = 12
				2: N = 21
	72	2.63	1.041	3: N = 21
				4: N = 18
				5: N = 0
Fall 2005				1: N = 9
				2: N = 12
	51	2.76	1.124	3: N = 12
				4: N = 18
				5: N = 0
Spring 2006				1: N = 7
				2: N = 7
	56	3.38	1.287	3: N = 11
				4: N = 20
				5: N = 11

Table 6.

Bivariate Correlations – Year 1

	Teacher Trained Year 1	Director Trained Year 1	# of Visits	# Children per classroom Year 1	T.F.F. Returned Year 1	Director/Mentor T.F.F. Returned Year 1	Teacher Turnover
Teacher Trained Year 1	*	r = -.133 p = .255 N = 75	r = .060 p = .608 N = 76	r = .110 p = .346 N = 76	r = .174 p = .133 N = 76	r = -.087 p = .496 N = 64	r = -.153 p = .158 N = 55
Director Trained Year 1	r = -.133 p = .255 N = 75	*	r = .010 p = .932 N = 75	r = .076 p = .516 N = 75	r = .416 p < .001 N = 75	r = .225 p = .052 N = 75	r = .003 p = .983 N = 55
Number of Visits	r = .060 p = .608 N = 76	r = .010 p = .932 N = 75	*	r = .510 p < .001 N = 76	r = .481 p < .001 N = 76	r = -.026 p = .823 N = 76	r = -.156 p = .255 N = 55

Number of	r = .110	r = .076	r = .510		r = .327	r = .012	r = .131
Children per	p = .346	p = .516	p < .001	*	p = .004	p = .921	p = .340
classroom Year 1	N = 76	N = 75	N = 76		N = 76	N = 76	N = 55
Teacher Feedback	r = .174	r = .416	r = .481	r = .327		r = .196	r = -.016
Forms Returned	p = .133	p < .001	p < .001	p = .004	*	p = .089	p = .907
Year 1	N = 76	N = 75	N = 76	N = 76		N = 76	N = 55
Director/Mentor							
Teacher Feedback	r = -.087	r = .225	r = -.026	r = .012	r = .196		r = .097
forms returned	p = .496	p = .052	p = .823	p = .921	p = .089	*	p = .482
Year 1	N = 64	N = 75	N = 76	N = 76	N = 76		N = 55
Teacher Turnover	r = -.153	r = .003	r = -.156	r = .131	r = -.016	r = .097	
	p = .158	p = .983	p = .255	p = .340	p = .907	p = .482	*
	N = 55	N = 55	N = 55	N = 55	N = 55	N = 55	

Table 7.

Year 2 Bivariate Correlations

	Teacher	Director	# of Visits	# Children	T.F.F.	Director/Mentor	Teacher
	Trained	Trained		per	Returned	T.F.F. Returned	Turnover
	Year 2	Year 2		classroom	Year 2	Year 2	
				Year 2			
Teacher Trained		$r = -.090$	$r = .115$	$r = -.082$	$r = .356$	$r = .193$	$r = -.193$
Year 2	*	$p = .511$	$p = .400$	$p = .546$	$p = .008$	$p = .153$	$p = .15$
		$N = 56$	$N = 56$	$N = 56$	$N = 55$	$N = 56$	$N = 55$
Director Trained	$r = -.133$		$r = -.118$	$r = .101$	$r = .280$	$r = .208$	$r = .114$
Year 2	$p = .255$	*	$p = .385$	$p = .460$	$p = .039$	$p = .124$	$p = .409$
	$N = 75$		$N = 56$	$N = 56$	$N = 55$	$N = 56$	$N = 55$
Number of Visits	$r = .060$	$r = .010$		$r = .510$	$r = -.010$	$r = .163$	$r = -.156$
	$p = .608$	$p = .932$	*	$p < .001$	$p = .941$	$p = .199$	$p = .255$
	$N = 76$	$N = 75$		$N = 76$	$N = 59$	$N = 64$	$N = 55$

Table 8.

Logistic Regression Year 1

	β	<i>Wald</i>	<i>df</i>	<i>p</i>	Odds Ratio
Constant	-1.025	10.813	1	0.001	.359

Table 9.

Logistic Regression Year 2

	β	<i>Wald</i>	<i>df</i>	<i>p</i>	Odds Ratio
Teacher Trained Year 2	-21.306	<.001	1	.999	<.001
Number of Teacher Feedback Forms Returned Year 2	.316	5.932	1	.015	1.372
Constant	-1.354	2.240	1	.134	.258

Table 8.

Mean initial ratings of attrition group and non-attrition group.

	Mean	Standard Deviation
Classrooms Completing Intervention	1.67	.81
Classrooms that did not complete intervention	1.63	.83

Table 9.

Model of Time and Director training on rating.

Effect	DirtrainedY1	Estimate	Standard Error	<i>df</i>	<i>t</i>	<i>p</i>
Intercept		1.3784	0.1474	72	9.35	<.0001
Time		0.5231	0.04802	176	10.89	<.0001
DirtrainedY1	0	-0.8180	0.3204	72	-2.55	0.0128

Figure Captions

Figure 1. Research design for the Classroom Literacy Interventions and Outcomes study

Figure 2. Partners for Literacy Site Locations

Figure 3. Boxplot of mean ratings at four time points.

Figure 1.

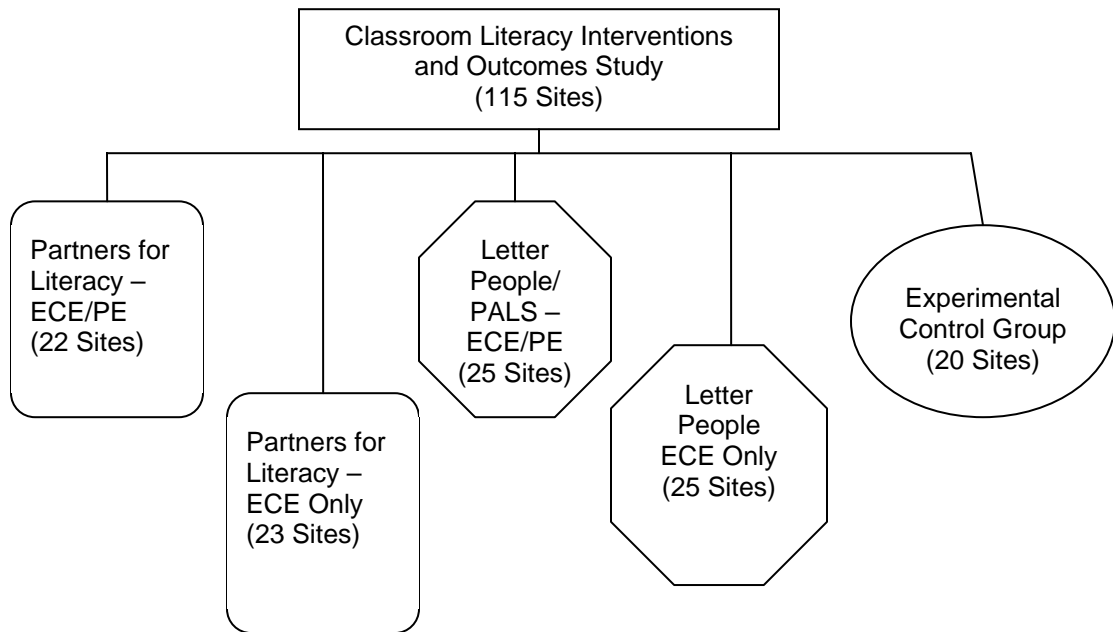


Figure 2

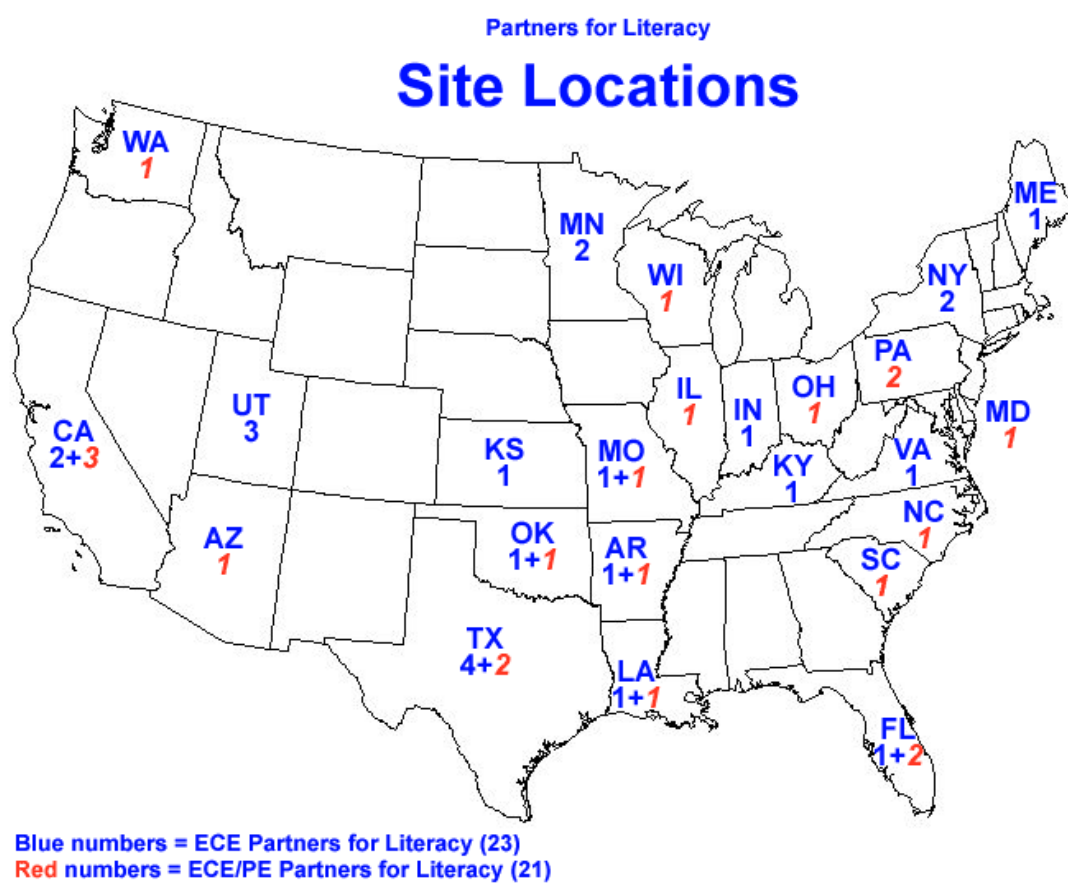
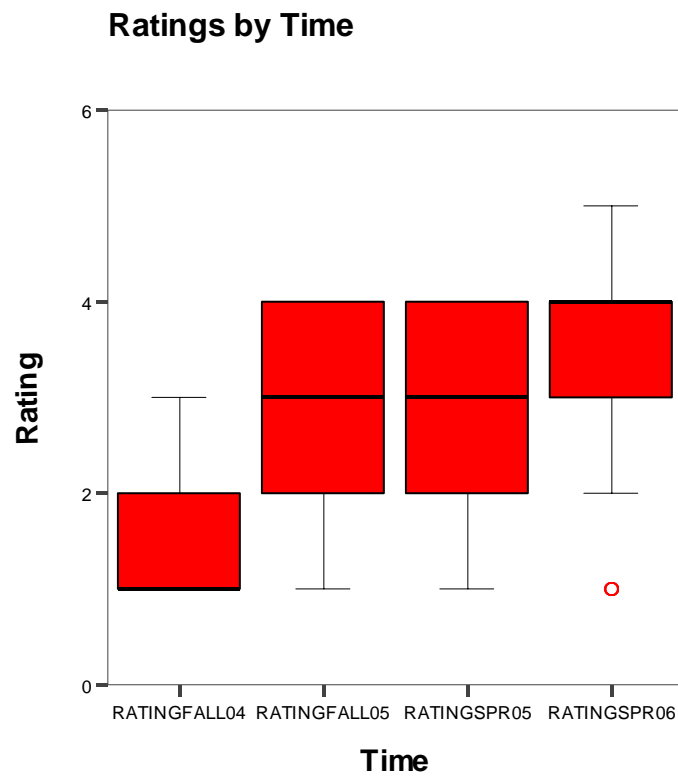


Figure 3



Appendix A:

Teacher Feedback Form and Manual

Early Childhood Education Teacher Feedback Form

Manual

It is important for us to obtain information on the performance of the children in the early childhood classrooms, especially related to the Partners for Literacy curriculum. We recognize that you will be rating children who come from a variety of backgrounds and different learning levels. You may have children in your program who are younger than others, just learning your program, developmentally atypical, or learning English. These children may have lower ratings in general. Even children who have been in your program for two years might not obtain the highest rating. We expect variability.

Some of this information should be checked *daily* and some should be filled out *every 4 weeks*. Please complete this form and mail or fax it to us at the end of each *4 week time period*.

Step 1: Fill out the top line on the form (site name, classroom number or name, your name, and your first and last name initials). *Please do not fill out the ID section that says "For office use only."*

Step 2: Fill out each child's first and last name and first and last initials in the first column. If you are using pages 2 and 3, you do not need to fill in the date again.

Please note: It will help you to copy the form now before using it. If children are added to the program, you can add their names and initials at the bottom of the form and then copy this revised form for later use.

Please fill out the date for the Monday of the first week that you use this form.

Step 3: Complete columns A-D *daily*, to show how many *LiteracyGames* are played each week.

Step 4: Write in the date next to Early Literacy on the day you complete columns E-L.

Step 5: Complete columns E-J *every 4 weeks* to summarize each child's performance.

Step 6: Complete column K by writing down how many days each child was present in that four weeks

Step 7: At the end of each *4 week period*, please *mail or fax* your forms to:

Leslie Moye, FPG Child Development Institute,
517 S. Greensboro Street, CB# 8040, Carrboro, NC, 27510
FAX: 919 966-1786

If you have a question about this form, please ask your Director to contact Leslie Moye at 919 966-4024 or by email at moye@mail.fpg.unc.edu.

Specific Directions:

Columns A, B, C, and D: *On a daily basis*, fill in a bubble each time a *LiteracyGame* is played.

Please be sure to fill the bubbles in from left to right and top to bottom.

Example

Tom Smith played *LiteracyGame* #111 – 4 times week A
Tom Smith played *LiteracyGame* #112 – 3 times week A
– 7 *LiteracyGames* week A

LiteracyGames				
1st Monday date:	2d Monday date:	3d Monday date:	4th Monday date:	
A. Tally number of Literacy- Games	B. Tally number of Literacy- Games	C. Tally number of Literacy- Games	D. Tally number of Literacy- Games	
● ● ● ●	○ ○ ○ ○	○ ○ ○ ○	○ ○ ○ ○	
● ● ● ○	○ ○ ○ ○	○ ○ ○ ○	○ ○ ○ ○	
○ ○ ○ ○	○ ○ ○ ○	○ ○ ○ ○	○ ○ ○ ○	
○ ○ ○ ○	○ ○ ○ ○	○ ○ ○ ○	○ ○ ○ ○	

Note that the number of times a child participates may vary from child to child, depending on their attendance and developmental level. You may, for example, provide an extra opportunity for a child who seems to need more time.

Column E: Oral Language

Fill in a number from the following rating to describe each child's performance on oral language:

(If the child's home language is not English, still rate the child's English language for this form.)

- 1 Child **does not speak**.
- 2 Child **speaks very little**, rarely initiates conversation, and responds with only a word or two when spoken to.
- 3 Child responds in **one or two word sentences** and occasionally initiates conversation or responds to a partner.
- 4 Child **sometimes initiates conversation** and may elaborate a sentence or two on a topic. They may also engage in some back and forth conversation.
- 5 Child **frequently initiates and responds to conversations** with adults and peers.

Column F: Letter Knowledge

Fill in a number from the following rating to describe each child's performance on letter knowledge:

- 1 Child has **very little or no** knowledge about letters.
- 2 Child can **point to 2 letters** correctly.
- 3 Child can **point to and name 3 to 5 letters** correctly.
- 4 Child can **point to and name 6 to 10 letters** correctly (when letters are not in order).
- 5 Child can **point to and name 11 or more letters** correctly (when letters are not in order).

Column G: Rhyming

Fill in a number from the following rating to describe each child's performance on rhyming:

- 1 Child **does not participate** in rhymes/songs.
- 2 Child **participates in saying** rhymes/songs.
- 3 Child **fills in blanks** for rhyming words on familiar rhymes/songs.
- 4 Child can identify simple words that sound alike (does "all" sound like "tall"?)
- 5 Child **matches pictures** of rhyming words by sound. (matches "hat" and "cat")

Column H: Interactive Book Reading

Fill in a number from the following rating to describe each child's performance on Interactive Book Reading:

- 1 Child **rarely attends to** a book being read (or not at all).
- 2 Child **usually attends to** a book being read.
- 3 Child **attends to a book** being read and can point to and name simple pictures.
- 4 Child **responds to simple questions** about the book like "who..." or "what..."
- 5 Child **asks questions, answers more complex questions, can retell a very simple story.**

Column I: Writing

Fill in a number from the following rating to describe each child's performance on writing:

- 1 Child **does not use** writing materials.
- 2 Child **uses a variety of writing materials** and scribbles.
- 3 Child can **write separate shapes** (non-letters).
- 4 Child can **write several letters** of his or her name.
- 5 Child **writes a simple word**.

Column J: Social Emotional

Fill in a number from the following rating to describe each child's performance on Social-Emotional Skills:

- 1 Child is **not aware** of how he or she or others feel.
- 2 Child can **label his or her feelings** of happy, sad, and mad.
- 3 Child can **recognize and label the feelings** of happy, sad, and mad in others.
- 4 Child can **tell events that make him or her feel** happy, sad and mad.
- 5 Child can **often respond to others** empathetically.

Column K: Attendance

Fill in the number of days each child was present for the **4 week** time period.

Please feel free to use the comment box provided on the last page of this form for suggestions or comments.

Site name

Classroom name

Teacher name

Initials

For office use >

Site ID

Classroom ID

Teacher ID

Page No. 1

Child Name

Initials

For office use
↓
ID

1st Monday date

2nd Monday date

3rd Monday date

4th Monday date

A. Tally number of Literacy-Games

B. Tally number of Literacy-Games

C. Tally number of Literacy-Games

D. Tally number of Literacy-Games

E. Oral language

F. Letter knowledge

G. Rhyming

H. Interactive book reading

I. Writing

J. Social Emotional

K. Attendance

Early Literacy

Date:

1. _____

2. _____

3. _____

4. _____

5. _____

6. _____

7. _____

8. _____

9. _____

10. _____

Site name _____

For office use >

Page No. 2			
Child Name	Initials	For office use ↓	ID
11. _____	_____	_____	_____
12. _____	_____	_____	_____
13. _____	_____	_____	_____
14. _____	_____	_____	_____
15. _____	_____	_____	_____
16. _____	_____	_____	_____
17. _____	_____	_____	_____
18. _____	_____	_____	_____
19. _____	_____	_____	_____
20. _____	_____	_____	_____

Site name _____

Classroom name _____

Teacher name _____

Initials _____

For office use >

Site ID _____

Classroom ID _____

Teacher ID _____

Page No. 3

Child Name

Initials

For office use
↓
ID

21. _____

22. _____

23. _____

24. _____

25. _____

26. _____

27. _____

28. _____

29. _____

30. _____

1st day of week:

1st day of week:

1st day of week:

1st day of week:

A. Tally number of Literacy-Games

B. Tally number of Literacy-Games

C. Tally number of Literacy-Games

D. Tally number of Literacy-Games

Early Literacy

Date:

F. Oral language

G. Rhyming

H. Interactive book

I. Writing

J. Social Emotional

K. Attendance

102

Appendix B:

Director/Mentor – Teacher Feedback Form and Manual

Director/Mentor Teacher Form

Purpose

The purpose of this instrument is to provide information and feedback on program implementation of the Partners for Literacy curriculum. We will use this data to summarize the quality of the Partners for Literacy curriculum after using it at all the Partners for Literacy sites.

Please fill out this form **every other week**. You will have to observe the teacher using the *LiteracyGames* activities with at least 3 pairs of children and you will be required to observe at least 10 minutes of book-reading with two children in the classroom, 8 children total.

These observations may be done in sequence, or they may be split up into two observation sessions.

You will be required to complete one observation and feedback form for each Teacher and Teacher Assistant in each ECE classroom at your site.

Please note that a separate form should be completed for the Teacher and the Teacher Assistant.

Step 1:

Please fill out the requested information at the top of the form. **Please make sure to mark the box that indicates if you are rating the Teacher or the Teacher Assistant.**

The classroom description refers to the group of children in the class separate from their teacher's name such as the room number or anything else that is a short description of the classroom that clearly identifies it to you. Please fill out the requested information at the top of the second page as well. **You do not need to fill out the section labeled "for office use only" at the top of the pages.**

Step 2:

At the end of each completed observation, please mail your forms to:

Partners for Literacy
Frank Porter Graham Institute
CB# 8040
University of North Carolina at Chapel Hill
Chapel Hill, NC 27599

or fax them to Partners for Literacy at 919.966.1786.

If you have a question about this form, please contact your program consultant.

Part I: LiteracyGames

This section is to be completed after watching at least 3 *pairs* of children play *LiteracyGames* with the Teacher or Teacher Assistant under observation.

For this part, please rate the extent to which you agree with certain statements about the Teacher or Teacher Assistant. If the Teacher or Teacher Assistant did not discuss or use an item, circle the rating “N” for not used.

The Teacher or Teacher Assistant:

- a. ***Was familiar with the LiteracyGames lesson plan and script*** – The teacher was prepared for the session, the teacher knew the *LiteracyGames* lesson plan and script well enough to smoothly move through the LiteracyGame.
- b. ***Was able to tailor the script to the child(ren)’s developmental needs*** – The teacher was able to modify the scripts provided on *LiteracyGames* to suit the needs of the children that they were working with.
- c. ***Had materials needed for LiteracyGame ready and accessible*** – The teacher had all of the materials that they needed for the entire session (all *LiteracyGames* played) close to them and prepared.
- d. ***Actively watched, listened, and waited for clues about the child’s readiness for an activity (Notice)*** – The teacher noticed what the child was doing, and began the 3N cycle.
- e. ***Easily got the LiteracyGame started and prompted the child to take the activity a step further (Nudge)*** – The teacher nudged the child during the LiteracyGame and therefore used the second “N” of the 3N strategy.
- f. ***Described the child’s action or response to the adult’s nudge thoroughly (Narrate)*** – The teacher used narrate, the last part of the 3N strategy.

Part II: Interactive Book Reading

This part is to be completed **after** watching reading activities with two children for at least 10 minutes each with the Teacher or Teacher’s Assistant under observation.

- g. ***Used the 3S strategy*** – used the See, Show, Say strategy while reading a book.

- h. **Used Wh Questions** – used Who, What, Where, When, Why and How questions when reading and discussing the story with the children.
- i. **Used Expanded Book Reading** – prompted the child to predict events in the story, re-tell the story in sequence and other expanded book reading practices.
- j. **Helped Child with Book Concepts** – integrated knowledge of book concepts into story reading, such as pointing out the front and back cover, pages, title, author, and other print concepts to the child.
- k. **Integrated Activities into Book Reading such as letter knowledge, rhyming, and alliteration** - pointed out letters in text, rhyming words, how letters and sounds correspond, similarities and differences in words and words that start with the same letter.
- l. **Tailored the book reading strategies to the child's developmental needs** - The teacher was able to modify the use of book reading strategies (3S, Wh questions, book concepts, expanded book reading) to suit the needs of the children that they were working with.

Comments

Directors, please reflect on the work of your teacher or teacher assistant and describe one aspect of the curriculum that you think that teacher has been using exceptionally well.

Reflect also on how the teacher or teacher assistant uses the Partners for Literacy curriculum, and provide us with one area where you think your site could benefit from some extra support or assistance in using the Partners for Literacy curriculum.

Please feel free to add additional comments about the ECE classroom for these two weeks.

Director/Mentor Teacher Form (101-1)

This form should be filled out for each teacher and teacher assistant.

Please note: observations for Parts I and II may be done at different times across a two week period.

Site Name _____ Classroom Description _____ Director Name _____ Initials _____ I am rating the: <input type="checkbox"/> Teacher <input type="checkbox"/> TA Teacher/ TA Name _____ Initials _____	<i>For Office Use Only</i> Site ID _____ Classroom ID _____ Director ID _____ ID _____ Teacher/TA ID _____
----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	---------------------------------------------------------------------------------------------------------------------------

PART I: LiteracyGames

Please fill out Part I every 2 weeks after you have observed the Teacher or Teacher Assistant use LiteracyGames with at least 3 pairs of children.

Date Observed: ____/____/____	The extent to which you agree (circle one from each row)				
Please rate the extent to which you agree or disagree with the following statements. Teacher or Teacher Assistant:	Strongly Disagree	Somewhat Disagree	Somewhat Agree	Strongly Agree	Not used
Was familiar with the <i>LiteracyGames</i> lesson plan and script	1	2	3	4	N
Was able to tailor the script to the children's developmental needs	1	2	3	4	N
Had materials needed for LiteracyGame ready and accessible	1	2	3	4	N
Actively watched, listened, and waited for clues about the children's readiness for an activity (Notice)	1	2	3	4	N
Easily got the LiteracyGame started and prompted the children to take the activity a step further (Nudge)	1	2	3	4	N
Described the children's action or response to the adult's nudge thoroughly (Narrate)	1	2	3	4	N

Part II: Interactive Book Reading

Fill out Part II every 2 weeks after you have observed the Teacher or Teacher Assistant for at least 10 minutes during *reading* activities with 2 different children (20 min. total).

Date Observed: ____/____/____	Child 1			Child 2		
How well did the teacher or teacher assistant do the following:	Fair	Good		Fair	Good	
	Excellent			Excellent		
Frequently used the 3S strategy (See, Show, Say)	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>
Frequently used Wh Questions (Who, What, Where, When, Why)	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>
Frequently used Expanded Book Reading	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>
Helped child with book concepts	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>
Integrated activities into book reading such as letter knowledge, rhyming and alliteration.	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>
Tailored the book reading strategies to the child's developmental needs	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>

Date _____	<i>For Office Use Only</i>
Site Name _____	
Classroom Description _____	
Director Name _____ Initials ____	
Staff Name _____ Initials ____	
	Site ID _____
	Classroom ID _____
	ID Director ID _____
	Staff ID _____

LiteracyGame 101 I Know About Books Cycle 1	None or a little	Sufficient amount	A lot
The teacher tried to elicit the following from the child...			
Hold the book correctly; turn pages one at a time			
Point to the front and back of the book; point to where to start reading			
The Teacher...	None or a little	Sufficient amount	A lot
Sat in a comfortable spot with 2 children			
Had an easy book prepared and read the book more than once			
Pointed out the front and back covers of the book, named the title and author before reading the story			
Pointed to the first word on the page before beginning to read that page			
Asked the child to point to the front and back of the book, and where to begin reading on a page			
Asked child to demonstrate pretend reading			
Encouraged the child to use the child vocabulary word (book)			
Circle each teacher vocabulary word that the teacher used:	Cover Front Back Page		
Expanded the game: Talked about the way they read, as they read			

Comments:

Appendix C:

Program Consultant Coding System – Ratings of Site Implementation and Classroom Quality

Partners for Literacy ECE Site Ratings January 15, 2005

Directions: use a rating of 1 to 5 for each site, with a rating of 1 being the lowest and 5 being the highest.

1 = no or almost no implementation (example TX site that was not using Pfl at all) or very low level of implementation, or using inconsistent with program procedures, such as in a large group setting. If you did not observe interactive book reading, the site cannot get a rating higher than 1.

2 = some examples of Pfl. Example, teachers used the LiteracyGames with two children at a time, but there was no evidence of interactive book reading during the day. Little evidence of working to involve all aspects of the Pfl curriculum. Little director involvement.

3 = Evidence of LiteracyGames with all children and Interactive Book Reading with all children. Evidence of enriched caregiving during part of the day, some Pfl activities during circle time. Some director involvement and mentoring.

4 = program can be described as doing very well...but still have obvious room for improvement. Should be conducting LiteracyGames and Interactive Book Reading with all children, using Circle time appropriately, reasonably strong centers, good evidence of 3N, 3S, Wh questions, expanded book reading, enriched caregiving, and problem solving. Using teacher feedback form. Director form may or may not have been initiated.

5 = Clear evidence of all components and aspects of the Pfl program. Program is doing an excellent job. **Each** staff is using the Pfl curriculum correctly. Evidence of strong staff coordination. Strong use of LiteracyGames, Interactive Book Reading. Strong evidence of 3N, 3S, Wh Questions, Expanded book reading, Enriched Caregiving, Problem Solving, Classroom Books. Using teacher feedback form and mastery sheets consistently and appropriately. Director is actively involved in mentoring. Site shows evidence of ability to expand on the Pfl in appropriate ways (i.e., using themes of LiteracyGames throughout the day, finding creative ways to use LiteracyGames. Modifying activities to be consistent with Pfl strategies and procedures.

References

- Bellg, A. J., Borrelli, B., Resnick, B., Hecht, J., Minicucci, D. S., Ory, M., Ogedegbe, G., Orwig, D., Ernst, D. & Czajkowski, S. (2004). Enhancing treatment fidelity in health behavior change studies: Best practices and recommendations from the NIH behavior change consortium. *Health Psychology, 23*, 443-451.
- Bierman, K.L., Coie, J.D., Dodge, K.A., Greenberg, M.T., Lochman, J.E., McMahon, R.J. & Pinderhughes, E. (2002). The implementation of the Fast Track Program: An example of a large-scale prevention science efficacy trial. *Journal of Abnormal Child Psychology, 30*, 1-17.
- Burchinal, M. R., Campbell, F. A., Bryant, D. M., Wasik, B. H. & Ramey, C. T. (1997). Early intervention and mediating processes in cognitive performance of children of low-income African American families. *Child Development, 68*, 935-954.
- Campbell, F. A., Pungello, E. P., Miller-Johnson, S., Burchinal, M. & Ramey, C. T. (2001). The development of cognitive and academic abilities: Growth curves from an early childhood educational experiment. *Developmental Psychology, 37*, 231-242.
- Campbell, F. A. & Ramey, C. T. (1994). Effects of early intervention on intellectual and academic achievement: A follow up study of children from low-income families. *Child Development, 65*, 684-698.
- Carroll, K. M., Nich, C., Sifry, R. L., Nuro, K. F., Frankforter, T. L., Ball, S. A., Fenton, L. & Rounsaville, B. J. (2000). A general system for evaluating therapist adherence and competence in psychotherapy research in the addictions. *Drug and Alcohol Dependence, 57*, 225-238.
- 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.
- Dickstein, S., Seifer, R., Eguia, M., Kuersten-Hogan, R. & Magee, K. D. (2002). Early Head Start MAP: Manualized assessment of progress. *Infant Mental Health Journal, 23*, 231-249.
- Domitrovich, C. E. & Greenberg, M. T. (2000). The study of implementation: Current findings from effective programs that prevent mental disorders in school-aged children. *Journal of Educational and Psychological Consultation, 11*, 193-221.
- Early Head Start National Resource Center (1999). *Technical Assistance Paper No. 1: Early Head Start Program Implementation: Start Up Planning*. Washington, DC: Early Head Start NRC.

- Goss, S., Noltemeyer, A. & Devore, H. (2007). Treatment integrity: A necessary component of Response to Intervention. *The School Psychologist*, 61, 34-38.
- Gresham, F. M., MacMillan, D. L., Beebe-Frankenberger, M. E. & Bocian, K. M. (2000). Treatment integrity in learning disabilities: Do we really know how treatments are implemented? *Learning Disabilities Research & Practice*, 15, 198-205.
- Halpern, R. (2001). Early Childhood Intervention for Low-Income Children and Families. In J. P. Shonkoff & S. J. Meisels, *Handbook of Early Childhood Intervention* (2nd Edition), p. 361-386. New York, NY: Cambridge University Press.
- Harachi, T. W., Abbott, R. D., Catalano, R. F., Haggerty, K. P. & Fleming, C. B. (1999). Opening the black box: Using process evaluation measures to assess implementation and theory building. *American Journal of Community Psychology*, 27, 711-731.
- Harbin, G., Herrmann, S., Wasik, B. H., Dobbins, D., & Lam, W. K. K. (2004). Coordinating services in family literacy programs. In B. H. Wasik (Ed.). *Handbook of Family Literacy*. Mahwah: NJ: Lawrence Erlbaum.
- Hutinger, P.L., & Johanson, J. (2000). Implementing and maintaining an effective early childhood comprehensive technology system, *Topics in Early Childhood Special Education*, 20(3), 159-173
- Kachigan, S. K. (1991). *Multivariate Statistical Analysis: A Conceptual Introduction* (2nd Edition). New York, NY: Radius Press.
- Kitson, A., Harvey, G. & McCormack, B. (1998). Enabling the implementation of evidence based practice: A conceptual framework. *Quality in Health Care*, 7, 149-158.
- Layzer, J. I. & St. Pierre, R. G. (May, 1996). *Early Childhood Programs: Adding a Two-Generation Perspective*. Cambridge, MA: Abt Associates Inc.
- Leventhal, H. & Friedman, M. A. (2004). Does establishing fidelity of treatment help in understanding treatment efficacy? Comment on Bellg et al. (2004). *Health Psychology*, 23, 452-456.
- Liaw, F. R., Meisels, S. J. & Brooks-Gunn, J. (1995). The effects of experience of early intervention on low birth weight, premature children: The Infant Health and Development Program. *Early Childhood Research Quarterly*, 10, 405-431.
- Lichstein, K. L., Riedel, B. W. & Grieve, R. (1994). Fair tests of clinical trials: A treatment implementation model. *Advances in Behavior Research and Therapy*, 16, 1-29.

- Love, J. M., Kisker, E. E., Ross, C., Raikes, H., Constantine, J., Boller, K., Brooks-Gunn, J., Chazan-Cohen, R., Tarullo, L. B., Brady-Smith, C., Fuligni, A. S., Schochet, P. Z., Paulsell, D. & Vogel, C. (2005). The effectiveness of Early Head Start for 3-year-old children and their parents: Lessons for policy and programs. *Developmental Psychology*, 41, 885-901.
- Masse, L. N. & Barnett, W. S. (1994). *A Benefit Cost Analysis of the Abecedarian Early Childhood Intervention*. ERIC Document ED479 989.
- McCormick, M. C., Brooks-Gunn, J., Buka, S. L., Goldman, J., Yu, J., Salganik, M., Scott, D. T., Bennett, F. C., Kay, L. L., Bernbaum, J. C., Bauer, C. R., Martin, C., Woods, E. R., Martin, A. & Casey, P. H. (2006). Early intervention in low birth weight premature infants: Results at 18 years of age for the Infant Health and Development Program. *Pediatrics*, 117, 771-780.
- Meyer, A., Nicholson, R., Danish, S., Fries, E. & Polk, V. (2000). A model to measure program integrity of peer-led health promotion programs in rural middle schools: Assessing the implementation of the sixth grade goals for health program. *Journal of Educational and Psychological Consultation*, 11, 223-252.
- Moncher, F. J. & Prinz, R. J. (1991). Treatment fidelity in outcome studies. *Clinical Psychology Review*, 11, 247-266.
- Nelson, G., Amio, J. L., Prilleltensky, I. & Nickels, P. (2000). Partnerships for implementing school and community prevention programs. *Journal of Educational and Psychological Consultation*, 11, 121-145.
- NICHD Early Childcare Research Network (2006). Child-care effect sizes for the NICHD study of early child care and youth development. *American Psychologist*, 61, 99-116.
- Noell, G. H., Witt, J. C., Slider, N. J., Connell, J. E., Gatti, S. L., Williams, K. L., Koenig, J. L. & Resetar, J. L. (2005). Treatment implementation following behavioral consultation in schools: A comparison of three follow-up strategies. *School Psychology Review*, 34, 87-106.
- Orrill, C.H. & Anthony, H.G. (2003). Implementing reform curriculum: A case of who's in charge. Paper presented at the *American Educational Research Association Annual Meeting*, Chicago, IL.
- Parkinson, J. L. (1975). *The impact of the first year of the Carolina Infant Curriculum: A study relating program implementation measures to out-come measures for the first two cohorts of the Abecedarian Project*. Unpublished doctoral dissertation, University of North Carolina at Chapel Hill.

- Paulsell, D., Kisker, E. E., Love, J. M. & Raikes, H. H. (2002). Understanding implementation in Early Head Start programs: Implications for policy and practice. *Infant Mental Health Journal*, 23, 14-35.
- Paulson, R. I., Post, R. L., Herinckx, H. A. & Risser, P. (2002). Beyond components: Using fidelity scales to measure and assure choice in program implementation and quality assurance. *Community Mental Health Journal*, 38, 119-128.
- Petchers, M. K., Biegel, D. E. & Snyder, A. (1991). Barriers to implementation of a cooperative model of respite care. *Journal of Gerontological Social Work*, 16, 17-32.
- Peterson, C. A. & McConnell, S. R. (1993). Factors affecting the impact of social interaction skills interventions in early childhood special education. *Topics in Early Childhood Special Education*, 13, 38-57.
- Ramey, C. T., Bryant, D. M., Wasik, B. H., Sparling, J. J., Fendt, K. H. & LaVange, L. M. (1992). Infant health and development program for low birth weight, premature infants: Program elements, family participation, and child intelligence. *Pediatrics*, 3, 454-465.
- Ricciuti, A. E., St. Pierre, R. G., Lee, W. & Parsad, A. (2004). *Third National Even Start Evaluation: Follow-Up Findings From the Experimental Design Study*. Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance.
- Rohrbach, L. A., Graham, J. W., & Hansen, W. B. (1993). Diffusion of a substance abuse prevention program: Predictors of program implementation. *Preventative Medicine*, 22, 237-260.
- Santacroce, S. J., Maccarelli, L. M. & Grey, M. (2004). Intervention fidelity. *Nursing Research*, 53, 63-66.
- Schweinhardt, L. J. (2003). *Benefits, Costs, and Explanation of the High/Scope Perry Preschool Program*. Paper Presented at the 2003 Biennial meeting of the Society for Research in Child Development (70th, Tampa, FL, April 24-27, 2003).
- Sparling, J. J. & Lewis, I. (1991). Partners: A curriculum to help premature, low birthweight infants get off to a good start. *Topics in Early Childhood Special Education*, 11, 36-56.
- Sparling, J. J. & Lewis, I. (1993). *Factors that help or hinder the delivery of early intervention*. Unpublished Manuscript, Frank Porter Graham Child Development Institute.

- St. Pierre, R. G., Ricciuti, A. & Creps, C. (2000). *Synthesis of Local and State Even Start Evaluations*. Washington, DC: U.S. Department of Education, Office of the Under Secretary, Planning and Evaluation Service.
- St. Pierre, R. G., Ricciuti, A., Tao, F., Creps, C., Swartz, J., Lee, W., Parsad, A. & Rimdzius, T. (2003). *Third National Even Start Evaluation: Program Impacts and Implications for Improvement*. Cambridge, MA: Abt Associates Inc.
- St. Pierre, R. G., Swartz, J. P., Murray, S. & Deck, D. (1996). *Improving Family Literacy: Findings from the National Even Start Evaluation*. Cambridge, MA: Abt Associates Inc.
- Tabachnick, B. G. & Fidell, L. S. (1996). *Using Multivariate Statistics (3rd Edition)*. New York, NY: HarperCollins College Publishers.
- Tao, F., Gamse, B. & Tarr, H. (1998). *National Evaluation of the Even Start Family Literacy Program: 1994-1997 Final Report*. Washington, DC: U.S. Department of Education, Planning and Evaluation Service.
- Waltz, J., Addis, M. E., Koerner, K. & Jacobson, N. S. (1993). Testing the integrity of a psychotherapy protocol: Assessment of adherence and competence. *Journal of Consulting and Clinical Psychology*, 61, 620-630.
- Wasik, B. H., Dobbins, D. R. & Herrmann, S. (2001). Intergenerational family literacy: Concepts, research, and practice. In S. Neuman & D. Dickinson. *Handbook of Early Literacy Development* (pp. 444-458). New York, NY: Guilford.
- Wasik, B. H. & Hermann, S. (2004). Family literacy: History, concepts, service. In B. H. Wasik (ed.) *Handbook of Family Literacy*. Mahwah, NJ: Lawrence Erlbaum Associates. pp. 3-22
- Wasik, B. H., Herrmann, S., Dobbins, D. R., & Roberts, J. E. (Fall, 2000). Family literacy programs: A promising practice for the 21st century, *North Carolina Association for Supervision and Curriculum Development*, 7-19.
- 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.
- Wasik, B. H. & Sparling, J. J. (2005). *Partners for Literacy Handbook 1: Early Childhood Education*. Chapel Hill, NC: Frank Porter Graham Child Development Institute.
- Westat & Abt Associates (December 2004). *Even Start Classroom Literacy Interventions and Outcomes Study: Study Design and Analysis Plan*. Unpublished Report

- submitted to the U.S. Department of Education Institute of Education Sciences, under contract number ED-01-CO-0120.
- Westat & Abt Associates (December 2005). *Even Start Classroom Literacy Interventions and Outcomes Study: Revised Analysis Plan*. Unpublished report submitted to the U.S. Department of Education Institute of Education Sciences, under contract number ED-01-CO-0120.
- Winton, P. (2006). The evidence-based practice movement and its effect on knowledge utilization. Unpublished Manuscript, Frank Porter Graham Child Development Institute.
- Witt, J. C. & Elliott, S. N. (1985). Acceptability of classroom intervention strategies. In T. R. Kratochwill (Ed.), *Advances in School Psychology* (pp. 251-288). Hillsdale, NJ: Erlbaum.
- Zins, J. E., Elias, M. J., Greenberg, M. T. & Pruetz, M. K. (2000). Promoting quality implementation in prevention programs. *Journal of Educational and Psychological Consultation*, 11, 173-174.