Abstract Title Page

Title: Examining the Efficacy of *Foundations of Science Literacy:* Exploring Contextual Factors

Author(s): Jess Gropen, Nancy Clark-Chiarelli, Stacy Ehrlich, and Yen Thieu

Abstract Body

Context: The current global economy continues to become more reliant on technology and scientific innovations. The U.S. has held a prominent place in the past, but in order to maintain our leadership position, we must become *more* competitive and foster the development of a new generation of scientifically literate citizens. This level of competitiveness requires thinking that is more logical, analytic, and abstract than in the past—an important hallmark of the contemporary world (Rutherford & Ahlgren, 1989). In order to fuel this competitiveness, there have been several recent calls for more innovated approaches to education in the STEM fields, beginning at the preschool level, including President Obama's recent announcement regarding the expansion of his Educate to Innovate campaign (The White House, 2010).

Not only has our educational system faltered in comparison to other countries, it has also faltered for particular populations within our own country. Children of color, children who are English language learners, and children of low-income backgrounds demonstrate lesser science proficiency than their peers, and have lower rates of post-secondary degrees and entry into a science-related field (National Science Foundation, 1999, 2007). Achievement gaps begin early (Denton & West, 2002; Lee & Burkam, 2002) and only grow larger over the years (Fryer & Levitt, 2006). To foster its future workforce's science literacy, the United States needs to improve science education for *all* children, at every grade level (Bowman, Donovan, & Burns, 2001; National Commission on Mathematics and Science Teaching for the 21st Century, 2000; Nelson, 1999). Teaching how to think scientifically in the preschool years provides all children with the skills necessary to continue learning and thinking critically throughout the school years, ultimately leading to a population of citizens that are scientifically literate and are able to innovate. Evidence is ample to suggest that young children are eager to learn and ready to be assisted in developing their early underpinnings of more advanced scientific thinking (Bowman et al., 2001; Shonkoff & Phillips, 2000).

Experience and research suggest that teachers' science knowledge is predictive of children's science learning (Loucks-Horsley, Love, Stiles, Mundry, & Hewson, 2003). However, many early childhood teachers are not ready to engage children in rich science experiences that lay the groundwork for later success. Because many early childhood teachers lack formal higher education (Barnett, 2003; Whitebook, 2003), professional development is key to assuring that teachers provide children with cognitively-challenging early learning experiences (Bowman et al., 2001; Dwyer, Chait, & McKee, 2000; Espinosa, 2002; Helburn & Bergmann, 2002). Recent research comparing different types of professional development for science teaching suggests that professional development is most effective when it uses a hybrid model—combining professional development with curriculum in ways that lead to intentional and informed use of curricular materials (Penuel & Gallagher, 2008). Effective teacher professional development should not only focus on developing the teachers' science content knowledge, but also their *pedagogical* content knowledge related to children's early science development (Ball, 2000; Shulman, 1987). The results of a solid professional development program should build on teachers' content knowledge of science and enhance their ability to teach scientific inquiry to young children.

Purpose of Research: We are responding to the critical need for empirical evidence on effective strategies to improve preschool science instruction in preschool. By 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 of great significance, 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.

The present study was designed to test the impact of *FSL* on (1) preschool teachers' knowledge of scientific concepts and pedagogy of teaching science to preschoolers, and ultimately (2) preschool children's understanding of science and improved scientific thinking. By implementing a randomized-controlled study, we can examine the impacts of *FSL* on preschool children, and begin to understand which factors effectively play a role in linking teacher professional development to children's scientific thinking capacities.

Research Questions: These four questions guided our analyses: 1) To what extent do 4-year-olds in *FSL* classrooms develop a greater understanding of physical science content and skills related to physical science inquiry, compared to peers in control classrooms? 2)Is the impact of *FSL* on children's understanding of physical science content, quantitative concepts, and inquiry skills mediated by teachers' classroom practice? 3) Do child characteristics (e.g., mother's education, gender, age) moderate the impact of *FSL* on their understanding of physical science content, quantitative concepts, and inquiry skills? 4) To what extent quality of science instruction impact teachers' classroom practice in teaching physical science to 4-year-olds?

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 the children in their classrooms. Initially, teachers in 78 classrooms participated in our study of *FSL*. By Spring 2010, there was a total of 72 classrooms, including 40 intervention classrooms and 32 control classrooms (note below that randomization was conducted as a 40/60 split). Our child sample consisted of children within these classrooms. Our final sample consisted of 508 children in Fall 2009 (208 in the control group and 300 in the intervention group) and 436 children in Spring 2010 (186 in the control group and 270 who received the *FSL* intervention). Children were eligible for our study if they spoke either English or Spanish, and turned four years old by 12/31/2009. Based on parental report, 12.1% of these children had an IEP. Among all the children in the sample, 24.7% had mothers with less than a high school degree, 47.8% with a high school diploma or some level of college, and 27.4% with a bachelor's degree or above.

Intervention: *FSL* is comprised of three components: 1) instructional sessions that are delivered face-to-face and 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.

The *FSL* course consists of 42 instructional hours divided into four face-to-face sessions, paced over a six-month period to allow ample opportunity for application during the inter-sessions through performance-based assignments and coaching. Four college credits are awarded upon participants' successful course completion. All four sessions employ key instructional elements: video reflection and analysis, analysis of children's work samples, and performance-based assignments. We supplement *FSL* course sessions with coaching delivered by instructors and coaches. On average, intervention teachers received 6 coaching visits (1.5 hours per session) paced to coincide with course assignments. Table 1 in Appendix B outlines the schedule of course sessions and coaching visits. (Please insert Table 1 here)

Central to our coaching approach are two strategies. *Content-focused coaching* involves a conference-observation-feedback cycle, which helps teachers learn new methods of structuring content and experiences to promote learning, how to plan instruction more effectively, and ways to assess children's learning (Feiman-Nemser & Parker, 1992; Harvard Family Research Project, 2004; Loucks-Horsley et al., 2003; Wood & McQuarrie, 1999). *FSL* coaching also involves *collaborative inquiry,* bringing together small groups of teachers together to plan and analyze their instruction.

Research Design: We used a randomized controlled trial (RCT) design with a total sample of 78 preschool classrooms. 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 (measure described below). 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 analysis.

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 of the tool is estimated to be at .94 as determined by Cronbach's alpha.

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 for preschool children. Classroom instruction is rated on a set of 30 statements (e.g., Teacher displayed an understanding of scientific concepts and phenomena). The individual items comprise seven scales: 1) Scientific Focus; 2) Setting the Stage for Inquiry; 3) Opportunity for Children to Explore Scientific Phenomena; 4) Meaning Making Experiences; 5) Planning; 6) Environment; and 7) Dosage.

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. Once the task starts, children are given no explicit reminders of the goal. In addition, if children pause after an attempt, we ascertain whether they think they have solved the challenge. Challenge tasks are a potentially rich source of information about how children deal with complex, temporally-extended tasks—in particular, those in which they must think about two dimensions in order to solve the problem. We measure whether children ultimately solve the challenge and whether the ultimate goal is remembered throughout the challenge trials. The *PAS'* internal consistency using Cronbach's alpha (α) is 0.727.

Analyses. We used multi-level modeling in order to assess the impact of *FSL* on children's spring *PAS* scores. In each model we entered a group of covariates at Level-1 including: fall score, gender, age, mother's educational level, and IEP-status. The main effect of Group (*FSL* or *non-FSL*) was examined and a set of additional moderating and mediating analyses were conducted. We also tested for the main effects of two classroom quality measures, *Science of Fidelity of Implementation Measure* and *STERS* on child outcomes.

Findings: There was a statistically significant main effect of Group on the average goal-remembered score. This finding indicates that the children in *FSL* classrooms were able to remember more goals, on average, than were children in control classrooms (B=0.128, $t(61)$=3.20, p=.003, effect size $\delta = 0.38$). (Please Insert Table 2 here)

A marginally significant interaction between Group and Gender on the average challenge-solved score was found. This finding suggests that boys in *FSL* classrooms were able to solve more challenges, on average, that were boys in control classrooms (B=0.154, $t(61)$=1.82, p=.073, effect size $\delta = 0.42$); in contrast, there was little difference between girls in *FSL* and control classrooms. (Please insert Table 3 here)

The statistically significant main effect of total Science Fidelity of Implementation Score on the average challenge-solved score indicates that the children in classrooms with higher fidelity to high-quality early science instruction, as embodied in FSL, were able to solve more challenges than children in classrooms with lower fidelity to high-quality science instruction ($B=0.099$, $t(36)=2.71$, $p=.011$, effect size $\delta = 0.28$). We also found a statistically significant main effect of total Fidelity Score on the average goal-remembered score ($B=0.090$, $t(36)=2.51$, $p=.017$, effect size $\delta = 0.25$. There was also a marginal main effect of total Fidelity Score on the total PAS score ($B=0.534$, $t(35)=1.75$, $p=.088$, effect size $\delta = 0.19$). (Please insert Tables 4-6 here)

Analyses also revealed a statistically significant main effect of STERS difference score on goal-remembered score ($B=0.069$, $t(54)=2.84$, $p=.007$, effect size $\delta = 0.20$). (Please insert Table 7 here)

Conclusions: Our results support the conclusion that exposure to *FSL* improves the ability of young children to remember the goals of their scientific inquiry. Findings also suggest that such exposure increases their ability to solve scientific challenges. We draw two broad implications from our results. First, the results clearly point to the relevance of contextual factors in understanding the mechanisms underlying early science development. These contextual factors include what are traditionally termed “moderators,” such as gender and level of maternal education, and “mediators,” such as the quality of science teaching in the classroom. Furthermore, the fact that classroom fidelity is a significant predictor of both of our focal variables (and is also marginally predictive of performance on the overall assessment) lends support to the efficacy of *FSL* as well as to the critical nature of implementation in this context. Indeed, in our experience many early childhood teachers have difficulty implementing an approach to early science instruction that emphasizes science knowledge in addition to inquiry skills.

A second broad implication stems from the complex relation between our focal variables, one of which taps the application of particular science knowledge in problem solving, and the other of which taps the ability of children to use their general cognitive capacity—executive function—to maintain a goal in mind. While the two constructs are clearly related in the context of *PAS* Challenge Tasks, the results suggest that more study is needed in order to understand how they interact as children grow in their capacity to solve scientific problems in an intentional way.

Appendices

Appendix A: References

Ball, D. L. (2000). Bridging practices: Intertwining content and pedagogy in teaching and learning to teach. *Journal of Teacher Education, 51*(3), 241-247.

Barnett, W. S. (2003). Better teachers, better preschools: Student achievement linked to teacher qualifications [Electronic version]. *Preschool Policy Matters*, (2), 1-12. Retrieved December 20, 2003, from http://nieer.org/resources/policybriefs/2.pdf

Bowman, B., Donovan, M. S., & Burns, M. S. (2001). *Eager to learn: Educating our preschoolers*. Washington, DC: National Academy Press.

Denton, K., & West, J. (2002). *Children's reading and mathematics achievement in kindergarten and first grade*. Retrieved October 23, 2007, from http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2002125

Dwyer, M. C., Chait, R., & McKee, P. (2000). *Building strong foundations for early learning: The U.S. Department of Education's guide to high-quality early childhood education programs*. Washington, DC: U.S. Department of Education, Planning and Evaluation Service.

Espinosa, L. M. (2002). High quality preschool: Why we need it and what it looks like [Electronic version]. *Preschool Policy Matters*, (1/November), 1-10. Retrieved July 16, 2003, from http://nieer.org/resources/policybriefs/1.pdf

Feiman-Nemser, S., & Parker, M. B. (1992). *Mentoring in context: A comparison of two U.S. programs for beginning teachers*. East Lansing, MI: National Center for Research on Teacher Learning, Michigan State University.

Fryer, R. G., & Levitt, S. D. (2006). The Black-White test score gap through third grade. *American Law and Economics Review, 8*(2), 249-281.

Harvard Family Research Project. (2004). Engaging with families in out-of-school time learning [Electronic version]. *Out-of-School Time Evaluation Snapshot*, *4*, 1-8

Helburn, S. W., & Bergmann, B. R. (2002). *America's child care problem: The way out*. New York: Palgrave.

Lee, V. E., & Burkam, D. T. (2002). *Inequality at the starting gate: Social background differences in achievement as children begin school*. Washington, DC: Economic Policy Institute.

Loucks-Horsley, S., Love, N., Stiles, K. E., Mundry, S., & Hewson, P. W. (2003). *Designing professional development for teachers of science and mathematics* (2nd ed.). Thousand Oaks, CA: Sage.

Myers, D., & Dynarski, M. (2003). *Random assignment in program evaluation and intervention research: questions and answers*. Retrieved April 18, 2007, from http://www.ed.gov/rschstat/eval/resources/randomqa.pdf

National Commission on Mathematics and Science Teaching for the 21st Century. (2000). *Before it's too late*. Washington, DC: U.S. Department of Education. Retrieved December 23, 2003, from http://www.ed.gov/inits/Math/glenn/report.pdf

National Science Foundation. (1999). *Women, minorities, and persons with disabilities in science and engineering: 1998* (NSF 99-87). Arlington, VA: Author. Retrieved June 18, 2008, from http://www.nsf.gov/statistics/nsf99338/pdfstart.htm

National Science Foundation. (2007). *Science and engineering degrees: 1966-2004*. Arlington, VA: Author. Retrieved June 18, 2008, from http://www.nsf.gov/statistics/nsf07307/pdf/nsf07307.pdf

Nelson, G. D. (1999). Science literacy for all in the 21st century [Electronic version]. *Educational Leadership*, *57*(2). Retrieved December 20, 2003, from http://www.project2061.org/research/articles/ascd.htm

Penuel, W. R., & Gallagher, L. P. (2008). *Comparing three approaches to preparing teachers to teach for deep understanding in Earth science: Short-term impacts on teachers and teaching practice*. Menlo Park, CA: SRI International. Retrieved June 16, 2008, from http://ctl.sri.com/publications/downloads/TIDES_ComparingThreeApproaches_EarthScience.pdf

Puma, M., Bell, S., Shapiro, G., Broene, P., Cook, R., Friedman, J., et al. (2001). *Building futures: The Head Start Impact Study. Research design plan*. Retrieved April 18, 2007, from http://www.acf.hhs.gov/programs/opre/hs/impact_study/reports/impact_study/impactstdy_resrch_plan.pdf

Rutherford, F. J., & Ahlgren, A. (1989). *Science for all Americans*. New York: Oxford University Press.

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

Shulman, L. S. (1987). Knowledge and teaching: Foundations for the new reform. *Harvard Educational Review, 57*, 1-22.

The White House. (2010, September 16). *President Obama to announce major expansion of "Educate to Innovate" campaign to improve science, technology, engineering and Math (STEM) education*. Retrieved October 1, 2010, from http://www.whitehouse.gov/the-press-office/2010/09/16/president-obama-announce-major-expansion-educate-innovate-campaign-impro

Whitebook, M. (2003). *Early education quality: Higher teacher qualifications for better learning environments - a review of the literature*. Berkeley, CA: University of California, Berkeley. Retrieved December 20, 2003, from http://iir.berkeley.edu/cscce/pdf/teacher.pdf

Wood, F. H., & McQuarrie, F., Jr. (1999). On-the-job learning. *Journal of Staff Development, 20*(3), 10-13.

Appendix B: Tables and Figures

Table 1. Schedule of *FSL* Sessions and Coaching

SESSION	PURPOSE / TOPICS
Course Session 1	**Introduction; Forces that influence water flow**
Mentor Meeting 1	Facilitating 'Open Exploration'; Creating environments that invite inquiry
Course Session 2	**Focused investigations of matter and force: Water flow and Drops**
Mentor Meeting 2	Facilitating 'Engage and Explore'; Plan 'Focused Explorations'
Course Session 3	**More Focused Investigations of Matter and Force: Sinking and Floating**
Mentor Meeting 3 (group)	Supporting understanding of Session 3 science, and planning to promote inquiry & learning
Mentor Meeting 4 (group)	Building understanding of children's theories through representation; Engaging in reflection on teaching & children's learning
Course Session 4	**Matter and Force with Balls and Ramps**
Mentor Meeting 5 (group)	Supporting understanding of Session 4 science; Analyzing children's 'Open Exploration'; Plan focused E-E-R cycle
Mentor Meeting 6 (group)	Supporting understanding of science and teaching of Balls and Ramps, and the reflective process with children

Table 2. HLM Regression Models Examining the Effect of Group on Average PAS Goal-Remembered Score

Effect	Model 1 $Outcome_{jk} = B_{0k} + B_1 Covariates + r_{jk}$ $B_{0k} = \gamma_{00} + u_{0k}$				Model 2 $Outcome_{jk} = B_{0k} + B_1 Covariates + r_{jk}$ $B_{0k} = \gamma_{00} + \gamma_{01} FSL + u_{0k}$				
	B	***Se***	***t***	***df***	***B***	***Se***	***t***	***df***	***δ***
Intercept	0.284***	0.039	7.20	62	0.203***	0.046	4.41	61	
Group					0.128**	0.040	3.20	61	0.38
$Var(r_{jk})$	0.104				0.104				
$Var(u_{0k})$	0.014				0.012				

p<.001, **p < .01, *p < .05; *B* = unstandardized regression coefficient. ***δ is defined as the ratio between the regression coefficient for FSL and the standard error of the outcome (Liu, Spybrook, Congdon, Martinez, & Raudenbush, 2006) . ***δ*** = γ_{01}/Sqrt [$Var(r_{jk}) + Var(u_{0k})$] for the size of the group effect on the intercept.

Table 3. HLM Regression Models Examining the Interaction between Group and Gender on the average PAS Challenge-Solved Score

Effect	Model 1 $Outcome_{jk} = B_{0k} + B_1 Gender + r_{jk}$ $B_{0k} = \gamma_{00} + u_{0k}$ $B_{1k} = \gamma_{10} + u_{1k}$				Model 2 $Outcome_{jk} = B_{0k} + B_1 Gender + r_{jk}$ $B_{0k} = \gamma_{00} + \gamma_{01} FSL + u_{0k}$ $B_{1k} = \gamma_{10} + \gamma_{11} FSL + u_{1k}$				
	B	***Se***	***t***	***df***	***B***	***Se***	***t***	***df***	***δ***
Intercept	0.437***	0.043	10.18	62	0.469***	0.057	8.21	61	
Group					-0.051	0.059	-0.85	61	0.14
Gender	-0.021	0.043	-0.49	62	-0.111	0.059	-1.87	61	0.31
Group×Gender					0.154	0.084	1.82	61	0.42
$Var(r_{jk})$	0.110				0.111				
$Var(u_{0k})$	0.021				0.022				
$Var(u_{1k})$	0.026				0.021				

p<.001, **p < .01, *p < .05; *B* = unstandardized regression coefficient. ***δ is defined as the ratio between the regression coefficient for FSL and the standard error of the outcome (Liu, Spybrook, Congdon, Martinez, & Raudenbush, 2006) . ***δ*** = γ_{01}/Sqrt [$Var(r_{jk}) + Var(u_{0k})$] for the size of the group effect on the intercept. ; ***δ*** = γ_{11}/Sqrt [$Var(r_{jk}) + Var(u_{1k})$] for the size of the Group × Gender Interaction.

Table 4. HLM Regression Models Examining the Effect of Fidelity Score on Average PAS Challenge-Solved Score

Effect	Model 1 $Outcome_{jk} = B_{0k} + B_1 Covariates + r_{jk}$ $B_{0k} = \gamma_{00} + u_{0k}$				Model 2 $Outcome_{jk} = B_{0k} + B_1 Covariates + r_{jk}$ $B_{0k} = \gamma_{00} + \gamma_{01} Fidelity + u_{0k}$				
	B	***Se***	***t***	***df***	***B***	***Se***	***t***	***df***	***δ***
Intercept	0.509***	0.051	10.045	37	0.503***	0.051	9.876	36	
Fidelity					0.099*	0.036	2.714	36	0.28
$Var(r_{jk})$	0.105				0.107				
$Var(u_{0k})$	0.014				0.015				

p<.001, **p < .01, *p < .05; *B* = unstandardized regression coefficient. ***δ is defined as the ratio between the regression coefficient for Fidelity and the standard error of the outcome (Liu, Spybrook, Congdon, Martinez, & Raudenbush, 2006) . ***δ*** = γ_{01}/Sqrt [$Var(r_{jk}) + Var(u_{0k})$] for the size of the fidelity effect on the intercept.

Table 5. HLM Regression Models Examining the Effect of Fidelity Score on Average PAS Goal- Remembered Score

Effect	Model 1 $Outcome_{jk} = B_{0k} + B_1 Covariates + r_{jk}$ $B_{0k} = \gamma_{00} + u_{0k}$				Model 2 $Outcome_{jk} = B_{0k} + B_1 Covariates + r_{jk}$ $B_{0k} = \gamma_{00} + \gamma_{01} Fidelity + u_{0k}$				
	B	***Se***	***t***	***df***	***B***	***Se***	***t***	***df***	***δ***
Intercept	0.377***	0.054	6.933	37	0.363***	0.055	6.628	36	
Fidelity					0.090*	0.036	2.514	36	0.25
$Var(r_{jk})$	0.104				0.107				
$Var(u_{0k})$	0.026				0.028				

p<.001, **p < .01, *p < .05; *B* = unstandardized regression coefficient. ***δ is defined as the ratio between the regression coefficient for Fidelity and the standard error of the outcome (Liu, Spybrook, Congdon, Martinez, & Raudenbush, 2006) . ***δ*** = γ_{01}/Sqrt [$Var(r_{jk}) + Var(u_{0k})$] for the size of the fidelity effect on the intercept.

Table 6. HLM Regression Models Examining the Effect of Fidelity Score on Total PAS Score

Effect	Model 1 $Outcome_{jk} = B_{0k} + B_1 Covariates + r_{jk}$ $B_{0k} = \gamma_{00} + u_{0k}$				Model 2 $Outcome_{jk} = B_{0k} + B_1 Covariates + r_{jk}$ $B_{0k} = \gamma_{00} + \gamma_{01} Fidelity + u_{0k}$				
	B	***Se***	***t***	***df***	***B***	***Se***	***t***	***df***	***δ***
Intercept	7.053***	0.422	16.715	36	7.012***	0.427	16.432	35	
Fidelity					0.534	0.305	1.752	35	0.19
$Var(r_{jk})$	7.263				7.309				
$Var(u_{0k})$	0.809				0.844				

p<.001, **p < .01, *p < .05; *B* = unstandardized regression coefficient. ***δ is defined as the ratio between the regression coefficient for Fidelity and the standard error of the outcome (Liu, Spybrook, Congdon, Martinez, & Raudenbush, 2006) . ***δ*** = γ_{01}/Sqrt [$Var(r_{jk}) + Var(u_{0k})$] for the size of the fidelity effect on the intercept.

Table 7. HLM Regression Models Examining the Effect of STERS difference score on Average PAS Goal-Remembered Score

Effect	Model 1 $Outcome_{jk} = B_{0k} + B_1 Covariates + r_{jk}$ $B_{0k} = \gamma_{00} + u_{0k}$				Model 2 $Outcome_{jk} = B_{0k} + B_1 Covariates + r_{jk}$ $B_{0k} = \gamma_{00} + \gamma_{01} STERSdiff + u_{0k}$				
	B	***Se***	***t***	***df***	***B***	***Se***	***t***	***df***	***δ***
Intercept	0.280***	0.043	6.521	55	0.287***	0.043	6.714	54	
STERSdiff					0.069**	0.024	2.843	54	0.20
$Var(r_{jk})$	0.103				0.102				
$Var(u_{0k})$	0.017				0.017				

p<.001, **p < .01, *p < .05; *B* = unstandardized regression coefficient. Both models also include level-1 covariates, not shown here, for IEP status and level of maternal education. ***δ is defined as the ratio between the regression coefficient for the STERS difference score and the standard error of the outcome (Liu, Spybrook, Congdon, Martinez, & Raudenbush, 2006) . ***δ*** = γ_{01}/Sqrt [$Var(r_{jk}) + Var(u_{0k})$] for the size of the STERS difference effect on the intercept.