Abstract Title Page

Title:

Creating a Successful Professional Development Program in Science for Head Start Teachers and Children: Understanding the Relationship between Development, Intervention, and Evaluation

Author(s):

Jess Gropen, PhD, Education Development Center; Nancy Clark-Chiarelli, EdD, Education Development Center; Ingrid Chalufour, MA, Education Development Center; Cindy Hoisington, MA, Education Development Center; Costanza Eggers-Piérola, EdD, Education Development Center

Abstract Body

Background/context:

While it has long been recognized that there is a clear and compelling need to foster young children's science literacy (B. T. Bowman, 1999; New, 1999; Hecker, 2001), our P–16 educational system is failing to provide students with a solid understanding of science and the capacity to succeed in tomorrow's labor market. Students are not demonstrating rapid progress in science achievement (National Center for Education Statistics, 2001), particularly when compared with other countries (Martin et al., 2000). Furthermore, an achievement gap in science persists, with children of color, some children who are English language learners, and children of low-income backgrounds demonstrating lesser science proficiency than their peers (Haycock, Jerald, & Huang, 2001; National Center for Education Statistics, 2001; Robelen, 2002).

To foster its future workforce's science literacy, the United States needs to improve science education for *all* children, at every grade level (Haycock et al., 2001; National Center for Education Statistics, 2001; Nelson, 1999; Singham, 2003; Sutman, 2001). In David Hawkins'(Hawkins, 1983) view, science can be "the great equalizer" when it is made more accessible by basing curriculum on everyday topics that are familiar to all children, regardless of their backgrounds. Because differences in children's achievement are apparent before they enter kindergarten (Shonkoff & Phillips, 2000), realizing science's potential for educational equity can bridge the achievement gap that begins in preschool.

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 (Dwyer, Chait, & McKee, 2000; Espinosa, 2002; Helburn & Bergmann, 2002; B. Bowman et al., 2001). Yet, few models of professional development build teachers' skills and knowledge in an ongoing way and provide access to higher education credits. Often, professional development consists of episodic workshops that do not reflect research-based knowledge about effective learning (Bransford, Brown, & Cocking, 1999; Darling-Hammond, 1996; Gallagher & Clifford, 2000; Hyson, 2001; Miller, Lord, & Dorney, 1994; Morgan et al., 1993) or build on teachers' current practice (Darling-Hammond, 1996; Morgan et al., 1993). Without ongoing feedback and content-focused mentoring, it is difficult for teachers to sustain changes in practice (Caruso & Fawcett, 1999; Darling-Hammond, 1996; Garet, Porter, Desimone, Birman, & Yoon, 2001).

Over the past three years, our team at Education Development Center, Inc. (EDC) has been researching a professional development program in science, *Foundations of Science Literacy* (*FSL*), for preschool lead and assistant Head Start teachers in Massachusetts and Rhode Island. Year 1 was a pilot year, so it is the data from Year 2 on which we report in this paper. *Foundations of Science Literacy* is designed to respond to the urgent call to prepare preschoolers for tomorrow and respond to Hawkins's eloquent plea for equity.

Purpose/objective/research question/focus of study:

Our research is designed to answer two important questions germane to this paper: 1) Does *FSL* impact Head Start teachers' practices in inquiry-based science instruction for four-year-old children? 2) Does *FSL* impact Head Start children's early science knowledge and skills? The objective of the presentation will be to report the principal teacher-, classroom-, and child-level findings from our Year 2 implementation, and to discuss these findings in the context of what makes for a successful professional development program in early science.

Setting:

The research took place in the metro-west and southern sections of Massachusetts.

Population/Participants/Subjects:

Working with five Head Start programs in Massachusetts, we recruited lead and assistant teachers from 50 classrooms to participate in the study. Within each program 60% of the classrooms were randomly assigned to the *FSL* intervention group and 40% to the control group. A description of the recruitment and analytic samples can be found in Table 1. (Please insert table 1 here).

Intervention/Program/Practice:

FSL has two main components: 1) instructional sessions that are conducted face-to-face and designed to build teachers' content knowledge in specific concepts in physical science and enhance their ability to teach science to young children; and 2) a mentoring component that provides coaching support to teachers as they master science content and implement inquiry-based science methods. Based on our experience training early childhood teachers, we deliver *FSL* over a six-month period. We know that it is essential to expand the timeframe for coursework beyond that typically allotted by institutions of higher education. Doing so paces the learning experience, allowing teachers to digest and apply new material while continuing to meet their job obligations (Dickinson & Brady, 2004). Initial sessions concentrate more of the instructional content on teaching teachers science content using an inquiry-based approach. As the program progresses, sessions focus more on the content and pedagogy appropriate for young children.

In addition, *FSL* includes a set of three key design features. First, teachers learn best when they see examples of the practices they are adopting. Videotape exemplars, coupled with teacher commentary, build teachers' capacity to analyze and reanalyze the effectiveness of practices in light of children's responses. Powerful vehicles for showing teacher-child and child-to-child interaction, they demonstrate the complex interactions among instruction, assessment, and children's learning. As "pictures of practice," they demystify how to introduce investigations, conduct rich discussions, and identify and work with children's naïve theories. They also build teachers' ability to engage in focused, professional dialogues. Second, young children express their science understandings and questions through conversations, drawings, narratives, and play. Yet, many teachers are not aware of the assessment opportunities that these sources of data provide. In *FSL*, we provide teachers with children's work samples that illustrate a range of understanding and a diversity of modes of expression. Such samples provide teachers with the experience they need to assess children's learning and prepare responsive curriculum activities that challenge children's thinking. Using work from their classrooms helps teachers move from "abstract" analysis, in which the children and classroom are unknowns, to "authentic" analysis in which the hypotheses they generate can be tested and reported on. Third, performance tasks that elicit what teachers know and are able to do help guide teachers' mastery of key concepts and strategies and assist course architects and instructors in evaluating the impact of teaching and learning events (Brady & Chalufour, 2004). In *FSL*, assignments are carefully sequenced to build a bridge between instructional sessions and teachers' classrooms. Participants are required to carry out application activities, set goals to improve their practice, and analyze the effectiveness of their teaching in terms of children's science learning, development, and engagement. All assignments center on children's work and/or videotapes of teachers' practices to provide direct evidence of classroom practices that allow us to evaluate teachers' learning.

Research Design:

Recruitment of Head Start teachers and assistant teachers was conducted in the summer and fall of 2006. We worked with program directors in Massachusetts to recruit teachers in their respective programs and centers. Our recruitment yielded 50 classrooms and 66 teachers (50 lead teachers and 16 assistant teachers). Teachers and assistant teachers interested in participating in the project, across the programs, were randomly assigned to one of two conditions: *FSL* Intervention and Control. We stratified by Head Start program, and classrooms were randomly assigned to one of the conditions. Furthermore, the randomized sample was not balanced for numbers of classrooms in the intervention and control groups (Myers & Dynarski, 2003). In particular, 60% of the classrooms were assigned to the intervention group, and 40% were assigned to the control group. This design 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). Moreover, an imbalanced sample "reduces the precision of the impact estimates by just 2%" (p. 33) when they employ a 60:40 ratio (Puma et al., 2001).

Data Collection and Analysis:

Because no existing instruments were available at the inception of this project to measure teachers' content knowledge, the quality of classroom science instruction, or young children's performance-based knowledge, we created three tools to measure these constructs. These newly developed instruments are described below. In addition, we also assessed global classroom quality using the *Early Childhood Environmental Rating Scale-Revised* (*ECERS-R*).

The *Science Teaching and Environment Observation Rating Scale* (*STERS*) is a classroom observation tool consisting of a framework for classroom observation and a teacher interview. The observation framework consists of five items corresponding to dimensions of quality science instruction in preschool classrooms: 1) Create a Physical Environment for Inquiry and Learning; 2) Facilitate Direct Experiences to Promote Conceptual Learning; 3) Promote Use of Scientific Inquiry; 4) Plan In-depth Investigations; 5) Assess Children's Learning. Each of the items is rated on a 4-point rating scale where "1" corresponds to *Inadequate* and "4" corresponds to *Exemplary*. The internal consistency of the *STERS* is high (Chronbach's alpha=.96). Six data collectors were trained by EDC staff to collect fall data and eight were trained in the spring. Training included an introduction to the *STERS*, and instructions for scoring each dimension based on classroom observation and teacher interview. Each data collector was accompanied by the trainer during the first classroom observation. Both the trainer and the data collector observed and scored the same classroom. Their scores were calibrated to ensure reliable scoring and consistency across data collectors.

To assess teacher's content and pedagogical knowledge, we created a set of four *Science Teacher Performance Tasks* (*TPT*s). These tasks were designed to assess a teacher's ability to: plan science curriculum, including both content and inquiry components (Planning a Science Experience); evaluate a child's science understanding based on his representation (Interpreting a Child's Work Sample) and his behavior during an exploration of water (Analyzing Misconceptions); and analyze teacher facilitation of a science exploration (Analysis of Science Teaching). We assessed teacher's performances based on their written analysis or explanation in response to a common prompt (e.g., video, child's work sample). Each of the responses are rated on a 4-point scale where "1" corresponds to *Little or no evidence of knowledge* and "4" corresponds to *Clear, consistent, and convincing evidence of knowledge*. The *TPT*s' reliability is .82 as measured by *Chronbach's alpha*.

As part of this effort, we also developed the *Preschool Assessment of Science* (*PAS*)—a performance-based instrument aimed at uncovering how 4-year-olds think about matter and the forces that act on matter. For example, what do young children think about the way that water naturally flows, or about how the direction and rate of flow may be changed by acting on the water (e.g., by using a squeeze bottle to expel water forcefully). Or, how far do they think a ball will travel after it rolls down a ramp, and do they think that the distance depends on the weight of the ball and the slope of the ramp? The *PAS* is organized in three main tasks: *Water Flow* (*WF*), *Marbles & Ramps* (*MR*), and *Floating & Sinking* (*FS*). Based on a reliability analysis, we separated the items in the *Floating & Sinking* task into two scales—one involving definitions and explanations (FS_{verbal}) and the other involving predictions in a sorting activity (FS_{sorting}). The other two *PAS* tasks were each represented by a separate scale.

For data analysis, we constructed a pair of regression models for each measure by first regressing the post-test score (obtained in spring) on the baseline pre-test score (obtained in fall), and then adding in the predictor of Group (*FSL* vs. control). Models were built using standard OLS for teacher/classroom measures and hierarchical linear modeling (HLM) for child measures. During the model building process, we also examined the impact of additional predictors as control or moderator variables. For each model, we present coefficients, standard errors, and indices of effect size, including R^2 and δ (δ is defined as the ratio between the regression coefficient for Group and the standard error of the outcome; Liu, Spybrook, Congdon, Martinez, & Raudenbush, 2006).

Findings/Results:

Teacher Outcomes. Group (*FSL* vs. control) was a significant predictor of spring *Teacher Performance Tasks* (*TPTs*) [$t(53) = 6.44$, $p < .001$, $R^2 = .34$, $\delta = 1.75$]. On average, *FSL* teachers scored 0.7 points higher than control teachers on the *TPT*s (for which scores range from 1-4). (Please insert table 2 here).

Classroom Outcomes. *FSL* classrooms showed stronger outcomes than control classrooms on both the *ECERS-R* and *STERS*. On the *ECERS-R*, we found statistically significant outcome differences between *FSL* and control classrooms on the *Language-Reasoning* (*LR*) subscale

[t(39) = 2.11, $p < .05$, $R^2 = .09$, $\delta = 0.68$], with ratings (which range from 1-7) averaging 0.8 points higher in *FSL* than in control classrooms. On the *STERS*, group differences were even more pronounced, reflecting the close alignment of the *STERS* with the focus of the *FSL* intervention on supporting science teaching and learning. In particular, *STERS* ratings (which range from 1-4) were 1.6 points higher in *FSL* than in control classrooms, a highly significant difference [t(39) = 9.43, $p < .001$, $R^2 = .63$, $\delta = 3.00$]. Interestingly, statistically significant correlations were found between *TPT* scores and classroom ratings [$r = .36$, $p < .05$ for *ECERS-LR* and $r = .57$, $p < .05$ for *STERS*], which is evidence that the *TPT* could be used as a proxy for classroom practice. (Please insert table 3 here).

Child Outcomes. Statistically significant positive effects for *FSL* were found for two blocks of *PAS* items having content that was most heavily emphasized in the intervention. In particular, on the *WF* block involving how force affects the direction and rate of water flow, containing nine items (scored 0-9 overall), *FSL* children averaged 0.6 points higher than did control children [t(43) = 2.36, $p < .05$, $R^2 = .08$, $\delta = .37$]. Similarly, *FSL* children scored 0.4 points higher than control children on the principal *MR* block [t(43) = 2.01, $p = .05$, $R^2 = .04$, $\delta = .28$], which includes six items (scored 0-6 overall) on how the speed of a ball rolling down a ramp, and the distance it travels, can be altered by changes in the slope of the ramp. *FSL* children also tended to have higher scores than control children in spring on the *FS* scales, but these effects were not statistically significant [$\delta = .20$ for FS_{verbal} and $\delta = .09$ for $FS_{sorting}$]. (Please insert table 4 here).

Figure 1 summarizes the important *pattern* of results on the *PAS* scales—namely, that students in *FSL* classrooms tended to show greater improvement in *PAS* scores compared with students in control classrooms. This pattern remained evident after controlling for other baseline child variables (e.g., gender, ELL status, and baseline performance on standardized measures). Furthermore, there was also evidence that improvements in classroom practices from fall to spring were associated with improvements in children's performance on the *PAS*. For example, controlling for fall scores, spring *STERS* scores were positively correlated with spring classroom mean *WF* [$r = .33$, p < .05)] and FS_{verbal} [$r = .33$, p < .05] scores. (Please insert figure 1 here)

Conclusions:

Our results indicate that *FSL* had a strong impact on teachers' science knowledge and classroom practices in inquiry-based science instruction, and that children in *FSL* classrooms showed a trend towards greater improvement in their understanding of basic physical science principles and their use of science inquiry skills. In addition, our findings have lead to specific revisions in *FSL*, including the use of more constrained classroom assignments, allowing for a better alignment between *FSL* and *PAS* content, and a stronger emphasis on the growth of children's reflective capacity in the context of *FSL*.

We conclude that successful development of professional development programs requires evaluation at every level in the "causal chain" that connects teaching and learning—from teacher's content knowledge, to their ability to apply it in their classrooms, to children's ability to engage in focused science activities with genuine conceptual content.

Appendixes

Appendix A. References

Barnett, W. S. (1998). Long-term effects on cognitive development and school success. In W. S. Barnett & S. S. Boocock (Eds.), *Early care and education for children in poverty: Promises, programs, and long-terms results* (pp. 11-44). New York: State University of New York Press.

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.

Bowman, B. T. (1999). *Policy implications for math, science, and technology in early childhood education*. Retrieved February 28, 2003, from http://www.project2061.org/tools/earlychild/context/bowman.htm

Brady, J., & Chalufour, I. (2004, June). *Performance assessment through assignments in credit-bearing professional development*. Paper presented at the National Association for the Education of Young Children's Institute for Early Childhood Professional Development, Baltimore, MD.

Bransford, J. D., Brown, A. L., & Cocking, R. R. (Eds.). (1999). *How people learn: Brain, mind, experience, and school*. Washington, DC: National Academy Press.

Caruso, J. J., & Fawcett, M. T. (1999). *Supervision in early childhood education: A developmental perspective* (2nd ed.). New York: Teachers College Press.

Cho, H., Kim, J., & Choi, D. H. (2003). Early childhood teachers' attitudes toward science teaching: a scale validation study. *Educational Research Quarterly*, *27*(2), 33-42.

Coley, R. J. (2002). *An uneven start: Indicators of inequality in school readiness*. Princeton, NJ: Educational Testing Services.

Copley, J., & Padrón, Y. (1999). Preparing teachers of young learners: Professional development of early childhood teachers in mathematics and science. In *Dialogue on early childhood science, mathematics, and technology education* (pp. 117-129). Washington, DC: American Association for the Advancement of Science.

Darling-Hammond, L. (1996). The quiet revolution: Rethinking teacher development. *Educational Leadership*, *63*(6), 4-10.

Denton, D. (2001). *Improving children's readiness for school: Preschool programs make a difference, but quality counts!* Atlanta, GA: Southern Regional Education Board.

Dickinson, D., & Brady, J. (2004). The role of preschool classrooms in supporting early literacy development. In M. Zaslow & I. Martinez-Beck (Eds.), *Early childhood professional development and children's successful transition to elementary school*. Baltimore: Brookes.

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
Gallagher, J., & Clifford, R. (2000). The missing support infrastructure in early childhood [Electronic version]. *Early Childhood Research & Practice*, *2*(1 Spring). Retrieved August 28, 2006, from http://ecrp.uiuc.edu/v2n1/gallagher.html
Garet, M. S., Porter, A. C., Desimone, L., Birman, B. F., & Yoon, K. S. (2001). What makes professional development effective? Results from a national sample of teachers. *American Educational Research Journal*, *38*(Winter), 915-945.
Hawkins, D. (1983). Nature closely observed. *Daedalus*, *112*(2), 65-89.
Haycock, K., Jerald, C., & Huang, S. (2001). Closing the gap: Done in a decade [Electronic version]. *Thinking K-16*, *5*(2), 3-21. Retrieved December 22, 2003, from http://www2.edtrust.org/NR/rdonlyres/85EB1387-A6B7-4AF4-BEB7-DF389772ECD2/0/K16_spring01.pdf
Hecker, D. E. (2001). Occupational employment projections to 2010 [Electronic version]. *Monthly Labor Review*, *124*(11), 57-84. Retrieved December 19, 2003, from http://www.bls.gov/opub/mlr/2001/11/art4full.pdf
Helburn, S. W., & Bergmann, B. R. (2002). *America's child care problem: The way out*. New York: Palgrave.
Hyson, M. (2001). Better futures for young children, better preparation for their teachers: Emerging from recent national reports. *Young Children*, *56*, 60-62.
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.
Martin, M. O., Mullis, I. V. S., Gonzalez, E. J., Gregory, K. D., Smith, T. A., Chrostowski, S. J., et al. (2000). *International science report*. Chestnut Hill, MA: The International Study Center, Boston College, Lynch School of Education. Retrieved December 22, 2003, from http://isc.bc.edu/timss1999i/pdf/T99i_Sci_All.pdf
Miller, B., Lord, B., & Dorney, J. (1994). *Staff development for teachers: A study of configurations and sustaining innovations* (Summary Rep.). Newton, MA: Education Development Center, Inc.
Morgan, G., Azer, S. L., Costley, J. B., Genser, A., Goodman, I. F., & Lombardi, J. (1993). *Making a career of it: The state of the states report on career development in early care and education*. Boston: The Center for Career Development in Early Care and Education at Wheelock College.
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 Center for Education Statistics. (2001). *The nation's report card: Science highlights 2000* (NCES 2002-452). Washington, DC: Author. Retrieved December 20, 2003, from http://nces.ed.gov/nationsreportcard/pdf/main2000/2002452.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

New, R. S. (1999). *Playing fair and square: Issues of equity in preschool mathematics, science, and technology*. Retrieved August 25, 2003, from http://www.project2061.org/tools/earlychild/fostering/new.htm

Peisner-Feinberg, E. S., Burchinal, M. R., Clifford, R. M., Culkin, M. L., Howes, C., Kagan, S. L., et al. (2001). The relation of child-care quality to children's cognitive and social developmental trajectories through second grade. *Child Development*, *72*(5), 1534-1553.

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

Ramey, C. T., & Ramey, S. L. (2003, February). *Preparing America's children for success in school*. Paper presented at the annual meeting of the National Governors Association, Washington, DC.

Robelen, E. W. (2002). *Taking on the achievement gap*. Retrieved December 22, 2003, from http://www.ncrel.org/gap/takeon/toc.htm

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

Shore, R. (1997). *Rethinking the brain: New insights into early development.* New York: Families and Work Institute.

Singham, M. (2003). The achievement gap: Myths and reality [Electronic version]. *Phi Delta Kappan*, *84*(8), 586-591. Retrieved December 22, 2003, from http://www.noycefdn.org/literacy/ecrwfiles/ECRWPrincipals/Achievement%20Gap.pdf

Smith, M. W., & Dickinson, D. K. (1994). Describing oral language opportunities and environments in Head Start and other preschool classrooms. *Early Childhood Research Quarterly, 9*, 345-366.

Sutman, F. X. (2001). Mathematics and science literacy for all Americans [Electronic version]. *ENC Focus*, *8*(3). Retrieved December 20, 2003, from http://www.enc.org/topics/realworld/literacy/document.shtm?input=FOC-002063-index

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

Appendix B. Tables and Figures

Table 1

Recruitment and Analytic Samples for FSL Pilot Study

	Recruitment Sample			Analytic Sample[1]		
	# classrooms	# teachers	# children	# classrooms	# teachers	# children
***FSL* group**	32	42	279	26	33	208
Control group	18	24	191	17	23	130
	50	66	470	43	56	338

[1]Defined as number of cases with data available on at least one post-test outcome measure

Table 2

Regression Models Examining the Effect of Group on Teacher Performance Task Outcomes

	Model 1				Model 2					
	B	***Se***	***b***	***t***	***B***	***Se***	***b***	***t***	δ	f^2
Intercept	0.823**	0.282		2.919	0.447*	0.221		2.020		
Fall Score	0.721**	0.185	0.469	3.906	0.680**	0.140	0.443	4.868		
Group					0.742**	0.115	0.586	6.438	1.75	0.52
R^2	0.220				0.562					
MSE	0.314				0.179					

**p < .01, *p < .05; *B* = unstandardized regression coefficient; *b* = standardized coefficient; *MSE* = Regression mean square error; $\delta = B_{Group}/\text{Sqrt}(MSE)$; $f^2 = \Delta R^2/(1-\Delta R^2)$, where $\Delta R^2 = R^2_{\text{Model 1}} - R^2_{\text{Model 2}}$.

Table 3

Regression Models Examining the Effect of Group on Classroom Outcomes

	Model 1				Model 2					
ECERS-LR	***B***	***Se***	***b***	***t***	***B***	***Se***	***B***	***t***	δ	f^2
Intercept	3.469**	0.950		3.653	3.258**	0.917		3.553		
Fall Score	0.424*	0.197	0.322	2.148	0.368	0.191	0.279	1.926		
Group					0.766*	0.363	0.306	2.108	0.68	0.10
R^2	0.103				0.195					
MSE	1.393				1.283					
ECERS-A	***B***	***Se***	***b***	***t***	***B***	***Se***	***B***	***t***	δ	f^2
Intercept	2.308*	1.059		2.179	2.206*	1.041		2.120		
Fall Score	0.609*	0.251	0.358	2.428	0.562*	0.248	0.331	2.269		
Group					0.480	0.299	0.234	1.605	0.51	0.06
R^2	0.128				0.182					
MSE	0.909				0.874					
ECERS-I	***B***	***Se***	***b***	***t***	***B***	***Se***	***B***	***t***	δ	f^2
Intercept	4.544**	0.862		5.270	4.397**	0.880		4.999		
Fall Score	0.265	0.154	0.262	1.717	0.259	0.155	0.256	1.673		
Group					0.290	0.323	0.138	0.899	0.29	0.02
R^2	0.069				0.088					
MSE	1.024				1.029					
STERS	***B***	***Se***	***b***	***t***	***B***	***Se***	***B***	***t***	δ	f^2
Intercept	1.978**	0.453		4.367	1.159**	0.268		4.325		
Fall Score	0.457	0.233	0.296	1.963	0.361**	0.131	0.234	2.764		
Group					1.609**	0.171	0.799	9.426	3.00	1.73
R^2	0.088				0.722					
MSE	0.917				0.287					

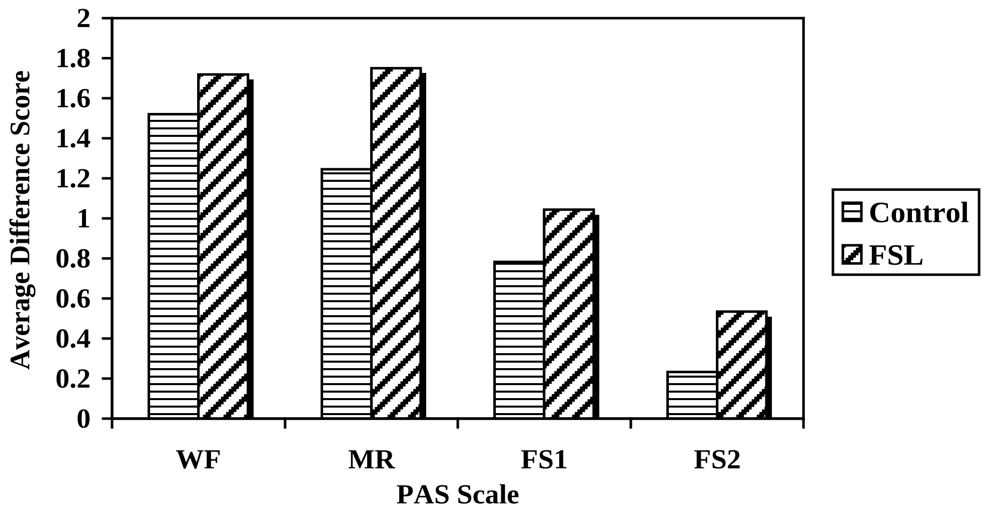
$^{**}p < .01$, $^{*}p < .05$; *B* = unstandardized regression coefficient; *b* = standardized coefficient; *MSE* = Regression mean square error; $\delta = B_{Group}/\mathrm{Sqrt}(MSE)$; $f^2 = \Delta R^2/(1-\Delta R^2)$, where $\Delta R^2 = R^2_{\text{Model 1}} - R^2_{\text{Model 2}}$.

Table 4

HLM Regression Models Examining the Effect of Group on PAS Child Outcomes

PAS Scale	Effect	**Model 1** $SpringPAS_{jk} = B_{0k} + B_1 FallPAS + r_{jk}$ $B_{0k} = \gamma_{00} + u_{0k}$				**Model 2** $SpringPAS_{jk} = B_{0k} + B_1 FallPAS + r_{jk}$ $B_{0k} = \gamma_{00} + \gamma_{01} FSL + u_{0k}$				
		B	***Se***	***t***	***df***	***B***	***Se***	***t***	***df***	δ
Water Flow	γ_{00}	4.494^{**}	0.383	11.73	44	4.251^{**}	0.429	9.92	43	
	B_1	0.576^{**}	0.051	11.32	260	0.571^{**}	0.051	11.19	259	
	γ_{01}					0.451	0.357	1.27	43	0.19
	$Var(r_{jk})$	5.421				5.408				
	$Var(u_{0k})$	0.357				0.358				
Water Flow -Block 2	γ_{00}	3.204^{**}	0.233	13.78	44	2.846^{**}	0.272	10.46	43	
	B_1	0.461^{**}	0.053	8.76	260	0.457^{**}	0.052	8.77	259	
	γ_{01}					0.598^{*}	0.253	2.36	43	0.37
	$Var(r_{jk})$	2.371				2.359				
	$Var(u_{0k})$	0.293^{**}				0.241^{*}				
Marbles & Ramps	γ_{00}	4.189^{**}	0.231	18.11	44	3.887	0.275	14.13	43	
	B_1	0.328^{**}	0.050	6.56	268	0.330	0.050	6.64	267	
	γ_{01}					0.480	0.241	1.99	43	0.29
	$Var(r_{jk})$	2.508				2.501				
	$Var(u_{0k})$	0.205^{*}				0.173^{*}				
Marbles & Ramps -Block 1	γ_{00}	3.230^{**}	0.163	19.86	44	3.000^{**}	0.197	15.19	43	
	B_1	0.326^{**}	0.047	6.88	268	0.329^{**}	0.047	6.98	267	
	γ_{01}					0.362^{*}	0.180	2.01	43	0.28
	$Var(r_{jk})$	1.666				1.656				
	$Var(u_{0k})$	0.069				0.057				
Floating & Sinking -Verbal	γ_{00}	1.708^{**}	0.174	9.82	44	1.506^{**}	0.233	6.45	43	
	B_1	0.618^{**}	0.063	9.74	269	0.615^{**}	0.063	9.69	268	
	γ_{01}					0.331	0.257	1.29	43	0.20
	$Var(r_{jk})$	2.386				2.383				
	$Var(u_{0k})$	0.268^{**}				0.263^{**}				
Floating & Sinking -Sorting	γ_{00}	5.628^{**}	0.369	15.27	44	5.500	0.412	13.34	43	
	B_1	0.166^{**}	0.055	3.01	269	0.167	0.055	3.03	268	
	γ_{01}					0.192	0.275	0.70	43	0.09
	$Var(r_{jk})$	4.707				4.716				
	$Var(u_{0k})$	0.007				0.006				

$^{***}p<.001$, $^{**}p < .01$, $^{*}p < .05$; B_1 = 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)□. $\delta = \gamma_{01}/\text{Sqrt}\,[Var(r_{jk}) + Var(u_{0k})]$



Note. Difference score = spring *PAS* score - fall *PAS* score; FS1 = FS_{verbal}; FS2 = $FS_{sorting}$

<u>Figure 1. Average Improvement in PAS Scores from Fall to Spring</u>