Colorado Humanities Motheread Study

Final Report 2017

Submitted To: Mile High United Way

Revised report submitted June 1, 2017

OMNI

Colorado Humanities Motheread Study

Final Report 2017

For more information, please contact:

Holen Hirsh, PhD

hhirsh@omni.org

303-839-9422, extension 122

Melissa Richmond, PhD

mrichmond@omni.org

303-839-9422, extension 166

For General Inquiries/Questions

p. 303-839-9422

f. 303-839-9420

OMNI Institute

899 Logan Street, Suite 600

Denver, CO 80203

www.omni.org

Acknowledgements: Program and evaluation activities described in this report were made possible by Mile High United Way, Colorado Humanities and Morgan County Early Childhood Council. This report is based upon work supported by the Social Innovation Fund (SIF), a key White House initiative and program of the Corporation for National and Community Service (CNCS). The Social Innovation Fund combines public and private resources to grow the impact of innovative, community-based solutions that have compelling evidence of improving the lives of people in low-income communities throughout the United States.

Table of Contents

Figures

Tables

Executive Summary

Overview: Colorado Humanities, in partnership with OMNI Institute and funded by the Corporation for National and Community Service via the Mile High United Way, conducted a randomized controlled trial impact study of the Motheread ® (Motheread/Fatheread Colorado; MFC) program. MFC is a community-based program delivered to parents to help enhance shared reading skills and the home literacy environment. The study was designed to test the impact of parents' participation in the MFC program on child literacy outcomes. The study also tests the impact of the program on parents' reading behaviors. OMNI also conducted an implementation study to track fidelity of program delivery over the course of the impact study.

Motheread, Inc., their affiliates, and independent researchers have conducted outcome evaluations across multiple states and through programs serving a variety of populations. Overall, results from the evaluations have demonstrated positive links between participation in the program and improvements in adult and child literacy outcomes and have provided 'preliminary' evidence of program impact on literacy outcomes. The current investigation targeted a 'moderate' level of evidence through a study with strong internal controls.

One hundred and forty-eight parents with children in five schools in Morgan County, Colorado were enrolled in the study between spring 2014 and spring 2016. Parents were randomly assigned to intervention or waitlist-control conditions (parents in the waitlist control were offered the program the following semester). Teachers and parents reported on children's literacy behaviors at the beginning and end of the semester. Teachers completed the TS Gold assessment and parents completed the Early Literacy Parent Questionnaire. Parents also reported on their reading behaviors at the beginning and end of the semester.

The following measures of child outcomes were used in the study:

- Teacher Assessment on TS Gold Literacy Scale
- Teacher Assessment on TS Gold Language Scale
- Parent Assessment on Language Awareness Scale
- Parent Assessment on Reading Scale
- Parent Assessment on Response to Print Scale
- Parent Assessment on Interest in Letters Scale
- Parent Assessment on Writing Scale

The following measures of parent outcomes were used in the study:

- Frequency of Reading Aloud
- Active Reading Scale
- Minutes per Week Spent Reading
- Designated Time for Reading
- Number of Books Read in a Sitting

Child Literacy Skills Results: Results for the impact of the MFC program on child literacy outcomes were mixed. There was not a statistically significant impact of the program on teacher-reported, child literacy outcomes, as measured by TS Gold. At the end of the semester, children whose parents participated in the program had similar TS Gold literacy and language scores as children whose parents did not participate in the program. However, exploratory analyses examining TS Gold scores approximately 15 months after parents completed the program suggests that the MFC program may have a delayed but meaningful effect on the TS Gold Language score. Children whose parents participated in the MFC program showed higher rates of growth on the TS Gold Language domain after the program ended than children whose parents were in the control group (those who never received the program and those who received the program after being waitlisted).

There was a statistically significant impact of the MFC program on parent-reported, child literacy outcomes, as measured by the Early Literacy Parent Questionnaire. At the end of the semester, children whose parents participated in the program had higher scores on the Reading, Language Awareness, and Print scales than children whose parents did not participate in the program. The Reading scale assesses how children interact around books, including engagement in shared reading. The Language Awareness scale assesses children's interest in and use of rhymes. The Response to Print scale assesses how children respond to words in their environment.

Parent Reading Behaviors: There was as statistically significant impact of the MFC program on parent reading behaviors. Parents who participated in MFC read aloud more often and indicated more active reading skills (e.g., stopping to point out letters, rereading stories, encouraging child to read with parent, etc.) after completing the program than parents who did not participate.

Results at a Glance

Outcome Measures		Impact
Child Literacy Skills	Teacher Assessment on TS Gold Literacy Scale: End of semester	—
	Teacher Assessment on TS Gold Language Scale: End of semester	—
	Teacher Assessment on TS Gold Literacy Scale: Longitudinal	—
	Teacher Assessment on TS Gold Language Scale: Longitudinal	✓
	Parent Assessment on Language Awareness Scale	✓
	Parent Assessment on Reading Scale	✓
	Parent Assessment on Response to Print Scale	✓
	Parent Assessment on Interest in Letters Scale	—
	Parent Assessment on Writing Scale	—
Parent Reading Behaviors	Frequency of Reading Aloud	✓
	Active Reading Scale	✓
	Minutes per Week Spent Reading	—
	Designated Time for Reading	—
	Number of Books Read in a Sitting	—

✓ *Statistically significant positive difference between intervention and control*

✓ *Statistically significant positive difference between intervention and control; results from exploratory analysis should be interpreted with caution*

— *No difference between intervention and control*

Conclusions: In sum, the results of the study indicate that MFC has a positive impact on the way parents read with their children. Parents in the study had more active reading behaviors and read aloud more often with their children after participating in MFC. Parents' participation in MFC also had an impact on their children's literacy skills, as measured by parent assessment. Finally, the results of the study indicate that there may be a delayed effect of parental participation in MFC on their children's literacy skills as measured by TS Gold.

Program Overview

Motheread ®, is an early childhood literacy intervention that aims to develop children's school-readiness, grow early literacy skills, and promote ongoing success in school by enhancing the reading environment in the home, including increasing the frequency and quality of reading with young children by parents. Motheread works to help close the third grade reading achievement gap by making books and reading a part of every home, including those in which parents cannot read or read well.

Motheread is delivered in Colorado through Colorado Humanities' program department, Colorado Center for the Book, under the Motheread/Fatheread Colorado (MFC) program umbrella. The mission of Colorado Humanities is to inspire the people of Colorado to explore ideas and to appreciate its diverse cultural heritage. The mission of Colorado Center for the Book, now a program department of Colorado Humanities, is to inspire Coloradans to read and write. Colorado Humanities & Center for the Book programs facilitate reading, listening, writing and learning in cultural activities and participation in civic dialogue. Colorado Humanities chose Motheread because it encourages parents and caregivers to read with children and talk about issues in their lives, which models civic dialogue. Moreover, Colorado Humanities recognizes that fostering language and literacy skills early in a child's life is critical to providing the foundation for a lifelong love of reading and learning in the humanities.

The MFC program was initiated by Colorado Humanities to address the challenges Colorado faces in ensuring that all children are reading on grade level by the end of third grade. In 2012, more than one-quarter of Colorado's third graders did not meet proficiency standards on Colorado's state test of reading, and these rates were higher for children living in poverty and English language learners.[1] Specifically, 79% of third graders in Colorado fluent in English demonstrated proficiency on the state reading assessment, compared to only 53% of children with Limited English Proficiency and only 11% of children who were Not English Proficient. Additionally, only 60% of Colorado children who met income eligibility for free/reduced lunch were reading at or above proficiency standards compared to 86% of children who were not eligible for free/reduced lunch.

The State of Colorado has recognized these gaps and the importance of quality early childhood programming that will provide children with the language and literacy skills that are critical for success in school. Colorado also has recognized that early childhood programming can be difficult to access and can vary in quality, and that "those of the highest quality serve relatively few children."[2] MFC aims to address these gaps by conducting trainings with adult educators that will provide adults with the necessary skills to improve children's reading environment and foster the skills needed to prepare children for success in school.

---

[1] 2012 TCAP data from the Colorado Department of Education, accessed at http://www.cde.state.co.us/assessment/CoAssess-DataAndResults.asp

[2] Colorado Reads: The Early Literacy Initiative, Spring 2012, accessed at http://www.colorado.gov/cs/Satellite/GovHickenlooper/CBON/1251617165703

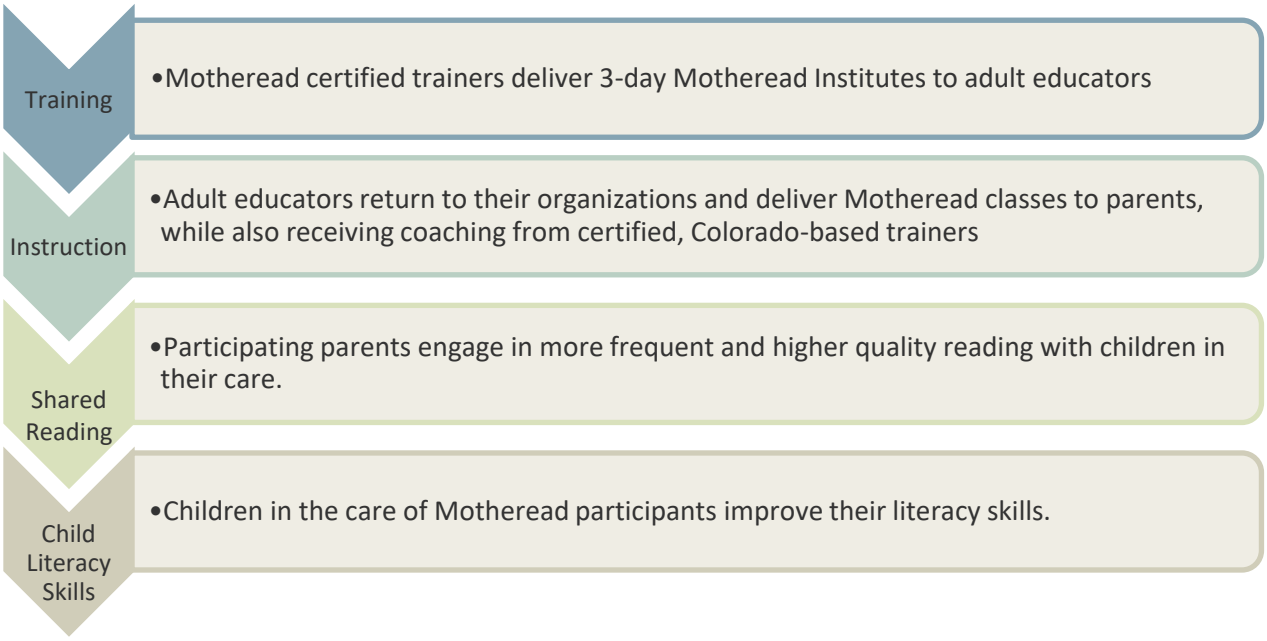
With the goal of closing the third grade reading achievement gap in Colorado, Colorado Humanities offers statewide trainings on the MFC program developed by Motheread, Inc. This program has a strong grounding in research and is aligned with Colorado Humanities' mission to engage Coloradans in the appreciation of our shared, yet diverse, culture and encourage a love of reading and books. Specifically, Motheread, Inc. developed a series of lessons that use award-winning children's literature as texts, many of which come from African-American, Latino, Native American and other ethnic cultures. The program uses great children's literature to instill reading as a way to access personal and cultural narratives and to introduce children to the humanities more generally. Each text has an accompanying curriculum that engages adults and children in story themes by promoting an interactive reading experience that relates to children's developmental experiences, such as expressing needs and feelings, establishing independence, understanding each family's values, and sharing and cooperating. The Motheread program focuses on adult participants' strengths; is learner-centered; includes reading, writing, listening, and speaking components; and promotes a social context supportive of collaborative learning.

MFC trains educators in Motheread, who then deliver it to adult caregivers. Adult educators trained in Motheread receive an initial 3-day training followed by one-on-one coaching during implementation. During MFC sessions, parents learn critical literacy skills, the "why" of reading, and specific instruction on how to "share stories" with children in an engaging and effective manner. These programs also seek to increase participants' frequency, enjoyment, comfort and skill in reading with children, even if the participants cannot read or have low reading levels. MFC is based on the theory that sharing stories in a manner that encourages child engagement is an effective and powerful instructional practice that improves children's literacy development.[3]

The following visual (Figure 1) portrays the structure of MFC. In brief, adult educators participate in a Motheread Institute, during which they are trained on the MFC curriculum by certified instructors. Adult educators then return to their organizations and provide classes to parents. It is recommended by MFC that sessions are led by two co-facilitators. Parents then apply what they learned in the classes when reading with children in their care, which in turn leads to improved children's literacy.

[3] Teale, W. H. (2003). Reading aloud to young children as a classroom instructional activity: Insights from research and practice. In A. van Kleeck, S. A. Stahl, & E. B. Bauer (Eds.), *On reading books to children* (pp.114-139). Mahwah, NJ: Erlbaum

Figure 1. Motheread Structure



MFC sessions are made up of the following components:

- ***Literacy as Experience.*** The goal of Literacy as Experience is to build adult literacy skills (comprehension, vocabulary, higher level questioning, discussion, etc.) through skill-building exercises. Along with skill-building, Literacy as Experience also is designed to help adults connect their life experiences to the story.
- ***Literacy as Art.*** In Literacy as Art, participants gain story-sharing skills such as reading with props, changing tone and inflection, etc.
- ***Literacy as Process.*** During Literacy as Process, adults explore the selected book's theme through other pieces of adult reading or writing.
- ***Story Extenders.*** The supplementary Story Extenders provide specific activities for parents to do with children that foster creative thinking and problem solving, and promote conversations between parents and their children.

MFC, in conjunction with Motheread, Inc., also developed a structured coaching model to help Colorado implementers solidify the techniques learned during Motheread Institutes and to deliver MFC with high fidelity. A Colorado-based trainer, trained and certified to conduct Motheread Institutes, delivers the coaching sessions. Specifically, each adult educator providing MFC to parents receives three in-person coaching sessions during their first six months of MFC implementation. Each coaching session includes the following activities:

- Observation of the full MFC session, using a structured observation instrument developed in partnership with Motheread, Inc.;
- A follow-up debrief/feedback session, approximately 30 minutes in length, immediately after the observation with the adult educator(s) being observed (MFC classes are often co-taught);
- An in-depth follow-up discussion (via phone or in-person) within 2 weeks of the observational session to review the completed observation tool, discuss strengths and opportunities for improvement in the delivery of MFC; and
- A short, written summary of the observational session that is shared with adult educators and program directors.

Motheread is based on research indicating that adult guidance and instruction are key factors in developing children's literacy skills.[4] The model teaches skills to adults to facilitate shared reading experiences with children that promote children's comprehension, the use of oral and written language, vocabulary development, and engagement in and enjoyment of reading. The program also promotes access to reading through the provision of high-quality books, accompanied by tools adults can use to facilitate ongoing reading experiences and utilize books effectively. Through high quality shared reading experiences and increased access to books, it is hypothesized that children reached by the program will increase language and literacy skills, leading to an increase in the number of children reading at grade level by third grade. Motheread bridges the academic gap by enabling all parents, even non-readers and second language learners, to explore books with the children in their care.

Evaluation Overview

The MFC impact study was a randomized controlled trial designed to test the impact of the program on child literacy outcomes with the aim of yielding a 'moderate' level of evidence for the efficacy of MFC. The impact study was designed to fill existing gaps in the literature base of Motheread. Specifically, existing studies have successfully demonstrated that adult participants improve their own literacy skills, and observe changes in their children's literacy behaviors. [5] However, none of the Motheread program evaluations utilized objective measures of changes in child literacy outcomes and none included a comparison or control group. To address these gaps, the Motheread impact evaluation was designed in

[4] Bus, A.G., van Ijzendorn, M.H., & Pellegrini, A.D. (1995). Joint book reading makes for success in learning to read: A meta-analysis on intergenerational transmission of literacy, *Review of Educational Research, 65*(1), 1-21.
[5] Gorham, Bertha. *Measuring Success in Motheread® Classes: Literacy and Parental Support Results*. Raleigh: Motheread, Inc., 2001.
[5] Wilder Research Center. *Minnesota Humanities Commission Motheread/Fatheread® Program Evaluation*. St. Paul, 2001.
[5] Clegg and Associates. *Motheread/Fatheread®: An Evaluation of Five Program Sites in Southwest Washington State*. Seattle, 1999. [5] Wilder Research Center. *Minnesota Humanities Commission Motheread/Fatheread® Evaluation*. St. Paul, 1995.
[5] Morgan, Marcyliena. *Motheread® Program Evaluation: Goals, Literacy and Training*. Los Angeles, 1994.
[5] Martin, Sandra L., Niki U. Cotton, Browne Dorothy C., Kurz, Branda, and Robertson, Elizabeth. Family violence and depressive symptomatology among incarcerated women. *Journal of Family Violence*, Dec 1995: 399-411.

a manner that 1) included objective, standardized measures of child literacy outcomes, and 2) increased the level of evidence through use of an experimental design.

The study recruited parents from four preschool sites (five schools with two schools combined to make one site) in Morgan County, Colorado (see Table 1 for study participation). Each semester in each preschool, parents were randomly assigned to an intervention or waitlist control condition, and they and their eldest child in preschool were assessed before and after the 12-week MFC program (see Figure 2 for study design for two cohorts as an example). The assessments came from parent surveys (intake and follow-up surveys, including measures of child literacy and parental behavior supportive of reading at home) and preschool teacher ratings of children's literacy and language skills (TS Gold scores). In tandem with the MFC impact study, an implementation study was also conducted to assess MFC program fidelity. This report details the final findings of the study.

Figure 2. MFC Impact Study Design

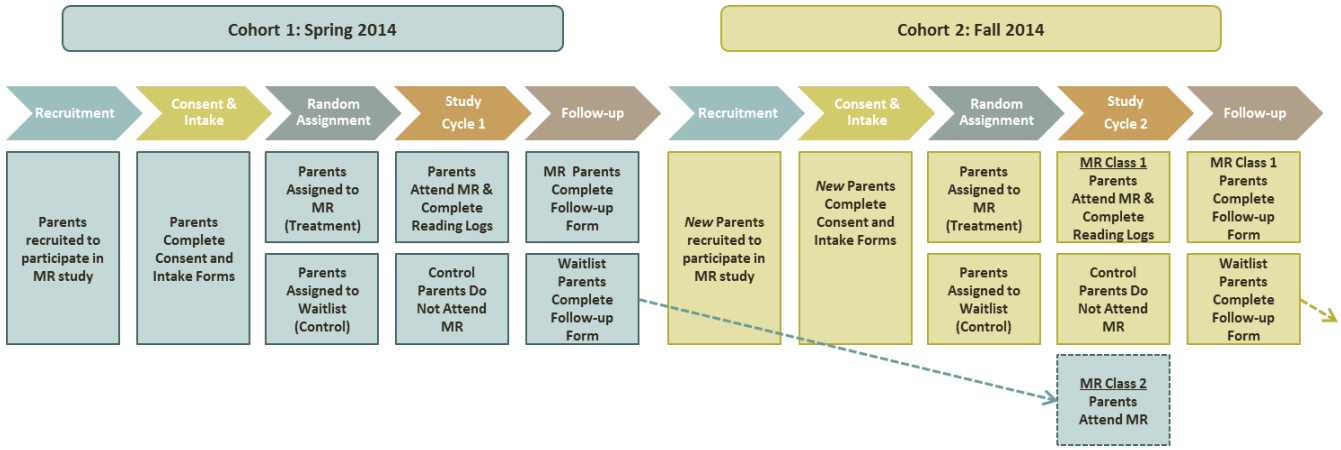


Table 1. MFC Impact Study Participation (Total N with Intervention and Control Group Ns in Parentheses)

Participating Schools	Spring 2014	Fall 2014	Spring 2015	Fall 2015	Spring 2016	Total Study Participants
Site #1 (Wiggins & Weldona)	18 (9, 9)	8 (4, 4)	6 (3, 3)	17 (8, 9)	1 (0, 1)	**50**
Site #2 (Thomson Primary)	-	14 (7, 7)	6 (3, 3)	10 (5, 5)	4 (2, 2)	**34**
Site #3 (Sherman Early Childhood Center)	-	-	22 (11, 11)	20 (10, 10)	12 (6, 6)	**54**
Site #4 (Trinity Lutheran Preschool	-	-	-	-	10 (5, 5)	**10**
Total	18	22	34	47	27	**148**

IMPLEMENTATION STUDY RESEARCH QUESTIONS

The implementation study addresses questions related to program fidelity, participant characteristics, and participant adherence to recommended practices. These questions are addressed through data

collected on participant forms, observations conducted by the research team, and coaching observations made by the MFC Coaches. The research questions are as follows:

1. How many parents and children were reached by the study and what are their demographic characteristics?
2. How much programming did parents receive (i.e., number of sessions)?
3. How frequently were parents reading with children between Motheread sessions?
4. Within each Motheread session, to what degree were all program components covered (i.e., Literacy as Art, Literacy as Experience, Literacy as Process, and Story Extenders)?
5. How many coaching sessions were delivered to Motheread adult educators and what was their timing with respect to program implementation (e.g., session # during which coaching occurred)?
6. What were the successes, barriers and challenges in delivering Motheread with fidelity?

IMPACT STUDY RESEARCH QUESTIONS

The impact study addresses a set of confirmatory questions about the efficacy of the MFC program on children's literacy and language outcomes and mediation of those outcomes via parents' outcomes. A set of exploratory questions further evaluate the impact of MFC, although they do not allow for confirmation from randomized comparisons.

Confirmatory Research Questions

1. To what extent do children exposed to Motheread through their parents' participation demonstrate greater gains in literacy and language outcomes than similar children whose parents did not participate in the program?
2. To what extent do parents participating in Motheread demonstrate behaviors more supportive of reading than similar parents not participating in the program?
3. To what extent do changes in reading behaviors of parents participating in Motheread mediate changes in literacy outcomes of children exposed to Motheread?

Exploratory Research Questions

4. To what extent does the exposure to Motheread benefit children who begin with low literacy and a poor literacy home environment more than other children?
5. To what extent does the amount of parents' participation in Motheread (i.e., number of sessions) correlate with gains in children's literacy?
6. To what extent are posttest improvements in Motheread children maintained over the next months and year?

Methods

DATA COLLECTION TOOLS

Questions for the implementation and impact studies are addressed using a variety of data collection tools and measures. The data collection tools used for the impact and implementation studies include the following:

- ***Participant Intake Forms***. The Participant Intake Form is used to gather demographic information about participating parents and their children. Additionally, this form asks parents to describe their current practices with respect to reading in the home and to obtain a baseline measure of parent reading behaviors and children's literacy skills in the areas of reading, language, knowledge of letters, writing, and response to print.
- ***Participant Follow-up Forms***. The Participant Follow-up form includes a post-measure of the parent reading behaviors and children's literacy skills assessed at intake. The form also asks about exposure to other literacy programs over the past 12 weeks and, for parents in MFC, satisfaction with the program.
- ***Parent Reading Logs***. The Parent Reading Logs document the amount of home reading using MFC books with children, non-MFC books with children, and home reading without children.
- ***Attendance Logs.*** Attendance Logs are used by adult educators delivering MFC to track the name and number of parents attending each of the 12 MFC class sessions.
- ***Educator Session Logs***. To support monitoring of implementation fidelity, the Educator Session Log is used by adult educators to provide information on whether the fundamental components of MFC (e.g., Literacy as Art, Experience, and Process) are delivered at each session.
- ***Observation Inventory (3 per adult educator)***. The Observation Inventory instrument includes similar items to the Educator Session Logs. Thus, coaches and the research team observe for the same aspects of program fidelity during their coaching visits as adult educators track on the session logs.
- ***Coaching Session Feedback Reports (3 per adult educator)***. Coaching Feedback reports utilize a standardized structure and provide documentation of the specific activities occurring during coaching sessions, as well as support recommendations offered by MFC.

OUTCOME MEASURES

Implementation research questions are addressed through information obtained on Participant Intake and Follow-up Forms, Educator Session Logs, and Observation Inventories. Impact research questions are addressed through three key measures, described below.

Child-Level Outcomes: Teacher Reports on Child (Teaching Strategies Gold)

Teaching Strategies (TS) Gold is a comprehensive observational measure that is designed to assess child development and learning from birth to kindergarten (Kim & Smith, 2010) and has been validated to predict later school success (Heroman, Burts, Berke, & Bickart, 2010). When the study was initiated, TS Gold was the only approved assessment of school readiness for the State of Colorado. [6] Teachers rate children on a 10-point scale across nine dimensions. For our purposes, the two relevant dimensions are language and literacy. Two other dimensions on English-language acquisition are relevant to English-language learners. The other five dimensions of socio-emotional development, physical development, cognitive development, science and technology, and mathematics knowledge are less relevant. Research has shown that the scale has adequate internal consistency reliability and construct validity[7, 8] and defines national norms to standardize the scales.[9] Of special value, teachers completing the TS Gold assessment are required to complete the assessment on all preschool children and upload data into the TS Gold tracking system as a part of school requirements. The teachers of children in the impact study will not be fully blinded, as they may learn informally from parents who is attending the program sessions. Yet they will have little stake in the program outcome or motivation to favor program children in their assessments.

In alignment with confirmatory impact questions #1 and #3, children in the program are predicted to increase their scores on the literacy and language domains to a greater degree than children in the control group. Our tests of the confirmatory questions thus rely on the TS Gold subscales that directly assess language and literacy gains.

Based on the TS Gold Technical Manual, both subscales of the TS Gold demonstrate strong reliability results.[10] Specifically:

- *Language Domain:* "Based on the Rasch reliability indexes, the scale appears to be highly reliable, as evidenced by person separation indexes of 7.09, person reliabilities of .98, item

---

[6] http://www.boarddocs.com/co/cde/Board.nsf/files/922PT7661E68/$file/SR%20presentation%20Nov%202012%20as%20of%2011-12-12%20final%20corrected.pdf

[7] Kim, D. & Smith, J. (2010). Evaluation of Two Observational Assessment Systems for Children's Development and Learning, NHSA Dialog, 13, 253 – 267.

[8] Lambert, R.G, Kim, D.H., Taylor, H., & McGee, J. (2010). Technical Manual for the Teaching Strategies GOLD™ Assessment System. Charlotte, NC: Center for Educational Measurement and Evaluation.

[9] Lambert, R.G. (2012). Growth Norms for the Teaching Strategies GOLD Assessment System. Charlotte, NC: Center for Educational Measurement and Evaluation.

[10] Lambert R., Do-Hong Kim, and Diane C. Burts. (2013). *Technical Manual (2nd Edition) for the Teaching Strategies GOLD Assessment System*. Center for Educational Measurement and Evaluation (CEME). Accessed here: http://ceme.uncc.edu/ceme-technical-reports.

separation indexes of 80.86, and item reliabilities of 1.00. The Cronbach's alpha reliability coefficient for this scale was .98, indicating high internal consistency reliability." (pg. 14)

- *Literacy Domain:* "Based on the Rasch reliability indexes, the scale appears to be highly reliable, as evidenced by person separation indexes of 4.90, person reliabilities of .96, item separation indexes of 69.73, and item reliabilities of 1.00. The Cronbach's alpha reliability coefficient for this scale was .98, indicating high internal consistency reliability." (pg. 14-15)

The scores for the language and literacy domains can also be collapsed into two categories: below the range of values considered appropriate for that age and at or above the values considered appropriate.

Child-Level Outcomes: Parent Reports on Child

To form a robust assessment of child literacy, parent-report measures of child-literacy skills are also included. These measures are used along with TS Gold to test confirmatory questions #1 and #3 (*"To what extent do children exposed to Motheread through their parents' participation demonstrate greater gains in literacy and language outcomes than similar children whose parents did not participate in the program?"* and, *"To what extent do changes in reading behaviors of parents participating in Motheread mediate changes in literacy outcomes of children exposed to Motheread?"*).

Five subscales from the Early Literacy Parent Questionnaire[11] are included on the intake and follow-up forms. The subscales, which are based on items offering five response categories ranging from not currently or never (1) to daily or several times per day (5), represent the following dimensions:

- **Reading Scale**: Five items on interaction around books that ask about the child's active involvement in reading (e.g., asking questions about stories, pointing to or asking about pictures, filling in words and lines). (Cronbach's alpha = .71)
- **Language Awareness Scale:** Four items on phonological awareness and rhyming skills that ask about the child playing rhyming games, telling nursery rhymes, and noticing rhymes. (Cronbach's alpha = .82)
- **Interest in Letters Scale:** Three items on knowledge of letters and sounds that ask about the child naming letters, attempting to make sounds from letters, and identifying some letters of the alphabet. (Cronbach's alpha = .80)
- **Writing Scale:** Five items on writing that ask about the child drawing, writing letters, writing words, and asking for help in writing. (Cronbach's alpha = .78)
- **Response to Print Scale:** Two items on response to print that ask about identifying words on signs and packaging and reading words on sight. (Cronbach's alpha = .64)

[11] Boudreau, D. (2005). Use of a parent questionnaire in emergent and early literacy assessment of preschool children. *Language, Speech & Hearing Services in Schools,* 36, 1-33.

The Early Literacy Parent Questionnaire measure developer reports strong relationships between formal assessments of early literacy skills (e.g., letter names, letter sounds, productive rhyme) and the parent-reported phonological awareness, response to print in the environment, alphabet knowledge, and orientation to literacy scales for a sample of children with language impairments. Despite limited validity testing of the tool, we determined that the Early Literacy Parent Questionnaire would be the best supplemental parent-report measure of child literacy for this study after an extensive review of available parent-report measures. The scales demonstrate good reliability, have strong relevance to the study population, tight alignment with the outcomes of interest and the Teaching Strategies Gold scales, and administration brevity in the context of other parent-report study measures.

Parent Behaviors

To measure parent behavior, eight questions from the Parent Survey of Home Literacy[12] were used to gauge the use of interactive book reading techniques by parents. The eight items reflect parent behaviors supportive of reading such as frequently pointing out objects, illustrations, and letters; discussion of what will happen next in the story; and encouragement of co-reading when there are repeated or rhyming phrases. Participants select which behaviors they engage in with their child. Parental behaviors reflect the key outcome to be tested in confirmatory question # 2 (impact on parents' behaviors) and to serve as the mediating variables in confirmatory question #3 (parents' behaviors leading to children's gains in literacy outcomes).

PROCEDURES

Data Collection

Participants were recruited to be in the study through newsletters to parents, home visits, conversations with preschool teachers, conversations with parents previously enrolled in the study, billboard advertising, radio advertising, print advertising, and through "demo" sessions that gave a brief introduction to the MFC program and the study. Parents who wished to enroll signed an informed consent and completed the Participant Intake Form. Parents were then randomly assigned to the intervention or control group by the research team (procedure discussed below).

Parents who were assigned to the intervention group participated in the 12-week long MFC class and completed a brief Parent Reading Log between MFC sessions. These forms were returned to the MFC educator at each session. Adult educators delivering MFC also completed an Attendance Log and the Educator Session Log documenting the activities completed during each MFC session. Each participating parent (intervention and control) completed a brief follow-up form approximately 12 weeks after study

[12] Smith, S., & Dixon, R. G. (1995). Literacy concept of low- and middle-class four-year-olds entering preschool. *The Journal of Educational Research, 88,* 243-253.

enrollment. MFC participants completed the follow-up form on the last day of class. Parents in the control group were contacted by the preschool teacher and the MFC coordinator to complete the follow-up form.

TS Gold data were requested for the most recent checkpoint prior to the beginning of the class and for the checkpoint most closely following the end of the class. During the last cohort of the study, an additional request for TS Gold data was made for all children in the study that was available up to the time they were in Kindergarten. Preschool staff downloaded relevant TS Gold data from the online TS Gold system and uploaded the data for the research team via a secure FTP server, following a protocol developed by the research team.

Throughout the course of the 12-week MFC session the MFC Coach visited the class three times and documented observations using the Observation Inventory. The MFC Coach also debriefed with the adult educators following each coaching visit. During the debriefing session, the coach documented information gathered from adult educators, including identified reasons for any deviations from the MFC curriculum and/or need for additional support from MFC. Following each coaching visit, the MFC coach developed a written feedback report that was shared with both the implementing adult educators and the program director. Coaching Session Feedback Reports and Site Visit Summaries were documented in electronic, written format and analyzed by the research team to address program implementation research questions.

Research team members also conducted site visits that included observations of the MFC sessions and semi-structured debriefing with MFC adult educators. The research team used the Observation Inventory to ensure the collection of consistent observation data. Site visits by the research team also included technical assistance when needed, typically surrounding TS Gold data collection. Each year of the study, the research team facilitated a more formal “debriefing” session with adult educators, the MFC Coach and Colorado Humanities to discuss the successes and challenges associated with the study. A total of three debriefing sessions were held.

All participant and session-level implementation data collection instruments were built into a customized, secure, web-enabled database configured in the Efforts-to-Outcomes™ (ETO) software platform. To support complete and accurate data collection and entry, the research team provided hardcopy versions of all data collection instruments for use by site staff. Site staff then provided all data collected on hard copies to the MFC coordinator for data entry into ETO.™

Randomization

Parent-child dyads served as the unit of assignment. After being informed of the study, consenting to participate in the program, and completing an intake form that assessed current reading behavior, parents awaited results of the randomization process that determined whether they started participation in the next month or at the next class cycle (i.e., in the subsequent spring or fall school semester). For example, for Cohort 1, parents assigned to the intervention condition participated in MFC in spring 2014 and parents assigned to the waitlist control participated in fall 2014. The OMNI research

team used a blocked randomization procedure to make assignments to intervention or control conditions. Within each preschool site, parent-child dyads were blocked on four key characteristics: child age, child gender, child race/ethnicity, and public or private payment for the preschool. Applying a ratio of 1:1 assignment within blocks, a random number generator assigned parents to the intervention or control group. OMNI informed parents via letter of their start date; preschool staff or adult educators conducted follow-up via phone or in-person at the preschool to remind parents before the first session.

By performing the randomization outside the school, the assignment maintained independence from teachers and parents. Further, by gaining consent to begin immediately or wait several months before the randomization occurs, the assignment limited dropouts among subjects not liking their assignment group. Statistical tests for baseline equivalence, discussed below, provide evidence of the success of the randomization procedure.

Data Preparation

All parent report surveys (intake, follow-up and reading logs) and MFC educator forms (attendance and Educator Session Logs) were examined by the research site coordinator for missingness and out-of-bounds values prior to data entry of the forms into OMNI's data collection (ETO System). The research site coordinator also scanned copies of all forms for the research team to review hard copies should there be an issue with data quality. For each measure on the parent report surveys, scales were created according to measure-developer guidelines (e.g., calculating means, sums, etc.,).

Preschool staff provided raw data from each TS Gold domain to the research team. Data were then restructured and merged with participant-level information from ETO. TS Gold scale scores were calculated from the raw data based on TS Gold standard guidelines.

Imputation of missing data is described below.

Implementation Findings

The purpose of the MFC implementation study was to 1) identify characteristics of program participants and children exposed to the program to ensure services were reaching the target population; 2) track the frequency and quality of services provided to participants to ensure the program was implemented with fidelity; and 3) provide data on program components to be used in the impact study to test associations between specific program components and children's outcomes.

IMPLEMENTATION QUESTION 1: HOW MANY PARENTS AND CHILDREN WERE REACHED BY THE STUDY AND WHAT WERE THEIR DEMOGRAPHIC CHARACTERISTICS?

A total of 148 families enrolled in the MFC Impact Study. Families were randomly assigned to participate in the MFC class intervention condition or the waitlist control condition, with 75 families assigned to intervention group and 73 to the waitlist group.[13]

In brief, 74% of the parents self-identified as White/Caucasian and 70% of the children were identified as White/Caucasian. Nearly half of participants (42% of parents, 52% of children) identified as Hispanic/Latino. Seventy-five percent of the families reported speaking English as a first language at home and 22% reported speaking Spanish. Out of the parents who reported an annual income (n=125), the average annual household income for families enrolled in the study was $43,810 (median = $36,000) and the average number of dependents was four. Sixty-six percent of families qualified for free or reduced-price lunch. Of the parents who reported how much higher education they hoped for their child to obtain (n=143), the majority (92%) of parents indicated that they *wanted* their child to finish college and 80% of parents *expected* that their child would complete college. See Appendices A1 – A11 for responses to each item and additional participant demographics, including the number of missing responses to each item.

IMPLEMENTATION QUESTION 2: HOW MUCH PROGRAMMING DID PARENTS RECEIVE (NUMBER OF SESSIONS)?

The MFC class is designed to be delivered in 12 sessions and Colorado Humanities has a target of participants completing eight or more classes. This target was set based on previous evaluations of Motheread that indicated that parents who received eight or more weeks of the program (at least 16 hours) improved their reading skills by one to four grade levels. [14]

Forty four out of 75 parents (59%) who participated in the intervention attended 8 or more session of the class. There were six parents who attended zero classes (the study employed an intent-to-treat design and these families remained in the intervention group). Classes ranged in length from 60 – 150 minutes. On average, classes lasted 102 minutes. Thus, with parents attending an average of eight classes, participation for each participant was about 816 minutes (13.6 hours) of MFC programming. Programming data were drawn from attendance logs and Educator Session Logs completed by MFC educators. MFC educators completed 89% of Educator Session logs and 100% of logs as planned; there were no missing logs.

---

[13] Randomization of condition assignment was conducted within each preschool, accounting for the slightly unequal distribution of waitlist and control participants.

[14] Measuring Success—Review of Research and Evaluation: http://0347dbd.netsolhost.com/wp-content/uploads/2012/05/MeasuringSuccess.pdf

IMPLEMENTATION QUESTION 3: HOW FREQUENTLY WERE PARENTS READING WITH CHILDREN BETWEEN MOTHEREAD SESSIONS?

Research indicates that parents should aim to read with their children every day for at least 30 minutes.[15] Colorado Humanities has set 30 minutes of reading daily as a target for parents in the MFC program. On average parents in the intervention group read to their children 6 days each week during the weeks of the intervention. Parents read the MFC book on average between 4 and 5 days a week and other books with their children between 4 and 5 days a week. Parents also reported reading on their own between 2 and 3 days a week.

For days that parents reported reading with their children, they reported reading over 25 minutes each day, on average. For days that parents reported reading on their own, they reported reading over 43 minutes each day, on average. Reading between Motheread sessions was measured by Reading Logs completed by parents. A completed Reading Log was required for each session of the program. The response rate for parent reading logs was 67%.

IMPLEMENTATION QUESTION 4: WITHIN EACH MOTHEREAD SESSION, TO WHAT DEGREE WERE ALL PROGRAM COMPONENTS COVERED (I.E., LITERACY AS ART, LITERACY AS EXPERIENCE, LITERACY AS PROCESS, AND STORY EXTENDERS)?

MFC aims to have the three main components of the program – Literacy as Art, Literacy as Experience, and Literacy as Process – covered in every Mothered session. The researchers and MFC Coaches reported that Literacy as Experience and Literacy as Process were covered in-depth in 100% of MFC observational sessions (n=42). Literacy as Art was covered in all but one observational session.

Story Extenders are an optional supplement to the program that allows parents the opportunity to practice reading skills with their children with the support of adult educators. Researchers reported that the Story Extenders portion of the curriculum was covered in 66% of MFC observational sessions. Finding time to complete the Story Extenders portion of the curriculum was a challenge that was noted early in the study and continued to be a challenge throughout the course of the study. However, because Story Extenders are optional, implementation findings indicate that the program was delivered with fidelity to the core components of the curriculum and that Story Extenders were used to supplement the curriculum about two-thirds of the time.

[15] Parents as Teachers (1999). *Start Early, Finish Strong: How to Help Every Child Become a Reader.* Accessed here: http://www2.ed.gov/pubs/startearly/ch_1.html.

IMPLEMENTATION QUESTION 5: HOW MANY COACHING SESSIONS WERE DELIVERED TO MOTHEREAD ADULT EDUCATORS AND WHAT WAS THEIR TIMING WITH RESPECT TO PROGRAM IMPLEMENTATION (E.G., SESSION # DURING WHICH COACHING OCCURRED)?

The fifth implementation question is specific to the delivery of MFC Coaching, which helps MFC implementers solidify the techniques learned during MFC Institutes and to deliver MFC with high fidelity. The coaching model is designed with a target of three coaching sessions for each new adult educator implementing the program and aims to follow best-practices from other early childhood program delivery coaching models, such as the Head Start CARES program.[16] A total of 18 coaching sessions were delivered across the first four cohorts. There were no coaching sessions conducted during cohort five since there were no new adult educators. Each new educator received a total of three coaching sessions, strategically planned throughout the MFC program such that one coaching session was offered at the beginning, middle and end of the 12-week course. The MFC Coaches used the coaching and observation tool during each session to provide data on the session and offer feedback to adult educators. The coach also spent time debriefing with educators and offered suggestions to educators following each observation session.

IMPLEMENTATION QUESTION 6: WHAT WERE THE SUCCESSES, BARRIERS AND CHALLENGES IN DELIVERING MOTHEREAD WITH FIDELITY?

To identify successes and challenges with implementation, OMNI conducted an analysis of data gathered during the class observations. A total of seven individuals from OMNI and MFC conducted 42 observations between March 2014 and April 2016. There were 16 observations conducted at Wiggins Preschool, 13 at Thomson Primary School, 11 at Sherman Early Childhood Center, and two at Trinity Lutheran Preschool. Of the total of 287 parents participating in the MFC program, 229 class participants were identified as female and 58 class participants were identified as male.

Successes

Overall, adult educators, the MFC coach, and the research team determined that MFC was delivered with high fidelity. The observations and educator session logs revealed that facilitators were successful at covering all three key components of the MFC curriculum. Indicators of successful program implementation for each key program component, as well as outcomes on these indicators are provided below:

[16] Lloyd, C.M & Modlin, E.L. (2012). Coaching as a Key Component in Teachers' Professional Development: Improving Classroom Practices in Head Start Settings. Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services. Accessed here: http://www.mdrc.org/sites/default/files/full_60.pdf

Literacy as Experience builds literacy skills and helps adults connect their life experiences to stories. This component includes saying the title of the book and the author and reading the book with the group. These activities were covered in all observed sessions. [17]

Literacy as Art includes dramatic reading, acting out the story, or using props or material to enhance story sharing between adults and children. These activities were covered in nearly all (41 out of 42) of the observed sessions.

Literacy as Process is designed to connect the theme of the MFC book to other literature and to build literacy skills. Literacy as process can include reading, interpreting poetry or building specific reading or writing skills through writing and drawing activities. These activities were noted in almost all the observed sessions (41 out of 42). Parents were often encouraged to write in their journals and share their entries.

Program Strengths

Session observers noted numerous strengths of the MFC program. The most commonly observed strengths included the following: the sense of a safe environment and community within the session classes, strong parent engagement, the development of self-improvement and personal growth, and overall program satisfaction reported by parents. Each of these factors are discussed in greater detail, with quotes illustrating these factors below.

Safe Environment/Sense of Community

Observers commonly referenced that a safe environment and sense of community were established in the Motheread/Fatheread classes. Observers attributed various factors to the development of this comfortable environment, including the facilitators, the tactful use of humor to build relationships between participants, the activities within the sessions designed to elicit participant engagement, and the specific topic content, which encouraged participants to engage in vulnerable, trust-building conversations.

The facilitators employed various skills critical for establishing a comfortable environment for program participants. Facilitators were described as skilled, well-prepared, and able to work well with other facilitators and participants. Observers noted that lesson plans were thought-out and clearly delivered, that facilitators effectively managed session discussions through measured opportunities to share experiences and encourage active listening, and that facilitators appropriately balanced participant contributions to ensure all participants were engaged. Furthermore, the facilitators were sensitive to language preferences and abilities within the sessions.

[17] The 2015 report indicated that, in one session, Literacy as Experience was not covered. This was a reporting error and the information provided in this final report is accurate.

Observers noted that many parents shared openly and respectfully listened to others regarding issues that helped them connect their everyday lives with the stories they were sharing with their children. Topics parents discussed included family values/traditions, memories, fears, relationships, cultural norms, dreams/goals, and personal choices. More specifically, some discussions were about bullying and prejudice, gender identity, learning disabilities, and commonly shared parental issues such as divorce/separation. Combined, these factors all contributed to fostering a safe environment and sense of community among program participants, as indicated in the quotes below.

"It is very clear that a safe environment has been created by the facilitators and participants as some very personal concerns and stories were shared. Participants discussed their concerns about keeping their children safe but yet prepared for what the world might throw at them. Stories were shared about personal experiences with bullying."

"There was a great deal of laughter throughout the night as well as several very serious, meaningful discussions. A safe and confidential environment has been created that provides a space for all these things to happen…[The parents] expressed how meaningful these Motheread sessions have become and their desire to be able to share that with other potential participants."

"[The facilitators] took turns asking questions and responding to answers that were shared. They listened well, and were able to connect with the participants. A great deal of humor was used throughout the conversation. Various levels of questions were posed, creating a rich discussion."

"All the facilitators employed active listening skills with great eye contact, nodding heads, reflecting on participant's comments, building on participant's comments and sharing laughter."

"This team has shown great growth throughout the series of Motheread sessions. They are well-prepared, demonstrating the thought that has gone into preparing for the lesson."

"This team work[s] well together. The responsibilities were evenly divided and I appreciate the efforts that go [into] providing materials in both English and Spanish. The lesson was well-planned and that showed in the delivery throughout the evening. All facilitators did a nice job of making eye contact with all participants and employing active listening strategies."

"It was great to observe their comradery and newly found friendships, they all seemed to bond immensely in the sessions – perhaps because of the icebreaker and sharing traumatic experiences, it opened everyone up and set an environment of trust."

"[I think] everyone open[ed] up right away during the icebreaker sharing many personal trials and tribulations. It was clear that all participants knew one another and furthermore were very comfortable with one another. The icebreaker questions were very silly for the most part - but this gave participants a chance to laugh."

Parent Engagement

Parent engagement was a strong theme throughout the observations. Most parents consistently participated in all discussions and activities. Observers documented strong engagement when parents had a willingness to share personal stories, actively listened to others' stories, asked insightful questions, and were actively involved in session activities. Observers attributed strong parent engagement to a variety of factors, including strong facilitation and the content of topics covered.

"Parents were asking each other questions about their experiences with the book-- great engagement!"

"Facilitators did a great job in calling on the quieter people to ensure everyone contributed. They asked thought-provoking questions and everyone contributed responses."

"Though some of the topics of conversation were sad – participants shared a lot of laughs and were very engaged with one another. You could feel the positive energy in the room."

"[The parents] then acted out Flossie and the Fox, and recorded it on an iPad after rehearsing one time with the plan to show the video to their kids. The parents LOVED it. They volunteered for all of the parts, joked, and laughed. Everyone was a great sport about it."

"The participants let their imaginations run wild and creative writing skills shine; many participants wrote stories of the day their child was born…and the love they have for their children. It was an extremely heartwarming experience, shared by all."

Self-Improvement

Self-improvement and personal growth and development by both the facilitators and program participants was another theme from the observations. Observers noted personal growth and skill development in three core aspects of the program. First, researchers observed how the facilitators positively responded to feedback generated throughout the course of the program and implemented recommendations to improve the facilitation process in subsequent sessions. Second, observers discerned that quieter, more reserved participants increasingly contributed as they attended more sessions. Finally, during sessions, parents reflected on how the skills and experiences they acquired through the program impacted their own personal development and their relationships with their children.

The facilitators implemented several of the suggestions that were provided after the last observation. There was a noticeable difference on the level of participation by all of the participants because of creating two smaller groups. This allowed for full participation. The amount of questions asked was adjusted, but the questions that connected the story with the participant's lives were maintained. In the post-observation discussion, the facilitators reported that they too can see that all participants are more engaged. Nice job on continuing to grow and develop as Motheread facilitators!

There has been growth in all participants, even for the more reserved ones, in their level of participation.

"One very quiet woman spoke a few times, for the first time, and they were quite thrilled that she is finally opening up."

"One participant commented that he 'doesn't do that' referring to dramatic reading. Other participants and the facilitators encouraged him by mentioning how he has grown in his skills. He was hesitant to read aloud at the beginning of the Motheread session and now reads with expression.

"[One] participant self-identified growth he has experienced in reading with his child because of participation in the Motheread sessions.

"One participant shared that learning about themes has helped her relationship with her kids, she 'not only reads to them, but she continues the discussion after the book is read.'

Program Satisfaction

The creation of a safe, comfortable environment, the high level of parent engagement, and the degree of personal development all contributed to overall program satisfaction. Observers noted that parents were satisfied with the following aspects of the program: using books with both English and Spanish translations; applying skills learned in the sessions to engage their children in reading; and sequencing their children's exposure to the stories in a way that maximizes their attention, involvement, and learning.

"One family reported that they loved the books that were English and Spanish, because their Grandma could participate by reading the Spanish to her grandchildren."

"Parents expressed the use of props was helpful, and children acting out noises was fun."

"Participants have reported how valuable the Motheread sessions have been to their family. One mom reported that her two young girls wanted to go to Motheread instead of going to school. Another father reported that he is reading to his two daughters for the first time."

"One parent commented that these activities help parents "practice their imagination" and "it helps us to relate to our kids since they get to be creative at school so it gives us chance to get creative." Parents talked about how hard it is to do activities like this with their kids at home."

"The participants expressed that the children enjoyed hearing the story while they were with their teachers and that being exposed to the story allowed them to have meaningful discussions in the car on the way home. They viewed this as being more effective than in taking the time to read together at the end of the night when the children were tired and distracted."

Barriers & Challenges

Although observers most commonly encountered program successes, three challenges were observed throughout the MFC sessions. These include the format and structure of session closings; the degree to which story extenders are incorporated into reading time with children; and the consistent attendance to the MFC program sessions. This section illustrates the dimensions of each challenge.

Session Close/Class Wrap-up

Observers routinely identified that session endings could be improved by taking time to reiterate learned skills and experiences. Observers commonly reported that sessions ended abruptly and that there often was no intentional class ending. Many observers posited that the lack of structured session closing resulted in lost opportunities for parents to reflect on the skills learned. Most observers attributed the weak session endings as a consequence of the time the sessions ended or to facilitators sensing the need to end quickly due to time constraints and other content covered earlier in the session.

"Closing was clear and brief, but did not give participants a chance to reflect."

"There wasn't much of a closing, more like a transition for the families to go read with their children."

"The closing was brief, but with that it also felt unfinished. The theme was never readdressed or discussed, there was no full circle established with the discussion, it all just kind of ended."

"The closing was basically transitioning to the activity with their kids but once everyone got their kid, it was past 8 and everyone decided to go home."

"At the end of class [the facilitators] reminded students about the importance of reading daily, handed out the two younger books and materials. They didn't really do formal closing. This reinforcement of the story and the Child and Family Development theme is really important, so they will work on a stronger but simple closing for next time."

Story Extenders

While the use of story extenders was identified as a useful aspect of the MFC program, observers found that some parents commented that it is difficult to integrate them into reading time with their children. One challenge with story extenders was the inattentiveness of the children, which parents related to the time of day that the program took place (in the evenings) as well as to the length of the program and when story extenders were covered within the sessions. The other challenge observers found was time constraints within the sessions to allow for the inclusion of story extenders in the curriculum. Although not a required component of the program, the omission of story extenders within sessions may have decreased the level of comfort and skill development parents acquired to implement this program aspect at home during story time with their children.

"Another participant mentioned that their child was beginning to read the story on his own using the pictures and many of the Story Extenders were mentioned. When the Story Extenders were distributed [the facilitator] shared with the parents that the children had already discussed one of the questions with their child care providers."

"Parents mentioned that they try doing the [story extenders], but it is usually difficult to get their children to pay attention."

"Parents reported using the story extenders at home - those that mentioned this also shared similar frustrations. They shared that their children often feel tired by that time of night (after sessions) so it is hard to keep them alert and involved when reading with them."

"The class has decided to stop doing the Story Extenders in class, since it is usually hard to get the kids to focus (It's late, and the kids just went over it in the room next door)."

"The story extender was given to the participants but the discussion on this section went no further than that."

"The facilitators reminded the parents about the story extender but that was it."

Attendance

A final challenge noted by observers was consistent attendance to the MFC sessions. School conflicts were observed to be the primary reason why parents missed MFC classes.

"Attendance was low because of a couple of conflicts associated with the school."

"There were only 3 parents in the class because it was parent teacher night and many of the other parents were at parent/teacher conferences."

"Attendance tonight [was] impacted by the Parent Advisory Council which involved two families in the group. One participant joined the group after her meeting for the last twenty minutes of the session."

Feedback from Motheread Coaches

MFC program coaches offered four main areas of constructive feedback for facilitators. First, coaches recommended focusing more on questions that connect the story to the parent's lives rather than the comprehension questions. Second, similarly to the observers, MFC Coaches also suggested that facilitators find an opportunity, especially during the Literacy as Art portion, to discuss the Story Extenders and Early Childhood Supplement books. Third, coaches recommended that facilitators address, clearly delineate, and allow specific time to practice reading and writing skills taught within each lesson. Finally, coaches proposed allotting more time to involve the children in the lesson and to provide parents with supervised opportunities to practice reading with their child.

Impact Findings

The Impact Study uses a randomized controlled trial to test whether participation in MFC leads to changes in parents' reading behaviors and child literacy outcomes. The experimental design is constructed to obtain a 'moderate' level of evidence.

PRETEST EQUIVALENCE ANALYSIS

We first test for baseline or pretest differences across conditions to determine whether the randomization succeeded in making the control and intervention groups equivalent (within the bounds of random chance). Evidence of equivalence would indicate that the random assignment worked as intended and that group differences at the start of the study are unlikely to account for group differences after the program trial. Such evidence is crucial for establishing the internal validity of a randomized controlled trial.

Table 2 lists the following for each pretest measure: the number of cases, minimum and maximum values, the means for the intervention and control groups, the probabilities from difference-of-means t-tests, and the size of group differences in standard deviation units.[18] The measures at pretest include family demographic characteristics, child language and literacy outcomes, and parent behavior outcomes. The difference-of-means t-tests, used for both continuous and dichotomous measures, identify statistically significant deviations from equivalence. The standardized differences (Cohen's d) indicate effect sizes. A common rule of thumb treats a d value of .2 as small, .5 as medium, and .8 as large.

The results in Table 2 suggest that the subjects were randomized successfully. The intervention and control groups do not differ significantly ($p < .05$) or even marginally ($p < .10$) on any of the pretest measures. Only a few d (standardized mean difference) values exceed .20, the common cutoff for a

[18] As some of the measures are not distributed normally, we also checked for differences using non-parametric techniques – chi-square tests for categorical measures and Mann-Whitney tests for continuous measures. These tests give much the same results but generally have less power to identify differences than the parametric tests. To provide more stringent tests for baseline group differences and the success of the randomization, we focus on the t-tests.

small difference, and these differences do not define a consistent pattern. For example, the intervention group is disadvantaged relative to the control group on income per dependent (d = .25) and TS Gold Language (d = .27) but advantaged on parent education (d = -.23).[19] Given the lack of statistical significance and the lack of a pattern consistently favoring one group over the other, we conclude that the pretest group differences result from random error rather than systematic bias. [20]

Table 2. Descriptive Statistics and Tests for Equivalence of Pretest Variables

Pretest Measures	N	Min	Max	Mean [a]		Mean Difference	
		Value	**Value**	**I**	**C**	**p**	**d**
Family Demographic Characteristics							
Child Age	143	3	5	4.00	4.03	.806	.04
Child Female	145	0	1	.54	.51	.677	-.07
Free Lunch (0 = do not qualify, 1 = reduced price lunch, 2 = free lunch)	124	0	2	1.2	1.2	.935	.01
Child Race[21] **(0 = white, 1 = non-White)**	148	0	1	.35	.30	.559	-.10
Child Hispanic (0 = non-Hispanic, 1 = Hispanic)	148	0	1	.49	.40	.243	-.19
Language (0 = Spanish or Other, 1 = English)	148	0	1	.72	.78	.396	.14
# Children	146	1	7	2.47	2.28	.319	-.17

[19] The d value is based on the control group mean minus the intervention group mean. A positive d means a higher control group mean and a negative d means a higher intervention group mean.
[20] Although condition differences in Hispanic ethnicity of the child are not statistically significant, they are relatively large (d = -.19) and potentially important for the outcomes. Also, since this variable proved important in previous reports, we continue to use it as a control in the models to follow.
[21] Child Race and Child Hispanic were coded as white and non-white and Hispanic and Non-Hispanic to reduce missing data.

Pretest Measures	N	Min	Max	Mean [a]		Mean Difference	
Family Income	125	750	100,000	42,083	45,096	.539	.11
# Dependents	139	1	9	4.24	4.12	.609	-.09
Income / Dependents	125	250	33,333	9,787	11,480	.167	.25
Parents Education (1 = no school, 2 = some school, 3 = high school, 4 = some college or technical school, 5 = college degree, 6 = post college degree)	136	1	6	4.06	3.81	.174	-.23
Child Educational Goals (0 = finish some schooling before college, 1 = finish college)	143	0	1	.91	.94	.441	.14
Child Education Expectation (0 = finish some schooling before college, 1 = finish college)	142	0	1	.78	.82	.554	.10
Child-Level Outcomes: Teacher Reports on Child (TS Gold)							
Language Domain (0=below age-level, 1 = at age level, 2 = above age level)	83	0	2	.56	.71	.228	.27
Literacy Domain (0=below age-level, 1 = at age level, 2 = above age level)	82	0	2	.66	.66	1.000	.00
Language Score	83	15	63	38.9	39.8	.689	.09
Literacy Score	82	9	78	32.9	33.0	.958	.01
Child-Level Outcomes: Parent-Reports on Child							
Reading Scale	147	1.00	5.00	3.61	3.64	.864	.03

Pretest Measures	N	Min	Max	Mean [a]		Mean Difference	
Language Scale	147	1.00	5.00	2.53	2.50	.878	-.03
Letters Scale	147	1.00	5.00	3.33	3.42	.637	.08
Writing Scale	147	1.00	5.00	2.72	2.54	.223	-.20
Print Scale	147	1.00	5.00	2.44	2.49	.816	.04
Parent Behaviors							
How Often Read Aloud	142	1	5	3.35	3.35	.993	.00
Minutes Read Aloud	126	1	500	178	195	.513	.12
Designated Time for Reading	131	0	1	.58	.49	.330	-.17
Books Read at One Setting	129	1	6	2.16	2.12	.839	-.04
Active Reading Scale – Sum	148	0	7	3.17	2.90	.379	-.14

I = Intervention Group, C = Control Group

MISSING DATA AND ATTRITION

Introduction

Even with evidence of successful randomization of subjects to intervention and control groups, loss of subjects can bias tests for group differences in the posttest outcomes. If the dropouts vary systematically across conditions, attrition can lead to misleading conclusions about program effectiveness. Through five cohorts of data gathering, our study has a maximum sample size of 148 subjects and, as is appropriate for an intent-to-treat analysis, we seek to include all subjects in their originally assigned group. However, the actual number of cases with valid data is reduced by 1) missing pretest scores on the teacher-rated TS Gold measures; 2) attrition from refusal or inability to complete the posttest survey or 3) missing data from not answering specific survey questions or inaccessible data. We therefore need to test for patterns in the loss of cases due to attrition and missing data.

Our attrition analysis consists of three steps. First, we use Little's global test to check whether data are missing completely at random or reflect a systematic pattern in observed subject characteristics.[22] Since Little's test has low power and is susceptible to Type II errors,[23] additional tests are helpful. Second, to check for non-random missing data in individual variables, we construct indicators of missingness (1 = missing, 0 = not missing) and then correlate the missingness indicators with each of the pretest measures. High correlations indicate that subjects with certain characteristics are more likely to be missing data or drop out. While suggestive of a problem, the test does not necessarily identify bias in comparisons across conditions. Third, for each pretest measure that is significantly correlated with missingness, the indicators of missingness are regressed on the pretest measure, an indicator of condition, and the interaction of the pretest measure-by- condition. A significant interaction term indicates that the pattern of attrition differs by condition, may compromise the randomization, and possibly biases tests for intervention effects. For example, if those prone to do poorly on the posttest outcomes drop out of the intervention group more than the control group, it would exaggerate the program effect. If, however, those prone to do poorly on the posttest dropout at the same rate for the intervention and control groups, it would not bias the program effect. A significant interaction term indicates the former, while an insignificant interaction term indicates the latter.

Pretest Missing Data

Most pretest variables have missing data but seldom more than 10% of all cases. Those measures with more substantial missing data include eligibility for free and reduced-price lunch (N = 124 or 84%), family income and family income per dependent (N = 125 or 84%), TS Gold language and literacy[24] (N = 82-83 or 55-56%), and several parent reading behaviors (N = 126-131 or 85-89%).

Little's global test indicates that the pretest data are missing completely at random. Using all the variables listed in Table 2, the chi-square value of 826.5 (df = 825, p = .479) cannot reject the null hypothesis that the data are missing completely at random. This means that the pretest missing data do not reflect any systematic pattern.

Given their importance to the study, the pretest TS Gold measures of language and literacy deserve special analysis, however. We correlated each of the other pretest measures with a missingness indicator for the TS Gold pretest measures (missing = 1, not missing = 0). The correlations are listed in the first column of Table 3. To be liberal in identifying possible bias, we use a significant level of .10 for these tests. Older children, children in large families, and children scoring higher on parent-reported letters and writing skills are significantly less likely to be missing data. However, if the missing data on these pretest measures are patterned by condition, they should differ across the intervention and control group. Since age, number of children, letters score, and writing score are distributed equally

---

[22] Little, R. J. A. (1988). A test of missing completely at random for multivariate data with missing values. *Journal of the American Statistical Association, 83*(404), 1198-1202.

[23] Enders, C. (2010). *Applied missing data analysis*. New York: Guilford.

[24] These children were not assessed the prior spring because they were new to the preschool (either three year olds entering preschool for the first time or new to the school district). Due to the study design, it was anticipated that TS Gold pretest data would be unavailable for a subset of participants in Cohorts 2 and 4, which both begin in the fall.

across conditions in Table 2, the correlations with missing data should not bias the tests for program impact. The more precise test of this inference uses the interaction of condition by each of the four measures. These results are listed in the fourth column of Table 3. None of the interaction coefficients are statistically significant. Thus, the more detailed analysis of missing data confirms the Little test in failing to identify a pattern of potential bias in missing pretest data on the TS Gold measures.

Based on these findings, we can impute pretest TS Gold data using the expectation maximization algorithm, a method appropriate for data missing completely at random.[25] The impact analysis to follow will examine outcomes for the subset of cases with missing data dropped and for all cases with missing data imputed.

Posttest Attrition and Missing Data

The posttest data face two problems with attrition and missing data: 1) 22 of 148 parents (14.9%) did not complete the posttest survey, and 2) up to 18 additional parents (12.1%) completed the survey but did not answer all the questions, mostly notably the question about parent reading behaviors.

Using the same strategy as for the pretest missing data, our attrition analysis consists of three steps. First, when testing all posttest outcomes plus condition together, the chi-square value of 916.7 (df = 862, p = .096) allows rejection of the null hypothesis at a marginal probability level ($p < .10$) that the data are missing completely at random. However, when limiting the sample to those who completed at least part of the survey and were not subject to attrition (N = 126), the chi-square value of 764.4 (df = 741, p = .268) is clearly not significant and indicates that item missing data are missing at random.

Second, columns 2 and 3 of Table 3 list the bivariate correlations from the differential attrition analysis.[26] The results indicate that several pretest characteristics are associated with missing data. In regard to attrition (N = 22), parents not completing the posttest survey have Hispanic children eligible for free lunch, are less likely to designate a time to read, and score lower on the active reading scale. In regard to missing items (N = 18), parents not answering all questions are more likely to have Hispanic children, have children with better writing skills, read less often to their child, and lack a designated time for reading.

Third, the last columns of Table 3 test for the interaction of condition and pretest characteristics associated with posttest attrition. The table lists coefficients for the interaction terms and associated significance. Only one of the tests reaches statistical significance. For designated reading time, the significant negative coefficient indicates that parents with a designated reading time and in the intervention group are less likely to be missing data. The evidence of differential attrition is modest and

[25] Schafer, J. L. & Olsen, M. K. (1998). Multiple imputation for multivariate missing-data problems: A data analyst's perspective. *Multivariate Behavioral Research, 33*, 545-571.

[26] We use a .10 criterion in Table 3 to help identify any possible evidence of differential attrition but otherwise use the usual levels of significance (.05, .01, .001) when testing for program effects.

the pattern of missingness would likely inflate the control group outcomes and hide condition differences. Still, care must be taken in analyzing the results for this measure of parent behavior.[27]

In summary, the attrition analysis is generally reassuring about the potential for missing data to distort the original randomization. Attrition is more likely among subjects with certain characteristics but generally similar across conditions and the data overall appear to be missing completely at random. To further check the results, we also present figures for all cases with imputed data from the expectation maximization estimation.

[27] To be cautious, we will check the results for parent behavior measures with controls for baseline designated reading time – the measure associated with significant differential attrition.

Table 3. Attrition Analysis of Missing Values on Three Sets of Measures

	Correlation (r) with Missing Values [a]						Regression (b) with Missing Values [b]					
Pretest Measures	Pretest TS Gold Missing N = 66		Posttest Survey Missing N = 22		Posttest Reading Missing N = 18		Pretest TS Gold Missing N = 66		Posttest Survey Missing N = 22		Posttest Reading Missing N = 18	
Condition	0.02		-0.04		0.08							
Age	-0.27	#	-0.04		0.13		-.08					
Female	0.08		-0.02		0.11							
Free Lunch	-0.13		0.17	#	0.12				.00			
Hispanic	0.10		0.20	#	0.17	#			-.18		.12	
# Children	-0.14	#	0.04		0.09		-.06					
Family Income	0.01		-0.14		-0.13							
# Dependents	-0.11		-0.04		0.01							
Income / Dependents	0.05		-0.13		-0.13							
Teacher Reports on Child												
Language Domain Categ.			0.08		-0.15							
Literacy Domain Categ.			0.01		0.06							
Language Score			.09		.06							
Language Score			.06		.10							
Parent Reports on Child												
Reading Scale	-0.07		-0.13		0.03							
Language Scale	-0.03		0.03		0.03							
Letters Scale	-0.19	#	-0.03		-0.06		.04					
Writing Scale	-0.24	#	0.02		0.23	#	-.01				.02	
Print Scale	-0.13		0.09		-0.08							
Parent Behaviors												
How Often Read Aloud	-0.02		-0.01		-0.17	#					-.06	
Minutes Read Aloud	-0.13		-0.01		-0.07							
Designated Time to Read	-0.00		-0.17	#	-0.24	#			-.10		-.14	#
Books Read at One Setting	0.10		-0.14		0.01							
Active Reading Scale 1	-0.05		-0.23	#	-0.11				.03			

a Bivariate correlation of missing =1 (non-missing = 0) with pretest measures

b Unstandardized regression coefficient of missing =1 (non-missing = 0) with pretest measures-by-condition

# $p < .10$

POWER ANALYSIS FINDINGS

With an analysis sample size ranging from 108 to 126 and a maximum sample size with imputation of 148, we can do a power analysis to quantify the ability to find significant program effects with the available data. Table 4 lists the power of the tests for the three sample sizes when the effect size is small (Cohen's d = .2), medium (d = .5), and large (d =.8). The sample sizes of 108, 126, and 148 are all powered for a medium effect size. The minimal detectible effect size for power = .8 ranges from .47 to .54.

Table 4. Power Analysis

Power of Test by	**Bivariate Test**		
Effect Size	N = 108	N =126	N = 148
0.2 (small)	0.18	0.20	0.23
0.5 (medium)	0.73	0.80	0.86
0.8 (large)	0.98	0.99	0.99
MDES at power = .8	0.54	0.50	0.47

TYPE OF ANALYSIS

The confirmatory research questions in summary form ask about the effect of the program on 1) children's language and literacy, 2) parents' behavior supportive of reading, and 3) the mediation of program impact on children by changes in behavior of parents. Exploratory questions ask about differential program benefits by child risk level at pretest and whether more involvement in the program leads to better outcomes. Corresponding to the random assignment of parents and children to the intervention and control groups, the unit of analysis is the parent-child dyad.

With data at two time points, regression models use the posttest score as the outcome and use intervention status, the pretest outcome score, and one selected covariate, Hispanic ethnicity. Additional time points would allow for use of repeated measures models, but two time points can be handled efficiently with least-squares regression for the outcome measures. With the pretest outcome control, the treatment coefficients reflect differences across conditions in the change from pretest to posttest.

The basic model can be represented as follows:

$$Y_{ik2} = \beta_{0k} + \beta_{1k} * I_i + \beta_{2k} * Y_{ik1} + \beta_{3k} * X_{ik1} + \beta_{ik}, \quad (1)$$

where Y_{ik2} and Y_{ik1} refer to outcome k for individual i at posttest (time 2) and pretest (time 1), I_i refers to a dummy variable for the intervention, X_{ik1} refers to the additional covariate of Hispanic ethnicity, and β_{ik} to a zero mean disturbance term that is normally and independently distributed. As noted earlier,

only one control variable, Hispanic ethnicity, is included in the model.[28] Tests of confirmatory questions #1 and #2 concerning the impact of the intervention come from the direction, size, and significance of the β_{1k} coefficients for the intervention measure. Support for the intervention will show in consistently significant and positive coefficients.

Tests of confirmatory question #3 involve mediation and a set of models estimated with multiple regression. Mediation models typically encompass four tests: 1) the intervention should meaningfully affect the child outcomes; 2) the intervention should meaningfully affect the mediating parent outcomes; 3) the parent outcomes should meaningfully affect the child outcomes; and 4) the overall effect of the intervention on the child outcomes should drop sharply when controlling for the mediating parent outcomes. The combined criteria for mediation imply a significant indirect effect of the intervention on child outcomes via parent outcomes.

Exploratory question #4 involves moderation. We expect that the program will most help those children with language and literacy problems at baseline. Children with already-developed skills and less room to improve may gain less from the program. We check for moderation effects by adding product interaction terms to the model. If the program effect varies with a pretest characteristic, the coefficients for the intervention times the characteristics will be large and consistent in direction.

Exploratory question #5 involves a test among intervention group subjects for the influence on the outcomes of session attendance and program participation. An effective program will show a meaningful difference between parents with low and high participation. We treat this test as exploratory because session attendance and participation were not randomly assigned. The test likely will be confounded by unmeasured characteristics such as motivation, conscientiousness, and commitment to the preschool. Benefits of program attendance and participation will suggest program effectiveness but without the internal validity of a comparison between randomized intervention and control groups.

Exploratory question #6 involves the long-term follow-up of the treatment group using the TS Gold measures of child literacy and language. Because subjects in the waitlist control group have the opportunity to enter treatment after the posttest, the statistical analysis of sustained, long-term effects differs from the analysis of the posttest effects. One key difference is that the intervention variable now has three categories: the treatment group subjects, the control group subjects who joined the treatment after the posttest, and the control group subjects who did not join the treatment after the posttest. Another key difference is that the analysis focuses on the changes from posttest to the long-term follow-up assessments rather than from pretest to posttest. The follow-up period thus measures the time since

[28] We originally proposed to include fixed effects for the pre-randomization blocking variables of cohort and school (4 dummy variables for school and 4 for cohort). However, checks show that neither schools nor cohorts had consistent significant effects on TS Gold or parent ratings. These results indicate that clustering within schools and cohorts is small. Including all the insignificant predictors would do little to improve the quality of the models, but it would reduce the power to identify treatment effects. We therefore have not used the fixed effects in the final models. Also, none of the parental characteristics differed significantly across conditions. Adding these controls similarly would not affect condition differences but may reduce the power of the models.

posttest and can range from 3 months to 15 months, depending on which year the subjects joined the study.

The longitudinal models adjust for these two issues. First, they contrast the treatment group with each of the two control groups. Second, they estimate the change in TS Gold language and literacy scores from the posttest through the follow-up period. The test of long-term program benefits thus comes from comparing the change over the follow-up period for the treatment group with the change for the two control groups (controlling for baseline scores, covariates, and fixed effects). If the change in scores is significantly lower for the two control groups, it will suggest long-term program benefits. More detail on the models and estimation follow in the section on Impact Question 6.

We treat this test as exploratory because the two control groups – one joining the intervention and one not joining the intervention – are no longer randomized. The lack of an exact comparison with subjects matched by randomization limits the internal validity of the over-time comparisons and means the results can be considered only exploratory.

IMPACT QUESTION 1 (CONFIRMATORY): TO WHAT EXTENT DO CHILDREN EXPOSED TO MOTHEREAD THROUGH THEIR PARENTS' PARTICIPATION DEMONSTRATE GREATER GAINS IN LITERACY AND LANGUAGE OUTCOMES THAN SIMILAR CHILDREN WHOSE PARENTS DID NOT PARTICIPATE IN THE PROGRAM?

Table 5 presents the key coefficients needed to test confirmatory question 1 concerning the program impact on child language and literacy. The child measures tap two teacher-reported skills and five parent-reported skills.

For each outcome, Table 5 presents the posttest means for the control and intervention groups, along with the probability that the mean difference could result from random error. These bivariate figures provide helpful descriptive information but do not control for pretest values and covariates. The next columns come from regression equation 1 (listed above). The columns list the regression coefficients for the intervention group adjusted for the controls and the Cohen's d coefficients (or standardized mean difference in the outcome) based on the regression.[29] The associated probability for a test of the

[29] The standardized mean difference can be obtained by standardizing the outcome in the regression model. The Cohen's d values differ from standardized regression coefficients or betas, which standardize both the predictors and outcomes.

coefficient significance is listed as well.[30] The last columns replicate the regression results but with missing data imputed using expectation maximization.[31]

[30] We originally proposed to use the Benjamini-Hochberg adjustment for multiple comparisons. However, the value of reducing the risk of a type I error with adjustments for multiple comparisons of must be balanced with the limitation of increasing the risk of a Type II error. Given the recruitment problems in the study and the much smaller sample than originally projected, our analysis is already underpowered; the Type I risk of rejecting a false null hypothesis is low but the Type II risk of failing to reject a false null hypothesis is high. As the Benjamini-Hochberg adjustment would further raise the risk of a Type II error and weaken the results, we decided it would be best to not use the adjustment.

[31] Given the finding that our data are missing completely at random, expectation maximization has some advantages over multiple imputation. Even for the TS Gold pretest outcome, the item with the most missing data, the missing values are due only to the age of program entrance, which was equalized in the randomization. The simpler expectation maximum estimates are well suited for the analysis of such data. Multiple imputation depends on adequate specification of an imputation model, which ideally includes auxiliary predictors of missingness. Misspecification of the imputation model can lead to biased results (see Sterne JAC, White IR, Carlin JB, Spratt M, Royston P, Kenward MG, Wood AM, Carpenter JR. Multiple imputation for missing data in epidemiological and clinical research: Potential and pitfalls. British Medical Journal, 2009, 338). We lack a strong imputation model and are uncertain if the assumptions needed for multiple imputation are met.

Table 5. Results from Analysis of Intervention on Posttest Child-Level Outcomes

Child-Level Outcomes	Posttest Means [a]		Mean Diff. Sig.[b]	Regression [c] No Imputed Data		Regression [c] Imputed Data [d]	
	I	**C**	**p**	**d**	**p**	**d**	**p**
Teacher Reports on Child (TS Gold) Post-test only: N=143-145							
Language Domain Categories	1.12	1.18	.518	-.07	.666		
Literacy Domain Categories	1.15	1.15	.983	.02	.882		
Language Scores	46.8	49.7	.062	-.27	.100		
Literacy Scores	46.2	48.9	.346	-.12	.463		
Teacher Reports on Child (TS Gold) Controls for Pre-test: N=81-83							
Language Domain Categories	1.27	1.21	.659	.27	.186	.05	.734
Literacy Domain Categories	1.18	1.24	.626	-.07	.730	.02	.926
Language Scores	51.0	51.1	.961	.16	.326	-.13	.368
Literacy Scores	52.4	52.0	.935	.06	.757	-.14	.322
Parent Reports on Child N=117-126							
Reading Scale	3.97	3.69	.053	.35	.011	.38	.005
Language Scale	3.13	2.89	.206	.29	.030	.26	.055
Letters Scale	3.75	3.72	.863	.16	.238	.14	.310
Writing Scale	3.18	3.16	.904	-.10	.500	-.08	.582
Print Scale	3.11	2.62	.044	.38	.011	.40	.004
a I = Intervention Group, C = Control Group							
b Mean Diff. Sig. = p value for mean difference							
c Regression controls for pretest outcome and Hispanic ethnicity							
d N = 148 for models with imputed data							

Given that pretest data are missing for many subjects, two sets of results are presented for the TS Gold, one for the posttest only (N = 143-145) and one for a reduced sample of 82-82 that includes the pretest. None of the tests indicate statistically significant benefits of the intervention. Comparison of the

posttest means without and with controls for pretest scores show small condition differences. The regressions with pretest scores indicate a positive effect of the program with standardized differences ranging from .02 to .27. However, given the sample size, these effects do not reach statistical significance. The model in the last columns that imputes missing TS Gold pretest data shows small and insignificant effects.

The results for parent reports in Table 5, in contrast, reveal several beneficial program effects. Controlling for pretest outcomes and Hispanic ethnicity, the program significantly improves the posttest scales for reading, language and print. The effect sizes are medium-small (d ranges from .29 to .38). The results are similar when imputing missing posttest data for these measures.

In sum, gains in literacy and language as a result of parent program participation are not significant on the teacher reported TS Gold measures, but are significant on 3 of 5 parent reported measures.

IMPACT QUESTION 2 (CONFIRMATORY): TO WHAT EXTENT DO PARENTS PARTICIPATING IN MOTHEREAD DEMONSTRATE BEHAVIORS MORE SUPPORTIVE OF READING THAN SIMILAR PARENTS NOT PARTICIPATING IN THE PROGRAM?

Using the same structure as Table 5, Table 6 presents the key coefficients needed to test confirmatory question 2 concerning the program impact on parent behavior. The parent behavior measures tap self-reported quantity and quality of reading to the child. Note that along with four key individual items on parent behavior, the table lists results for the scale summing seven other parent-behavior items in the survey (i.e., the Active Reading Scale).

The bivariate posttest results demonstrate significant intervention benefits on two outcomes: reading often to their children, and having a designated time for reading. A third, scoring high on the active reading scale, has a marginally significant effect ($p = .055$). For reading aloud and the active reading scale, the effects are significant in the regression models, both without and with imputation of missing data. For having a designated reading time, the effect drops below significance in the regression without imputed data but remains significant in the model with imputed data. Moreover, there was some evidence of a problem with differential attrition for this outcome.[32]

[32] Recall that this baseline measure was differentially related to missing data by condition. The insignificant effect without imputation may stem from the differential attrition, while the significant effect for the imputed data is more reliable because all subjects are included. Still, to be cautious, we view the results for this outcome as mixed.

Table 6. Results from Analysis of Intervention on Posttest Parent Behaviors

Parent Behaviors	Posttest Means [a]		Mean Diff. Sig.[b]	Regression [c] No Imputed Data		Regression [c] Imputed Data [d]	
	I	**C**	**p**	**d**	**p**	**d**	**p**
How Often Read Aloud (N = 121)	3.98	3.56	.008	.56	.000	.53	.000
Minutes Read Aloud (N = 100)	235	228	.817	.12	.524	.20	.192
Designated Time for Reading (N = 102)	.81	.61	.019	.27	.137	.46	.002
Books Read at One Setting (N = 100)	2.39	2.31	.640	.09	.624	.08	.598
Active Reading Scale – Sum (N = 126)	4.45	3.82	.055	.33	.046	.33	.032
a I = Intervention Group, C = Control Group							
b Mean Diff. Sig. = p value for mean difference							
c Regression controls for pretest outcome and child Hispanic ethnicity							
d N = 148 for models with imputed data							

IMPACT QUESTION 3 (CONFIRMATORY): TO WHAT EXTENT DO CHANGES IN READING BEHAVIORS OF PARENTS PARTICIPATING IN MOTHEREAD MEDIATE CHANGES IN LITERACY OUTCOMES OF CHILDREN EXPOSED TO MOTHEREAD?

Confirmatory question 3 posits that changes in the behavior of parents mediate the intervention effect on children. However, the small effects found for the teacher-rated TS Gold child reading outcomes limit the value of a mediation analysis. Without a strong relationship between the intervention and the ultimate outcome, there is little to mediate and the evidence of mediation is necessarily weak. We therefore do not report tests for the TS Gold outcomes. However, the results for several of the parent-reported measures of child reading skills do show a significant association with the intervention and are good candidates for analysis of the mediation effects of parent behavior on child outcomes.

To simplify tests for the mediation analysis, we create a summary scale for the child outcomes. The scale takes the mean from the five parent-rated measures listed in Table 5 (language, reading, letters, writing print); it has an alpha reliability of .81. This combined child literacy scale will serve as the ultimate outcome. For the mediators, Table 6 shows significant program effects on two parent behaviors – how

often they read aloud to the child and the active reading scale. We designate each of these measures as a separate mediator, as they do not form a reliable scale.[33]

Table 7 presents the results from each of the four steps in a mediation analysis (described in the "Type of Analysis" section). The sample uses the 126 cases with available posttest data. As is common in mediation analyses, we present standardized regression coefficients (betas) rather than d coefficients.

Step 1 in the mediation analysis regresses the newly created child literacy scale on the intervention (with controls for the pretest outcome and child Hispanic ethnicity). As shown in Table 7, the intervention significantly improves the outcome (as it did for most of the separate items in Table 5). Step 2 summarizes previous results in showing that the intervention significantly increases the read aloud measure and active reading scale; the standardized coefficients equal .280 and .164, both medium-sized effects. Step 3 examines the effects of the posttest read aloud and active reading measures on the posttest child outcome scale. The read aloud measure significantly increases child literacy but active reading does not. Step 4 estimates the effects of the intervention on the child outcome scale with controls for the mediator. The intervention effects should drop significantly from Step 1 (the intervention effect without the control). In Table 7, the intervention coefficient drops to insignificance when controlling for parent reading aloud, but it changes little when controlling for active reading.

A final step computes the indirect effect of the intervention on the child literacy outcome scale via the two mediators. Both indirect effects are small and not statistically significant. For the read aloud measure, there is some evidence of mediation. It meets all four criteria for mediation but has a small mediation effect that is only marginally significant ($p = .077$). For the active reading scale, it fails to significantly improve the child outcome scale and to have a significant mediation effect.

[33] Using the other parent behavior measures in Table 6 also does not form a strong scale, and these other measures have the limitation of substantial missing data.

Table 7. Results of Mediation Analysis for Indirect Influence of Program on Child Literacy Scale through Parent Reading Aloud and Active Reading

Outcome	**Child Literacy Scale**			
Mediator	Reading Aloud		Active Reading	
1. Intervention: Outcome	.141	*	.141	*
2. Intervention: Mediator	.280	**	.164	*
3. Mediator: Outcome	.128	*	.044	
4. Intervention: Outcome (Net)	.114		.137	*
Indirect Effect	.036		.007	

Note: Standardized coefficients listed, * p < .05, ** p < .01

IMPACT QUESTION 4 (EXPLORATORY): TO WHAT EXTENT DOES THE EXPOSURE TO MOTHEREAD BENEFIT CHILDREN WHO BEGIN WITH LOW LITERACY AND A POOR LITERACY HOME ENVIRONMENT MORE THAN OTHER CHILDREN?

The first exploratory research question asks about differential program benefits by child risk level at pretest. It requires tests for moderation that specify different program effects across levels of the pretest characteristics. The tests involve adding product terms of the intervention indicator times the pretest measure of risk. To be thorough, we examined 10 pretest risk measures: 1) TS Gold language categories, 2) TS Gold literacy categories, 3) a combined scale for the five parent ratings of child reading skills, 4) the active reading scale, 5) eligibility for free and reduced-price lunch, 6) Hispanic ethnicity, 7) English language, 8) family income, 9) parent education, and 10) parent expectation for child education attainment. We then added the 10 interaction terms one at a time to models for the three outcomes of posttest TS Gold language and literacy categories,[34] and the parent-rated child literacy scale combining the five separate parent-rated child measures.

The results show that one of the 30 interaction terms reaches statistical significance. The effect of the program on the parent-rated child literacy scale was significantly stronger (p = .038) for Hispanic children than others. Further, significant or marginally significant interaction terms for Hispanic ethnicity and TS Gold language (p = .098) and literacy (p = .045) suggest that the program has weaker effects for Hispanic children. Given the numerous tests, however, these findings could occur by chance. Overall, then, the program appears to work similarly across groups defined by pretest risk.

[34] To maximize cases, we used the imputed TS Gold pretest scores as a control but use the unimputed TS Gold scores as outcomes.

IMPACT QUESTION 5 (EXPLORATORY) TO WHAT EXTENT DOES THE AMOUNT OF PARENTS' PARTICIPATION IN MOTHEREAD (I.E., NUMBER OF SESSIONS) CORRELATE WITH GAINS IN CHILDREN'S LITERACY?

The next exploratory question asks whether more of the program leads to better outcomes among intervention participants. As noted above, this analysis has some methodological limitations relating to the inability to assign participants randomly to low and high participation. Worsening the problem, the analysis can use only the intervention subjects with data on program participation and posttest TS Gold measures – a number ranging from only 60-67.

Our tests focus on four measures of program participation and the three outcome measures of the TS Gold language categories, the TS Gold literacy categories, and the child literacy scale combining the five separate parent-rated child measures. Table 8 presents the standardized coefficients for the three outcomes regressed on each of the participation measures with a control for the pretest outcome.[35]

Table 8. Influence of Participation on TS Gold Language and Literacy Categories (N = 60-67)

Participation Measure	TS Gold Language Categories		TS Gold Literacy Categories		Parent-Rated Child Literacy	
	beta	**p**	**beta**	**p**	**beta**	**p**
Attendance Rate	.02	.856	.05	.680	-.02	.811
Average minutes each week reading the MR book	.05	.701	-.13	.296	.09	.345
Average minutes each week reading another book	.06	.602	.16	.214	-.08	.458
Average minutes each week reading books without child	.08	.474	-.03	.827	-.12	.231
Note: beta = standardized regression coefficient controlling for pretest outcome						

The results show weak relationships. All the coefficients are small and nonsignificant. Examining difference among the intervention group only, parent involvement in the program appears to do little to improve teacher ratings of child language and literacy skills.

[35] To maximize cases, we again used the imputed TS Gold pretest scores as a control but use the unimputed TS Gold scores as outcomes.

IMPACT QUESTION 6 (EXPLORATORY): TO WHAT EXTENT ARE POSTTEST IMPROVEMENTS IN MOTHEREAD CHILDREN MAINTAINED OVER THE NEXT MONTHS AND YEAR?

The final exploratory research question addresses the extent to which improvements gained by children whose parents receive MFC program are sustained in the months and year following participation in the program.

The longitudinal data set for the MFC program includes all TS Gold assessments and in that way differs from the pretest-posttest data set used in the previous section. In addition to the pretest and posttest assessments, the longitudinal data for some cohorts include an interim assessment between the pretest and posttest (while the program is ongoing). It also contains up to four additional assessments after the posttest. The first follow-up assessment occurred as soon as three months after the posttest, while the last follow-up assessment occurred as late as 15 months after the posttest. Given the additional follow-up assessments, the longitudinal data differ in another way. The subjects comprise three groups: the original treatment subjects, the original control subjects who accepted the opportunity after the posttest to join the program, and the original control subjects who declined the invitation after the posttest to join the program. While the pretest and posttest assessments can compare the randomized treatment and control groups, the follow-up assessments can compare the three groups.

Statistical Issues

The goal of the longitudinal data analysis is to see if the program has sustained or newly emergent effects in the period following the program. However, this analysis faces two special statistical problems corresponding to the unique features of the longitudinal data.

First, after the posttest, the division of the originally randomized control group subjects into those who joined and did not join the program was non-random. That makes it hard to separate follow-up changes due to the program from changes due to self-selection of some control subjects into the program. For example, those who accepted the opportunity to join the program may be more motivated to help their children than those who did not. Tests of program efficacy therefore need to examine and possibly control for background differences between control group subjects who join and do not join the program after the posttest.

The original randomization assigned 73 to the treatment group and 72 to the control group. As shown earlier, these groups are quite similar on the background variables. The three posttest groups consist of 73 in the treatment group (as before), 30 in the control-treatment group, and 42 in the continued-control group. Because the three groups are no longer randomized, it is necessary to identify how they may differ from one another – significant differences could bias comparisons of the three groups.

One-way analyses of variance examine if the means for each of the baseline measures differ across the three groups. Only one baseline measure, parent education, is significant. However, a more thorough examination of group differences is found in Table 9, which lists the means for the three groups and the standardized mean differences (Cohen's d) between each of the control groups and the treatment

group. Six baseline measures have relatively large standardized differences (d > .25): 1) parent education, 2) the active reading scale, 3) female child, 4) education goals for child, 5) parent-rated language awareness, and 6) parent-rated writing. The continued-control group has lower scores than the treatment group on parent education and active reading, while the control-treatment group has fewer girls, lower language awareness, and poorer writing scores than the treatment group. More surprising, the control-treatment parents have higher education goals than the treatment group. It appears, then, that one control group is not consistently more disadvantaged than the other: The continued-control group does worse on parent education and active reading, while the control-treatment group does worse on child language awareness and writing. Still, the longitudinal tests for group differences will need to control for these baseline measures.

Table 9. Group Means and Standardized Differences (d) for Three Study Groups

Baseline	Continued-Control		Control-Treatment		Treatment
Measure	Mean	d [a]	Mean	d [b]	Mean
Child Age	4.10	0.17	3.93	-0.08	3.99
Child Female	0.57	0.06	0.40	-0.29	0.54
Free Lunch	1.23	0.00	1.25	0.03	1.23
Race	0.33	-0.05	0.27	-0.19	0.36
Hispanic	0.40	-0.18	0.40	-0.19	0.49
Language	0.76	0.12	0.80	0.20	0.71
# Children	2.35	-0.12	2.23	-0.23	2.48
Family Income	$45,594	0.12	$43,343	0.04	$42,354
# Dependents	4.16	-0.07	4.07	-0.15	4.26
Income / Dependents	$11,409	0.23	$11,288	0.21	$9,844
Parent Education #	3.58	-0.44	4.07	0.02	4.05
Child Education Goals	0.89	-0.03	1.00	0.38	0.90
Child Education Expectation	0.81	0.08	0.83	0.14	0.78
Reading Scale	3.75	0.17	3.47	-0.15	3.61
Language Scale	2.70	0.15	2.21	-0.28	2.52
Letters Scale	3.46	0.14	3.30	-0.02	3.32
Writing Scale	2.65	-0.06	2.32	-0.44	2.70
Print Scale	2.48	0.04	2.45	0.02	2.43
Active Reading Scale	2.60	-0.33	3.30	0.07	3.18

a Continued-Control minus Treatment

b Control-Treatment minus Treatment

Means differ significantly (P < .10)

Second, a potential statistical problem comes from some subjects having more follow-up assessments than others. The number differs because children started earlier and stayed longer in the preschool. Attrition that differs across the treatment and two control groups may bias the comparisons. Tests therefore need to examine group differences in the number of assessments.

Another one-way analysis of variance shows significant differences across groups in the number of assessments. The control-treatment group has the higher mean number of assessments (2.1), followed by the treatment group (2.0) and the continued-control group (1.4). Post-hoc differences-of-means tests with a Bonferroni adjustment show significant differences of the continued-control group with the other two groups but no difference between the treatment and control-treatment groups. A control for number of assessments will be needed in the longitudinal tests for group differences.

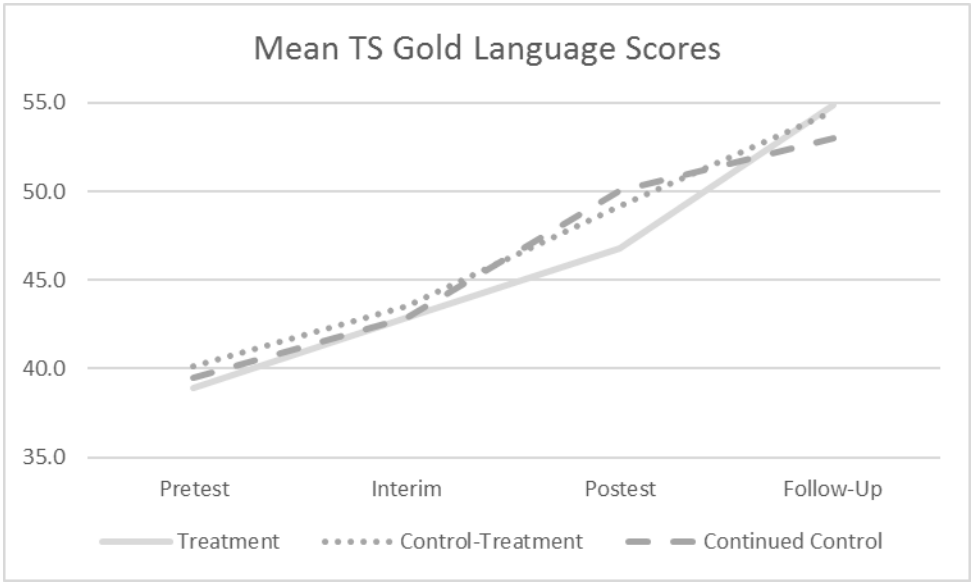
Models and Hypotheses

For the longitudinal analysis, we can use two approaches. First, we can use all assessments, including the pretest and interim assessment that precede the posttest and follow-up. However, the previous analyses focusing on pretest to posttest change showed no treatment effect. As an illustration of the value of a follow-up analysis, Table 10 shows the means for the four TS Gold measures across the three groups and four periods (at pretest, interim assessment, posttest, and the follow-up period), and Figure 3 graphs the means for the TS Gold language score. At posttest, the means differ little, if anything showing higher scores for the control group. These results simply confirm the previous findings. More intriguing are the changes after the posttest. The means for the TS Gold language score and category indicate a larger increase in the treatment group than either of the control groups. The change after posttest rather than from pretest to posttest needs to be investigated in more depth. Given these figures and previous findings of no program posttest effects, we devote most attention to another approach and present results for the full period as a supplement.

Table 10. Descriptive Statistics for TS Gold Measures by Treatment Group and Assessment Period

			TS Gold Language			TS Gold Literacy	
Assessment	Treatment Group	N	Mean Score	Mean Category	N	Mean Score	Mean Category
Pretest	Continued Control	25	39.5	0.68	24	32.5	0.58
	Control-Treatment	17	40.1	0.76	17	33.8	0.76
	Treatment	41	38.9	0.56	41	32.9	0.66
Interim	Continued Control	24	42.9	0.79	24	35.5	0.71
	Control-Treatment	17	43.5	0.76	17	42.8	0.88
	Treatment	45	42.9	0.87	45	37.6	0.87
Posttest	Continued Control	42	50.1	1.17	42	48.3	1.12
	Control-Treatment	30	49.2	1.20	30	49.8	1.20
	Treatment	73	46.8	1.12	71	46.2	1.15
Follow-up	Continued Control	35	53.0	1.09	29	53.0	1.21
	Control-Treatment	37	54.5	1.27	33	63.2	1.39
	Treatment	88	54.9	1.35	85	61.2	1.27

Figure 3. Trend by Group in Mean TS Gold Language Score



Second, a more focused strategy for the longitudinal analysis is to examine the change from posttest to follow-up only. We can compare changes in TS Gold from posttest to the follow-ups for the randomly

assigned treatment group and the control-treatment and continued-control groups. This strategy has the advantage of avoiding two different contrasts, one between the original treatment and control groups for pretest to posttest and the other between the three groups for posttest to follow-up. The focus on the posttest to follow-up change for the three groups thus both simplifies the analysis as well as focuses on the period in which program effects may be observed.

A repeated-measures, mixed model is well suited for studying group differences in TS Gold change from posttest to follow-up. The fixed effects component of the mixed model in which time is nested within individual children takes the following form:

$$Y_{ti} = \beta_0 + \beta_1 * A_{ti} + \beta_2 * CC_i + \beta_3 * CT_i + \beta_4 * A_{ti} * CC_i + \beta_5 * A_{ti} * CT_i + \Sigma \beta_k X_{kti},$$

where Y_{ti} refers to the TS Gold score at assessment t for child i, A_{ti} refers to the assessment date, CC_i to the continued-control group, and CT_i to the control-treatment group, and X_k to k control variables. The models include a random intercept, and a random slope for the assessment date, which reflects variation across children in the growth rate of the outcome. With this specification, the models include error terms both within and between individuals, adjust the standard errors for the clustering of assessments within individuals, and allow for uneven numbers of assessments across individuals. The specific model terms include:

1. Two dummy variables for the control groups (CC_i and CT_i) and an omitted category for the treatment group; the effects β_2 and β_2 show the differences between the treatment and each of the control groups *at posttest*.

2. A term for the assessment date (A_{ti}), ranging from 0 for the posttest, 1 for 3 months after the posttest, 2 for 6 months after the posttest, and so on until 5 for 15 months after the posttest; the effect β_1 shows the growth in TS Gold scores over time *for the omitted treatment group*.

3. Two interaction terms for the assessment period times each of the two control-group dummy variables ($A_{ti} * CC_i$ and $A_{ti} * CT_i$); the effects β_4 and β_5 show the *difference in the growth* in the TS Gold scores for each control group relative to the treatment group.

4. Controls include baseline measures that were part of the evaluation design and that were found to differ across groups. The first group includes cohort and assessment season in the form of dummy variables, and child age at the first assessment. The second group includes the six measures identified above as differing across groups.

If the program has effects after posttest, we would predict the following: 1) no differences between the groups at posttest ($\beta_2 = \beta_2 = 0$), 2) an upward trend in scores for the treatment group after the posttest ($\beta_1 > 0$), and 3) weaker upward trends in scores for the two control groups ($\beta_4 < 0$ and $\beta_5 < 0$).

The model is estimated for four outcomes: 1) the TS Gold interval-level language score, 2) the TS Gold interval-level literacy score, 3) the TS Gold ordinal language classification based on grade-based expectation (0 = below, 1 = meets, 2= exceeds), and 4) the TS Gold ordinal literacy classification based on grade-based expectation (0 = below, 1 = meets, 2= exceeds). The first two scales are not grade

adjusted, but the models control for age at the first assessment and then use the assessment date measure to capture improvements in subsequent assessments at older ages and higher grades. Assessment date captures the growth in scores as children grow older. The last two ordinal outcomes are adjusted directly for grade standards by the teacher ratings, but the control covariates are still used for age at first assessment and assessment date. Linear mixed models are used for the interval and ordinal outcomes. Tests for the ordinal outcomes using mixed models for ordinal, multinomial, or binary logistic regression do not converge.[36] Although not ideal, the linear mixed models can provide a useful approximation of program effects.

Results

Table 11 presents the basic mixed model coefficients for the TS Gold language and literacy scores. The models control for age at first assessment, cohort, and season of assessment (fall, winter, and spring).[37] The sample size of 319 cases for TS Gold language includes 144 subjects and an average of 2.2 assessments per subject. The additional control variables, which reflect baseline differences across groups but have more missing data, will be added subsequently.

[36] The below expectations category includes only about 15% of the cases and is perfectly predicted by several of the independent variables.

[37] Controls for the five preschools in the sample overlap with cohort, have no influence on the TS Gold scores, and are not included.

Table 11. Repeated-Measures Mixed-Model Estimates for TS Gold Language and Literacy Scores: Posttest to Follow-up

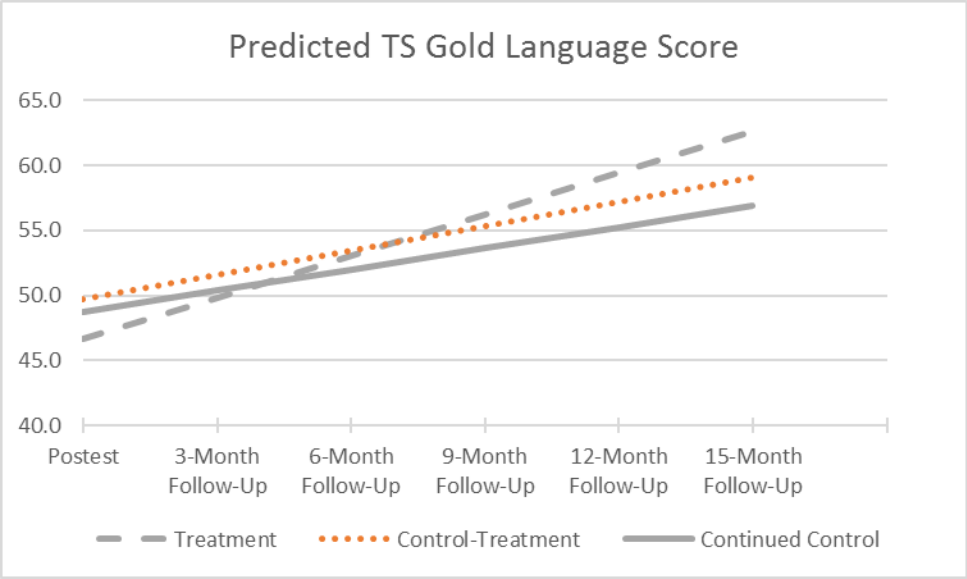
	TS Gold Language Score		TS Gold Literacy Score	
Predictors	b	t	b	t
Intercept	15.08		-17.34	
Control Variables				
Age at Start	6.9	6.51***	13.31	6.34***
Cohort 1 (omitted)				
Cohort 2	-0.46	-0.2	4.93	1.08
Cohort3	-0.8	-0.38	-2.91	-0.7
Cohort 4	-4.57	-2.26*	-10.38	-2.64**
Cohort 5	-2.08	-0.89	0.46	0.1
Fall Assessment (omitted)				
Winter Assessment	2.53	3.21**	6.75	4.91***
Spring Assessment	5.68	7.61***	11.17	8.65***
Program Variables				
Treatment (omitted)				
Control-Treatment	3.02	1.89	6.18	1.97
Continued Control	2.07	1.44	-0.3	-0.11
Follow-up Assessment	3.19	8.83***	6.83	10.34***
Follow-up Assessment x				
Control-Treatment	-1.32	-2.03*	-1.8	-1.45
Follow-up Assessment x				
Continued-Control	-1.56	-2.31*	-2.52	-1.86
Variance intercept	40.88	6.40***	168.2	6.89***
Variance Follow-up	1.69	2.62**	6.3	2.70**
Assessment				
N Children	144		142	
N Children-Assessments	319		289	

* p < .05; ** p < .01; *** p < .001

First, the results for the TS Gold language scores confirm the predictions. The coefficients for the control-treatment and continued-control groups are not significant (although the control-treatment group is marginally significantly higher than the treatment group). This indicates that all three groups have similar scores at posttest. The coefficient for the assessment date is positive, indicating that the treatment group increases the TS Gold score by 3.19 points every three months. Most importantly, the coefficients for both interaction terms are negative and significant, indicating that the increase in the TS Gold scores is weaker for each control group relative to the treatment group. The results thus show similar program impacts for both groups of control participants, those who went on to receive the treatment and those who did not. Both groups did significantly worse than the original treatment group.

This can be illustrated by graphing the linear slopes from posttest to the follow-ups for each group. Figure 4 depicts the linear change in TS Gold language scores over the full follow-up period. Although the groups start from similar points, the treatment group has a greater growth rate than both control groups. The faster growth would be expected for the treatment group relative to the continued-control group, which may include less motivated parents. More notable is the faster growth for the treatment group relative to the control-treatment group. This group of parents shows high motivation in choosing to join the program, and thus defines a higher standard of comparison.[38] Yet, the treatment group still does better in the growth of the TS Gold language scores from posttest to follow-up.

Figure 4. Trend by Group in Predicted TS Gold Language Score



[38] The control-treatment group parents also enjoy the benefits of participating in the program during the follow-up period. Previous findings suggest that these benefits do not emerge immediately, but they may contribute to the high standard of comparison.

The results for the TS Gold literacy scores reveal less evidence of treatment group benefits. The growth rate is higher for the treatment group than both control groups, but neither of the interaction terms reaches statistical significance and the differences fall within the range of random error.

Table 12 presents the mixed models with the same set of predictors for the two ordinal outcomes. Given the adjustment for grade in the TS Gold measure, the assessment date significantly influences the language rating but not the literacy ratings.[39] The growth in the language rating for the treatment group is significantly greater than that for the continued-control group but not for the control-treatment group. The results for the language rating suggest some benefit of late treatment for the original control group, and differ somewhat from the previous results for the language scores. For growth in the literacy rating, however, neither of the control groups differs significantly from the treatment group.

[39] The variance of these slopes was not significant, and the models therefore specify random effects only for the intercept.

Table 12. Repeated-Measures Mixed-Model Estimates for TS Gold Language and Literacy Grade-Based Categories: Posttest to Follow-up

	TS Gold Language Category		TS Gold Literacy Category	
Predictors	b	t	b	t
Intercept	0.49		0.63	
Control Variables				
Age at Start	0.08	1.13	0.06	0.86
Cohort 1 (omitted)				
Cohort 2	0.03	0.21	0.23	1.46
Cohort3	-0.25	-1.80	-0.23	-1.58
Cohort 4	-0.22	-1.62	-0.24	-1.74
Cohort 5	-0.25	-1.53	-0.25	-1.46
Fall Assessment (omitted)				
Winter Assessment	0.34	4.00***	0.25	3.13***
Spring Assessment	0.63	7.80***	0.55	7.44***
Program Variables				
Treatment (omitted)				
Control-Treatment	0.07	0.63	0.09	0.74
Continued Control	-0.01	-0.11	-0.09	-0.89
Follow-up Assessment	0.08	3.25***	0.01	0.35
Follow-up Assessment x				
Control-Treatment	-0.06	-1.39	-0.02	-0.43
Follow-up Assessment x				
Continued-Control	-0.11	-2.65**	-0.03	-0.59
Variance intercept	0.12	4.23***	0.16	5.09***
N Children	144		142	
N Children-Assessments	319		289	

* p < .05; ** p < .01; *** p < .001

Influence of Control Variables

The basic results provide some evidence of faster growth in TS Gold language scores after the program ends but the evidence is less clear for the TS Gold literacy scores. We checked these results by adding controls for baseline measures that differed significantly between groups and adjusting for the varied

number of follow-ups across subjects and groups. We estimated these models for the two interval-scale outcome measures of TS Gold language and literacy scores.

First, the additional control variables substantially reduce the number of cases. The six variables to add include parent education, the active reading scale, female child, educational goals for the child, parent-rated language awareness, and parent rated writing skills. Reflecting loss of cases for subjects with missing pretest data, the sample sizes drop from 319 cases and 144 subjects to 277 cases and 128 subjects for the TS Gold language score model and from 289 cases and 142 subjects to 259 and 126 for the TS Gold literacy score model.

Of the added baseline controls, the parent-rated language awareness scale significantly influences the TS Gold literacy scores, but none of the other controls has a significant effect on either outcome. It is therefore not surprising that the program effects change little from those presented in Table 11. For the TS Gold language score, both control groups grow significantly more slowly than the treatment group, while for the TS Gold literacy score, the groups do not differ significantly.

Second, the models also control for the number of assessments and the number of assessments squared (to allow for non-linearity), and both significantly influence both outcomes. TS Gold scores increase with additional assessments but given the non-linearity, the effects level off at the highest number of assessments. Again, however, the different growth rates between groups shown in Table 11 change little with the controls. The higher growth rate in TS Gold language scores after posttest for the treatment group does not disappear.

Models for All Assessments

We checked the results using the full period of assessments, from pretest to the last follow-up. The analysis strategy can take two approaches. First, the pretest to posttest period can be treated separately from the posttest to follow-up period. This approach assumes that little change in child language and literacy can emerge until after the program has been in place for a sufficient amount of time and that the pretest to posttest effects will be substantially weaker than the posttest to follow-up effects. The models accordingly use separate growth terms for the two periods. The models also include interaction terms of the control-treatment and continued-treatment groups times the terms for the early and late periods. The results for the TS Gold language scores largely replicate the earlier results. The growth rate during the follow-up period is significantly greater for the treatment group than the continued-control group ($p = .025$), and marginally significantly greater for the treatment group than the control-treatment group ($p = .079$). The results for the TS Gold literacy scores also replicate the earlier results, but in this case showing no group differences in the follow-up period.

A second approach assumes that the program brings about steady, incremental changes from the very start to the end of the follow-up period. This approach uses a single linear growth term covering the full assessment period, which reflects continual change from program start to after the program end.[40] The

[40] Models with linear and squared terms to capture non-linearity perform significantly worse than the linear models and are not presented.

results for the TS Gold language scores show that the growth rates do not differ significantly across groups. It appears that when averaged across the full period, the program shows little benefits. This result contrasts with the previous results, which show no benefits up to the posttest but significant benefits afterward. The results for the TS Gold literacy scores likewise differ from the previous results, but in a more positive direction. The growth term is significantly greater for the treatment group than the continued-control group.

Conclusion

In brief, the Motheread program appears to have a delayed but meaningful effect on one objectively and independently rated measure – the TS Gold language score. The child improvements do not show up immediately after the program, but it appears that parents continue the Motheread activities and that children eventually benefit from the activities. The results for the other TS Gold measures are less consistent. The TS Gold language rating relative to grade expectations grew faster for the treatment group than the continued-control group but not faster than the control-treatment group. Further, the program did not significantly improve growth in the TS Gold literacy scores or ratings. The findings are confirmed with additional controls for baseline measures and for number of assessments completed. However, models covering the full time period – rather than from posttest through follow-up – produce mixed results.

Given mixed findings, we cannot assert with certainty the treatment results are real – they might have resulted from selection bias or random error. Yet we can say that some evidence favors the conclusion that the treatment leads to real gains in TS Gold language scores. To the extent that the results are reliable, it suggests a delayed or cumulative effect of the treatment. Those children in the original treatment do not show higher TS Gold language scores right away, but the small initial benefits provide a foundation that fosters later gains. The original control group that later received the treatment – and was followed for only a short period after treatment – also does not appear to benefit from the program right away, but they might show gains had they been followed longer.

Lessons Learned & Recommendations

Program implementation and evaluation activities during the Social Innovation Fund implementation and impact study yielded important lessons for the research team which are outlined below. Along with each challenge, we note recommendations for future work.

RECRUITMENT CHALLENGES

Recruitment of families to participate in the study was the most significant challenge faced by the research team. Despite calculating conservative recruitment numbers in collaboration with research site partners based on the population in Morgan County, CO where the study was conducted, recruitment was lower than anticipated for each cohort of the study. Midway through the study a revised power analysis was conducted, based on preliminary findings, that determined that a minimum sample size of 128 would be necessary to detect significant results. The research team worked in close collaboration with Mile High United Way, Colorado Humanities and Morgan County Early Childhood Council to determine appropriate steps to recruit a sufficient sample size to complete the study. Efforts included:

- Leveraging additional parent forums/events (e.g., back to school night) to talk about the program and start recruitment early for the next cohort.
- Enlisting parents who had already participated in the program to talk with their friends and peers about the benefits of participation in the study. This effort included having parents who had participated in the study call a list of parents who had signed up for demonstration sessions, but who did not attend, to discuss participation. Parents making the calls were incentivized for their time with $15 gas gift cards.
- Offering the opportunity for parents to take the class in a different community or on a different day than the session being offered at the child's preschool site.
- Enrolling an additional preschool site for Cohort 5 of the study to increase enrollment numbers.
- Employing standard marketing techniques, including billboards, radio and print advertising to increase participation.

Ultimately 148 participants were recruited for the study, which exceeded the 128 minimum sample size determined by the power analysis. Nonetheless, the sample size was smaller than hoped. In future studies, it may be helpful to offer the class during a day time and/or weekend timeslot and to identify other creative ways to motivate busy families to attend.

CONDUCTING RIGOROUS RESEARCH IN AN APPLIED SETTING

The nature of the randomized controlled design of the study necessitated that some participants wait to be enrolled in the Motheread/Fatheread program. Although, from a methodological perspective, this was the design with the most rigorous test of the impact of the program, it was difficult for the research

site coordinators to have to communicate to individuals in their community that they could not participate in the program right away. The research team worked closely with the research site team to devise strategies to minimize disappointment. For example, random assignment was conducted by the research team and communicated directly from the research team to participants in order to take research site partners out of the "middle" in terms of group assignment communication. The research team also worked to make communication about random assignment accessible to members of the community who did not have a research background, for example, but explaining random assignment as analogous to "flipping a coin." Being intentional about straightforward communication using language that non-researchers could understand helped mitigate concerns around random assignment. The research team recommends in person meetings as much as possible to allow for direct contact with potential participants and explaining research in terms that non-researchers can easily understand.

CREATE MORE INTENTIONAL "CLOSINGS" OF THE MOTHEREAD CLASS

Observations from MFC coaches and researchers revealed that often the closing section of the MFC class was abrupt. MFC coaches noted that program may benefit from adjusting the format of the sessions to maximize opportunities to practice and develop skills learned in class and allow time for reflection and discussion that connects the stories and the skills practiced to their lives. This could include structuring a more formal class closing/session ending, allotting time to work with Story Extenders during class, and allowing time to practice skills that parents may feel more vulnerable or uncomfortable with, such as reading aloud or acting out sections of a story so they have more confidence employing these techniques at home.

LONGITUDINAL TS GOLD COLLECTION EXTENSION

The results of the study revealed a delayed but potentially meaningful impact of the program on children's TS Gold scores. Unfortunately, the study design, small sample size and missing data limited the ability to draw strong conclusions from the findings. Furthermore, TS Gold data were available at most 15 months from participants' entry into the study. A replication study should consider a longer period to test program impact and additional resources to increase the sample pool.

Considerations and Limitations

There are several considerations to keep in mind when interpreting the results of this impact study.

First, the study design included child literacy outcomes as measured by both teacher and parent assessments of child literacy and language skills. However, findings were mixed. At posttest, parents in the intervention group reported better literacy outcomes for their children than parents in the control condition, whereas the teacher report measures were not significantly different between intervention and control children until later in the follow-up period. Because a double-blind methodology was not possible (i.e., we could not blind parents to participation in the literacy intervention), it is possible that parents in the intervention group expected their children to do better than parents in the control group and that findings were a result of a reporting bias (an 'expectancy effect') rather than true change in children's skills. However, it is also possible that parents in the intervention group had higher standards when assessing child literacy skills as a result of their increased education surrounding literacy. That is, the more that parents read with their children and their children's skills improved, the more parents developed a high standard for children's literary competency. With this rationale, it is possible that the true difference between the intervention and control groups in literacy skills is even greater than what was reported. Unfortunately, we do not have the appropriate data to further assess these possibilities and acknowledge that definitive conclusions of program impact as assessed by parent reports of child behaviors are limited.

Second, parent reports of their own reading behavior are also subject to bias. Parents in the intervention group may have overstated their reading behaviors or the 'novelty' or newness of the program, rather than the intervention itself, may have accounted for the change in parents reports of how they read to their children. Although parent behaviors were not measured longitudinally after participation in the intervention, longitudinal findings from the TS Gold suggests that parent participation in the MFC program has long term effects on child literacy outcomes. As such, these data also suggest it is likely that parent behavior change is not just short term, but rather sustained beyond the course of the program.

Third, the study was conducted in a single county in rural Colorado. The rigorous design supported detection of a moderate, rather than high, level of evidence per CNCS guidelines. As such, results of the current study are not generalizable beyond the community in which the study was conducted. We do not have reason to expect that findings would differ in other communities, but studies in additional settings are necessary to increase confidence that the program is effective across settings and populations.

Finally, recruitment challenges led to a smaller sample size than expected for this study and limited the statistical power necessary to conduct some of the planned statistical analyses. The limited sample size may also have limited our ability to detect statistically significant effects of the program.

Next Steps

To our knowledge, MFC is the only program in Colorado that trains educators to work with parents to increase parents' engagement in their young children's literary education. Colorado Humanities plans to continue to scale up implementation of the MFC program in Colorado to the school district and community-wide level. Thus far, Colorado Humanities has had success with program scaling efforts in two counties in Colorado. As part of their scaling efforts, Colorado Humanities also plans to build in more support for MFC program infrastructure, including fundraising for trainings, providing direct support for implementation costs, Motheread Coach trainings, and evaluation. Furthermore, Colorado Humanities plans to develop strategies for continuing to collect impact data from partners implementing the MFC program.

Should funding be available, future studies of the MFC program will seek to expand on the generalizability and longitudinal effects of findings from the current impact study. Generalizability will be addressed by expanding the subject population to broader and more diverse participant pools. Longevity of the impact of the program will be addressed by lengthening the timeline in which outcomes for both parents and children are tracked after parent completion of the MFC program. Specifically, assessing children's academic achievement through third grade would be important for determining long term effects of the program on children's third-grade reading proficiency. Because of the promising findings from this study, future reports would likely target a high level of evidence for the MFC program.

Study Administration

HUMAN SUBJECTS PROTECTION

IRB approval for the study was granted in December 2013. Approval for continuing review was granted in September 2016. There have been no issues with securing or maintaining IRB approval. Data analysis and reporting will be complete by June 2017 and the research team does not plan to apply for annual continuing IRB approval in 2017.

TIMELINE

Data collection was completed on time. The final report of the study is being delivered as scheduled in March 2017.

EVALUATION BUDGET

Proposed and actual budgets for the evaluation are included in the table below. Although actual budgets were somewhat reduced due to a reduction in funding, the evaluation team was able to complete all core components of the evaluation without compromising study rigor by streamlining evaluation processes during the early years of the project.

	Year 1	**Year 2**	**Year 3**	**Year 4**
Proposed Budget	$112,817	$118, 683	$118,683	$95,077
Actual Budget	$100,125	$97,000	$89,892	$90, 025

EVALUATOR /SUBGRANTEE ROLE AND INVOLVEMENT

OMNI Institute supported the research for this study. Dr. Melissa Richmond directed the study and Dr. Holen Hirsh led the study. Dr. Fred Pampel was the statistical consultant on the project.

Appendix A

Table A.1. Parent Race

	All		Intervention		Control	
Parent Race	**N**	**%**	**N**	**%**	**N**	**%**
White/Caucasian	106	74%	54	74%	52	74%
Black/African American	5	3%	2	3%	3	4%
Asian	0	0%	0	0%	0	0%
American Indian or Alaska Native	2	1%	2	3%	0	0%
Native Hawaiian/Pacific Islander	0	0%	0	0%	0	0%
Multiracial	5	3%	1	1%	4	6%
Unknown	0	0%	0	0%	0	0%
Other	25	17%	14	19%	11	16%
Sum	**143**	**100%**	**73**	**100%**	**70**	**100%**
Missing	**5**		**2**		**3**	

Table A.2. Parent Ethnicity

	All		Intervention		Control	
Parent Identifies as Hispanic/Latino	**N**	**%**	**N**	**%**	**N**	**%**
Yes	60	42%	33	45%	27	38%
No	84	58%	40	55%	44	62%
Sum	**144**	**100%**	**73**	**100%**	**71**	**100%**
Missing	**4**		**2**		**2**	

Table A.3. Family Language Spoken at Home

	All		Intervention		Control	
Language Spoken at Home	**N**	**%**	**N**	**%**	**N**	**%**
English	111	75%	54	72%	57	78%
Spanish	33	22%	20	27%	13	18%
Other	4	3%	1	1%	3	4%
Sum	**148**	**100%**	**75**	**100%**	**73**	**100%**
Missing	**0**		**0**		**0**	

Table A.4. Parent Education Status

	All		Intervention		Control	
Parent Education Status	**N**	**%**	**N**	**%**	**N**	**%**
No schooling completed	4	3%	2	3%	2	3%
Some school (but less than high school)	1	1%	0	0%	1	1%
High school graduate/GED	44	32%	19	28%	25	37%
Some college or technical school	49	36%	26	38%	23	34%
College degree	27	20%	13	19%	14	21%
Post college degree	11	8%	8	12%	3	4%
Sum	**136**	**100%**	**68**	**100%**	**68**	**100%**
Missing	**12**		**7**		**5**	

Table A.5. Number of Children in Household

	All		Intervention		Control	
# of Children in Household	**N**	**%**	**N**	**%**	**N**	**%**
1	31	21%	10	13%	21	30%
2	58	40%	31	41%	27	38%
3	37	25%	26	35%	11	15%
4	13	9%	5	7%	8	11%
5 or more	7	5%	3	4%	4	6%
Sum	**146**	**100%**	**75**	**100%**	**71**	**100%**
Missing	**2**		**0**		**2**	

Table A.6. Annual Household Income

	All	Intervention	Control
Household Income (n=125)			
Mean	$43,810	$43,474	$44,163
Median	$36,000	$36,000	$36,000
Min	$750	$750	$6,000
Max	$160,000	$160,000	$104,000

Table A.7. Family Free and Reduced Lunch Status

	All		Intervention		Control	
FRL Status	**N**	**%**	**N**	**%**	**N**	**%**
No, we do not qualify	42	34%	22	34%	20	33%
Reduced-price lunch	14	11%	7	11%	7	12%
Free lunch	68	55%	35	55%	33	55%
Sum	**124**	**100%**	**64**	**100%**	**60**	**100%**
Missing	**24**		**11**		**13**	

Table A.8. Child Race

	All		Intervention		Control	
Child Race	**N**	**%**	**N**	**%**	**N**	**%**
White/Caucasian	100	70%	49	68%	51	72%
Black/African American	4	3%	1	1%	3	4%
Asian	2	1%	1	1%	1	1%
American Indian or Alaska Native	2	1%	2	3%	0	0%
Native Hawaiian/Pacific Islander	0	0%	0	0%	0	0%
Multiracial	9	6%	6	8%	3	4%
Unknown	0	0%	0	0%	0	0%
Other	26	18%	13	18%	13	18%
Sum	**143**	**100%**	**72**	**100%**	**71**	**100%**
Missing	**5**		**3**		**2**	

Table A.9. Child Ethnicity

	All		Intervention		Control	
Child Identifies as Hispanic/Latino	**N**	**%**	**N**	**%**	**N**	**%**
Yes	66	52%	37	55%	29	48%
No	62	48%	30	45%	32	52%
Sum	**128**	**100%**	**67**	**100%**	**61**	**100%**
Missing	**20**		**8**		**12**	

Table A.10. Hopes for Child Longevity in Formal Schooling

	All		Intervention		Control	
	N	**%**	**N**	**%**	**N**	**%**
How far do you *want* your child to go in formal schooling?						
Finish Elementary School	0	0%	0	0%	0	0%
Finish Middle School	0	0%	0	0%	0	0%
Finish High School	2	1%	2	3%	0	0%
Attend Trade School	1	1%	1	1%	0	0%
Attend College or University	8	6%	4	5%	4	6%
Finish College or University	132	92%	67	91%	65	94%
Sum	**143**	**100%**	**74**	**100%**	**69**	**100%**
Missing	**5**		**1**		**4**	

Table A.11. Expectations for Child Longevity in Formal Schooling

	All		Intervention		Control	
How far do you *think* your child to go in formal schooling?						
Finish Elementary School	0	0%	0	0%	0	0%
Finish Middle School	0	0%	0	0%	0	0%
Finish High School	6	4%	4	5%	2	3%
Attend Trade School	4	3%	3	4%	1	1%
Attend College or University	18	13%	9	12%	9	13%
Finish College or University	114	80%	58	78%	56	82%
Sum	**142**	**100%**	**74**	**100%**	**68**	**100%**
Missing	**6**		**1**		**5**	