Paper 2: Foundations of Science Literacy

Context: The need for a scientifically-literate American population has gained great prominence on the educational landscape of priorities over the past decade; young children will be increasingly exposed to the STEM fields and encouraged to excel in these areas (The White House, 2009). While experience and research suggest that teachers’ science knowledge is predictive of children’s science learning (Loucks-Horsley, Love, Stiles, Mundry, & Hewson, 2003), many early childhood teachers are not ready to engage children in rich science experiences that lay the groundwork for later success. Instead, teachers often rely on “spur of the moment” planning or provide activities that are loosely connected to a theme (Bowman, Donovan, & Burns, 2001). Even quality programs tend to emphasize language and social development, at the expense of science learning (Smith & Dickinson, 1994). These challenges are exacerbated by some teachers’ apprehension towards teaching science based on uncertainty of both the science content and pedagogy (Clark-Chiarelli & Gropen, 2008). When science instruction does occur, it usually focuses either on the recall of facts with insufficient attention to student understanding or, alternatively, on activities that are only loosely related to conceptual goals. Neither approach promotes teacher-student interaction that furthers science learning (Duschl, Schweingruber, & Shouse, 2007).

Professional development is key to assuring that early childhood teachers provide children with cognitively-challenging early learning experiences (Bowman, et al., 2001; Dwyer, Chait, & McKee, 2000; Espinosa, 2002; Helburn & Bergmann, 2002). Unfortunately, few models of professional development build preschool teachers’ skills and knowledge in an ongoing way and provide access to content knowledge (Barnett, 2003; Whitebook, 2003). Recent research comparing different types of professional development for science teaching suggests that the most effective models are a hybrid, combining professional development with curriculum in ways that lead to intentional and informed use of curricular materials (Penuel & Gallagher, 2008). In addition, effective teacher professional development should not only focus on developing teachers’ science content knowledge, but also their pedagogical content knowledge related to children’s early science development (Ball, 2000; Shulman, 1987). When it does so, teachers are prepared to provide developmentally appropriate scaffolding and experiences that support early science learning.

Purpose of Research: We are responding to the critical need for empirical evidence on effective strategies to improve science instruction in preschool. Focusing on the Head Start community, Foundations of Science Literacy (FSL) is a credit-bearing professional development course that directly addresses the achievement gap in early science education. The program not only addresses an urgent need, it also integrates the resources, structure, and support that preschool teachers need to improve early science learning and teaching. Based on many years of experience, we have learned that episodic workshops, offered without a sound curriculum or credit, do little to change teachers’ classroom practice. In sharp contrast, the great promise of FSL is that it includes several features that create a comprehensive approach.

Research Questions: The present study was designed to test (1) the overall fidelity of implementation of FSL that was achieved relative to control classrooms, and (2) the positive impact of FSL on preschool children’s understanding of science and scientific thinking as mediated by improvement of preschool teachers’ pedagogical content knowledge and quality science instruction within the classroom. By implementing a randomized-controlled study, we can examine the impact of FSL on preschool children, and the mediating factors that link teacher professional development to children’s scientific thinking capacities.
Setting: FSL was implemented in two communities in the greater New York City area. The large majority of participating programs were Head Start programs. Implementation and research were conducted during the 2009-2010 school year.

Participants: Our population included preschool teachers and a sample of children in their classrooms. The analytic sample is comprised of 40 intervention classrooms and 32 control classrooms, with 436 children (270 in intervention classrooms and 186 in control). Children were eligible if they spoke either English or Spanish, and turned four years old by 12/31/2009.

Intervention: FSL consists of three components: 1) instructional, face-to-face sessions that build teachers’ pedagogical content knowledge in the physical sciences; 2) a coaching component that supports teachers as they master content and methods of inquiry in the physical sciences; and 3) one unit of the Young Scientist Series (YSS), a unique preschool science curriculum for 4-year-olds in widespread use and with recognition from national science education organizations. While each component takes on a different role in supporting changes in teacher practice that lead to child outcomes, they all work together to enhance teachers’ abilities to: 1) engage all children in exploration through the use of effective plans, strategies, and materials; 2) focus children’s investigations on concepts related to matter and forces; and 3) surface naïve theories and support children’s ability to represent and reflect on those ideas.

Research Design: We used a randomized controlled trial (RCT) design with a total sample of 78 preschool classrooms, with 40 classrooms assigned to the intervention group and 32 classrooms assigned to the control group. We employed a randomized sample that was intentionally not balanced for numbers of children and classrooms in the intervention and control groups (Myers & Dynarski, 2003). This degree of imbalance in the random assignment plan has a negligible impact on the precision of the impact estimates, and is often preferable as it potentially maximizes cost effectiveness, increases statistical power, and limits the number of individuals who potentially will not benefit from the intervention (Puma et al., 2001). During randomization, classrooms were blocked by program location (one of two locations) and by center. Children were then selected within each classroom. If we received consent forms for more than 10 eligible children within a classroom, the study team randomly selected 10 to be in the main sample.

Data Collection and Analysis: Measures. Observations were conducted in each classroom, teachers completed a performance task, and children were assessed one-on-one before and after the implementation of the FSL professional development course. Data collection occurred during the months of October 2009- November 2009 for the Fall and April 2010-June 2010 for the Spring. Fidelity of implementation was only measured once throughout the year, in Spring 2010. All intervention classrooms were observed for fidelity. In addition, 30 percent of the control classrooms were also measured using the fidelity of implementation observation measure. The following describe the key measures that are essential to the current analyses.

Science Teacher Performance Tasks (STPT). The STPTs are a measure of science pedagogical content knowledge and include four 30-minute performance tasks: (1) analyzing a video vignette of a science experience in the classroom, (2) interpreting a child’s work sample, (3) analyzing misconceptions of water flow, and (4) planning a science experience. These tasks require teachers to analyze different aspects of science instruction and respond to a set of prompts that are designed to gauge teacher knowledge of science content and pedagogy. Based on each of the tasks, teachers construct written responses that are scored using a rubric with a 4-point scale. The scores for the four tasks were averaged to create a composite score.

Science Teaching and Environment Rating Scale (STERS). The STERS is a classroom observation tool originally designed with NIH funding to measure the quality of early childhood
science teaching and learning environments. Using a 1 to 4 rating scale, the STERS measures the following aspects of science teaching in the preschool setting: 1) Physical Environment for Inquiry and Learning; 2) Direct Experiences to Promote Conceptual Learning; 3) Use of Scientific Inquiry; 4) Collaborative Climate that Promotes Exploration and Understanding; 5) Opportunities for Extended Conversations; 6) Children’s Vocabulary; 7) In-depth Investigations; and 8) Assessment of Children’s Learning. Internal consistency estimated to be at .94 (Cronbach’s alpha). The ratings for the eight items were averaged to create a composite score.

**Science Fidelity of Implementation Measure.** The conceptual framework for the Science Fidelity of Implementation measure is based on the key dimensions of inquiry-based science instruction for preschool children. Classroom instruction was rated on a set of 29 statements (e.g., “Teachers learning objectives were clear in the introduction to the exploration”), each of which was scored by an observer on a four-point scale ranging from 0 (if the observer “strongly disagreed” with the statement or if the statement could not be evaluated because the relevant aspect of science teaching was not exemplified at all) to 3 (if the observer “strongly agreed” with the statement based on what was happening in the classroom). The individual items comprise six scales: 1) Focus (Scientific Focus); 2) Engage (Setting the Stage for Inquiry); 3) Explore (Opportunity for Children to Explore Scientific Phenomena); 4) Reflect (Meaning Making Experiences); 5) Plan (Planning Scientific Inquiry); 6) Environment.

**Preschool Assessment of Science (PAS).** For children, we used the PAS, a measure of preschoolers’ concepts, facts, knowledge, and skills in physical science (Gropen, Clark-Chiarelli, & Hoisington, 2006). The PAS includes two “types” of tasks: prediction tasks, and challenge tasks. Prediction tasks measure children’s predictions of a scientific concept, their ability to test that prediction against an observed occurrence, and finally their ability to revise an incorrect prediction based on conflicting observational evidence. The second type of corresponds to a challenge cycle, in which children are presented with a set of materials and a particular problem to solve within two minutes. The PAS’ internal consistency using Cronbach’s alpha (a) is 0.727. The current analyses focus on the prediction task only.

**Analyses:** First, the fidelity of implementation of FSL was assessed. Following Hulleman & Cordray (2009), we calculated the percentage of absolute fidelity of FSL and control classrooms by dividing their mean score on each of the fidelity scales, and overall, by the maximum possible average score of 3. We also calculated a binary complier index, based on the proportion of classrooms achieving an average score of 2 (“Agree”) or higher on each scale. Second, a causal hypothesis of the impact of FSL was evaluated in which FSL improved preschool children’s understanding of science and improved scientific thinking by improving preschool teachers’ pedagogical content knowledge and their implementation of quality instruction within the classroom. In order to evaluate this hypothesis, a series of mediation models were tested using structural equation modeling with Amos 19.0 (Arbuckle, 2010). Nested models were compared for relative goodness of fit using chi-square difference tests and overall goodness of fit was evaluated using Hu & Bentler’s (1999) recommendations for the Comparative Fit Index (CFI > .95) and the Root Mean Square Error of Approximation (RMSEA < .06).

**Findings:** Results of the fidelity analyses, presented in Table 1, indicate a higher percentage of absolute fidelity for FSL classrooms (ranging from 44% to 79%) than for control classrooms (ranging from 12% to 41%). Note that the lowest fidelity for both FSL and control classrooms involved “Meaning Making Experiences.” The binary complier index showed a stronger contrast between FSL and control classrooms. The binary index ranged from 41 to 85 percent for FSL classrooms, compared to 0 to 20 percent for control classrooms.
For the mediation analyses, three variables from a prediction task (PAS) were included as children’s science outcomes, including children’s initial prediction of the outcome of a scientific phenomenon (prediction), their observation of the actual outcome of that phenomenon (observation), and their prediction of a similar phenomenon, revised based on the previous observation (revision). Teacher’s pedagogical content knowledge (STPT) and science teaching quality (STERS) were included as mediators. First, a “direct effects” model was tested in which the direct effect of FSL on children’s science outcomes were specified and the indirect paths through the mediators were constrained to be zero (See Figure 1). This model had poor fit, $\chi^2(71) = 409.04$, CFI = 0.496, RMSEA = .104. FSL did not significantly predict children’s initial prediction, but did significantly predict their observation, $B = 0.111(.045)$, $p = .013$, and their revision, $B = 0.145(.043)$, $p < .001$, controlling for fall scores and other child covariates. Next, a mediation model was tested, specifying both a direct relationship between FSL and children’s science outcomes and also indirect relationships mediated by teacher pedagogical content knowledge and science teaching quality (See Figure 2). This model had good fit according to one of the indices, CFI = .882, RMSEA = .054, and had significantly better fit than the direct effects model according to a chi-square difference test, $\chi^2_{\text{diff}}(8) = 266.65$, $p < .0001$. FSL was a significant predictor of both teacher pedagogical content knowledge, $B = 0.486(.046)$, $p < .001$, and science teaching quality, controlling for spring scores, $B = 1.08(.062)$, $p < .001$. Science teaching quality, in turn, significantly predicted the likelihood of children’s observation, $B = 0.077(.037)$, $p = .038$, and teacher pedagogical content knowledge significantly predicted the likelihood of children’s revision, $B = 0.103(.040)$, $p = .009$. Neither mediator was a significant predictor of children’s initial prediction. The direct effects of FSL on children’s observation and revision were no longer significant. Next a full mediation model was tested in which the indirect paths from FSL to children’s science outcomes were estimated but the direct paths were constrained to be zero. This model had good fit according to one of the indices (CFI = .881, RMSEA = .052) and did not have significantly different goodness of fit from the previous model, $\chi^2_{\text{diff}}(3) = 3.62$, $p = .306$. Finally, nonsignificant indirect paths were removed to produce the final model (See Figure 3). This model did not significantly differ in goodness of fit from the previous model, $\chi^2_{\text{diff}}(4) = 1.77$, $p = .777$, and had good fit according to one index, CFI = 0.884, RMSEA = .050.

**Conclusion:** Our analysis indicates that, compared to control teachers, teachers participating in FSL achieved a higher degree of fidelity to key dimensions of inquiry-based science instruction for preschool children. The results also indicate that future implementations should bolster the fidelity of classroom implementation to aspects of the program involving Meaning Making Experiences—especially, those involving reflection. Structural equation modeling results provide strong evidence for a positive impact of FSL on preschool teachers’ pedagogical content knowledge and quality science instruction within the classroom. These results also provide evidence for a positive impact on FSL on preschool children’s understanding of science and scientific thinking, based on results for two of three items within the prediction task. Finally, mediation analyses provide strong preliminary evidence that preschool teachers’ pedagogical content knowledge and the quality of science instruction are causal mechanisms through which FSL impacted children’s learning. These results shed light on the components of early childhood science intervention that are vital for improving outcomes. Implications for measurement of key outcomes for preschool science interventions and effective practices will be discussed.
Appendix A: REFERENCES


Appendix B. Tables and Figures

Table 1. Fidelity indices.

<table>
<thead>
<tr>
<th>Index</th>
<th>Fidelity Scale</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Focus</td>
</tr>
<tr>
<td>Absolute Fidelity Index (Mean Raw Score/3), %</td>
<td>74.51%</td>
</tr>
<tr>
<td>FSL (N=34)</td>
<td>74.51%</td>
</tr>
<tr>
<td>Control (N=10)</td>
<td>26.67%</td>
</tr>
<tr>
<td>Binary Complier Index (proportion ≥ 2), %</td>
<td>82.35%</td>
</tr>
<tr>
<td>FSL (N=34)</td>
<td>82.35%</td>
</tr>
<tr>
<td>Control (N=10)</td>
<td>20.00%</td>
</tr>
</tbody>
</table>

Figure 1. Direct effects model, $\chi^2 (71) = 409.04, p < .001, CFI = .496, RMSEA = .104$
(Note: Child covariates included in the model are not presented here for simplicity of figure.)

*** $p < .001$, ** $p < .01$, * $p < .05$
Figure 2. Direct and indirect effects model, $\chi^2 (63) = 142.39$, $p < .001$, CFI = .882, RMSEA = .054 (Note: Child covariates included in the model are not presented here for simplicity of figure.)

Figure 3. Final full mediation model, $\chi^2 (70) = 147.78$, $p < .001$, CFI = .884, RMSEA = .050 (Note: Child covariates included in the model are not presented here for simplicity of figure.)