SREE 2011: Symposium submission

Contact (session chair) email
Kerry G. Hofer (kerry.g.hofer@vanderbilt.edu)
Peabody Research Institute, Vanderbilt University

Title The effects of pre-kindergarten and pre-kindergarten curricula on emergent math and literacy skills

First and second choice of conference section
Early Childhood Education
School and Classroom-Based Educational Practices

Symposium justification – 500 word limit
Pre-kindergarten is widely expected to boost the emergent math and literacy skills of economically disadvantaged children in ways that improve their ‘school readiness’ in the short term and their overall academic achievement in the long term. Whether such effects occur depends on the experiences children have in the pre-k classroom and those, in turn, depend largely on the classroom practices adopted by the teacher. One factor that can shape those experiences and practices is a defined curriculum that guides teachers’ selection of instructional activities and the relative emphasis given to each. There are many pre-k curricula in use but research to date has identified few that show significantly better outcomes than their alternatives.

This symposium will first present the results from studies of the effects of two distinctive curricula that extend the growing body of research on the effectiveness of pre-k curricula by investigating particularly innovative and diverse approaches to pre-k instruction. While most pre-k curricula focus on early reading and language skills, the first paper presents the outcomes for a well-developed math curriculum, Building Blocks, that is supported by equally well-developed professional development and coaching components. This curriculum showed promise in efficacy trials and has now been studied in a multi-site field experiment to assess its effectiveness at scale. The second paper is also about this curriculum but focuses on its effects on teachers’ classroom practice, particularly the relationship between fidelity of implementation and effects on math and language outcomes.

The third paper reports on the effects of a literacy-oriented curriculum, Opening the World of Learning (OWL), supplemented with a specialized program for the lowest performing students in the class, Enhanced Milieu Teaching (EMT). The experimental design was configured to allow estimation of the effects of OWL alone as well as those of the OWL-EMT combination. Fidelity of implementation and other classroom measures were also used to investigate moderators of the intervention effects.

The last paper in this symposium takes a different tack. It reports preliminary results from a randomized study of the effects of the Tennessee statewide Voluntary Pre-Kindergarten programs (TN-VPK). This study assesses whether a pre-k program implemented at scale with certified teachers, a well defined target population of economically disadvantaged children, and curricula drawn only from an approved list produces better learning outcomes than the available alternatives.

Collectively, these presentations will provide a picture of what some of the current generation of large-scale studies are finding about the effectiveness of pre-k and pre-k curricula.
Presentation #1

Title: An Examination of the Building Blocks Math Curriculum: Results of a Longitudinal Scale-Up Study

Authors:
Douglas H. Clements, PhD, University at Buffalo, The State University of New York (PRESENTER) (clements@buffalo.edu)
Julie Sarama, PhD, University at Buffalo, The State University of New York (jsarama@buffalo.edu)
Dale Farran, PhD, Vanderbilt University (dale.farran@vanderbilt.edu)
Mark Lipsey, PhD, Vanderbilt University (mark.lipsey@vanderbilt.edu)
Kerry G. Hofer, PhD, Vanderbilt University (kerry.g.hofer@vanderbilt.edu)
Carol Bilbrey, PhD, Vanderbilt University (carol.bilbrey@vanderbilt.edu)

Background / Context:
Studies show that the mathematics test-score gap is evident at every level of schooling and can be linked to students’ earlier performance. For example, a mathematics performance gap was found in children as young as three years of age (Case & Griffin, 1990; Jordan, Huttenlocher, & Levine, 1992). This gap has effects into kindergarten and 1st grade (Denton & West, 2002; Entwisle & Alexander, 1993; West, Denton, & Reaney, 2001) that continue into middle school and high school (Berkner & Chavez, 1997; Brasswell, Lutkus, Grigg, Santapau, Tay-Lim, & Johnson, 2001). However, the gap when children are young is relatively smaller than the gap at older grades (Bodovski & Farkas, 2007; Kilpatrick, Swafford, & Findell, 2001). Thus, children who come to school with less knowledge about mathematics are most at risk of falling behind in elementary mathematics, which affects their overall math achievement across the K-12 years. Addressing the mathematics performance gap early on, before children start school, has therefore become a priority for preschool programs serving children from low-income backgrounds (Clements, 2004).

Although the successes of some research-based educational practices have been documented, equally recognized is the “deep, systemic incapacity of U.S. schools, and the practitioners who work in them, to develop, incorporate, and extend new ideas about teaching and learning in anything but a small fraction of schools and classrooms” (see also Berends, Kirby, Naftel, & McKelvey, 2001; Cuban, 2001; Elmore, 1996, p. 1; Tyack & Tobin, 1992). There may be no more challenging educational and theoretical issue than scaling up educational programs across a large number of diverse populations and contexts in the early childhood system in the U.S., avoiding the dilution and pollution that usually plagues such efforts to achieve broad success. We created a research-based model to meet this challenge in the area of mathematics, with the intent that the model generalize to other subject matter areas and other age groups. The field needs transferable, practical examples of scale up (McDonald, Keesler, Kauffman, & Schneider, 2006); empirical evidence of the effectiveness of these examples; and focused research on critical variables—all leading to refined, generalizable theories and models of scale up. This paper examines the longitudinal effects of a randomized field trial involving the scale-up of a preschool mathematics curriculum.

Purpose / Objective / Research Question / Focus of Study:
The specific goal of our implementation of the TRIAD (Technology-enhanced, Research-based, Instruction, Assessment, and professional Development) model was to increase math achievement in young children, especially those at risk, by means of a high-quality field-centered implementation of the Building Blocks math curriculum, with all aspects of the curriculum—mathematical content, pedagogy, teacher’s guide, technology, and assessments—based on a common core of learning trajectories. The TRIAD intervention provides (a) these curriculum materials; (b) ongoing professional development, including scalable distance education, a web-based application with extensive support for teaching based on learning trajectories, and classroom-based coaching during the school year; and (c) supportive roles and materials for parents and administrators. In this project, we evaluated a large-scale implementation of the TRIAD intervention in distant geographical areas with diverse populations. The primary research question of interest is as follows:

Do children who are exposed to the Building Blocks mathematics curriculum in preschool perform better on measures of mathematics skills through the end of first grade than do children who were not exposed to that curriculum?

Setting:
This scale-up intervention took place in preschool classrooms in three urban school districts: the Buffalo Public School system in Buffalo, NY, the Boston Public School system in Boston, MA, and a combination of the Metropolitan Nashville Public School system and the Metropolitan Action Council Head Start system in Nashville, TN. A total of 62 sites (26 in Buffalo, 16 in Boston, and 20 in Nashville including 16 public schools and 4 Head Start centers) were randomly assigned to one of two conditions. This process yielded 103 classrooms that participated in the new math curriculum training and 60 classrooms that conducted business as usual. The original study sample included over 2,100 children who had at least partial pretest information collected on them, whether by direct assessment, teacher ratings, or observations.

Population / Participants / Subjects:
Participants in this study included primarily at-risk preschoolers between the ages of 4 and 6 from low-income households. The analysis sample, defined as those students who had at least partial pretest information, was comprised of 2076 students. The sample was roughly half male and half female, and predominately African American. Attrition rates were low throughout the study.

Intervention / Program / Practice:
Building Blocks, the I in the TRIAD acronym, was based on a comprehensive Curriculum Research Framework (Clements, 2007) and its efficacy validated by two Cluster Randomized Trial (CRT) evaluations, yielding effect sizes ranging from .5 to over 2 (Clements & Sarama, 2007, 2008). The Assessment component of TRIAD includes both formative assessment performed by the teachers training to use learning trajectories for this purpose, supplemented by the Building Blocks Software management system. TRIAD’s professional Development includes multiple forms of training (15 full days over two years, the first year a “gentle introduction” with no data collection) and support (coaching and mentoring). Each of these uses the software application, Building Blocks Learning Trajectories (BBLT), which presents and connects all components of the innovation. BBLT provides scalable access to the learning trajectories via descriptions, videos, and commentaries. The two main aspects of each
learning trajectory—the developmental progressions of children’s thinking and connected instruction—are linked to the other. Building Blocks is a supplemental mathematics curriculum designed to develop preschool children's early mathematical knowledge through various individual and small- and large-group activities. The curriculum embeds mathematical learning in children's daily activities, ranging from designated math activities to circle and story time, with the goal of helping children relate their informal math knowledge to more formal mathematical concepts. The Building Blocks curricular intervention in this scale-up study was implemented during the preschool year after teachers had a year of training and practice. Children from both treatment and control classrooms were followed through their first grade years.

Research Design:
This scale-up study was a cluster randomized field trial in which schools/centers were randomly assigned to experimental conditions.

Data Collection and Analysis:
Child outcomes were measured with the Research-based Early Mathematics Assessment (REMA), which uses an individual interview format, with explicit protocol, coding, and scoring procedures. It assesses children’s thinking and learning along research-based developmental progressions within areas of mathematics considered significant for preschoolers, as determined by a consensus of participants in a national conference on early childhood mathematics (Clements & Conference Working Group, 2004), rather than mirroring objectives or activities from any curriculum or state. Topics in number include verbal counting, number recognition and subitizing, object counting and counting strategies, number comparison and sequencing, number composition and decomposition, and adding and subtracting; geometry topics include shape identification, shape composition and decomposition, congruence, construction of shapes, and turns; and finally there are items on measurement and patterning. Content validity was assessed via expert panel review; concurrent validity was established with a .86 correlation with another instrument (Klein, Starkey, & Wakeley, 2000). For this study, Rasch scores for the total instrument were computed on correctness scores and logits transformed to T-scores (M = 50, SD = 10) for ease of interpretation. These T-scores were used for all statistical analyses. The REMA was given to children at the beginning and end of preschool, the end of kindergarten, and the end of first grade, as was the Renfrew Bus Story, a measure of children’s narrative recall skills (although only administered in the preschool year). In Nashville, children’s math and literacy achievement were also assessed using two math subtests and one literacy subtest from the Woodcock Johnson III. Each of these tests was given to children twice during the preschool year, once at the end of the kindergarten year, and again at the end of the first grade year. In order to examine the effectiveness of the curriculum in enhancing children’s mathematics skills, a series of linear mixed models were conducted. Independent models predicted children’s skills on each of the assessments at the end of prekindergarten, the end of kindergarten, and the end of first grade from curriculum condition, controlling for children's pretest skills and a host of demographic covariates. Children were nested in their prekindergarten classrooms, schools, and sites. Due to the nested nature of the design, the effective sample size for analyses is decreased by a factor related to the Intraclass Correlation Coefficient (ICC), or the degree to which classroom and school units are non-independent. Because of this reduction in analytical sample size, a p-value of .10 was held as the significance marker rather than the more conservative .05.
Findings / Results:

HLM analyses revealed that the two experimental groups differed significantly in math achievement in preschool. The Building Blocks group outperformed the control group on REMA measures with effect sizes ranging from .35 to .69. The Nashville site found effect sizes for Woodcock Johnson math subtests that ranged from .18 to .32. At no site were there significant effects on children’s letter/word identification. However, significant effects were found for expressive language in some sites but not others. While some significant curricular effects were found at the end of kindergarten, most differences between experimental groups had dissipated by the end of first grade.

Conclusions:

Our original project scaled up the implementation of a prekindergarten mathematics intervention that had been demonstrated in several randomized trials of increasing scope to increase foundational mathematics skills (Clements & Sarama, 2008). The key question for the scale-up project was whether the curriculum could be effective when provided under circumstances of routine practice on realistic scale—critical if it is to have any potential to help preschools across country improve math instruction. Many early childhood programs developed in universities or specialized research centers have proved initially effective but, when scaled up to be used by a “second-generation” (Farran, 1990, pg. 508), the effects have been diluted or proved non-existent [e.g., the Infant Health and Development Program (Brooks-Gunn, et al., 1994), Even Start (St. Pierre & Swartz, 1996), Head Start (U. S. GAO, 1997), and the Comprehensive Child Development Program (St. Pierre, Layzer, Goodson, & Bernstein, 1997)]. In contrast to that reported dilution, we found significant effects across outcomes in the children who participated in the original scale-up project, including different measures of mathematics achievement and one expressive language measure.

Research has suggested that early curricular effects may fade over time, resulting in very little, if any, discernable difference in elementary school between students who had been exposed to a given curriculum prior to formal schooling and students who were not exposed to such a program, as those without early curriculum exposure “catch up” to their peers (Barnett et al., 1995). In the Preschool Curriculum Evaluation Research (PCER) project, across all 14 curricula, kindergarten effects were nonexistent, prompting a decision not to collect any further longitudinal data. Similarly, with the scale-up project, we saw evidence of curricular effects across outcomes at the end of prekindergarten, but very few differences at the end of kindergarten, and virtually none at the end of first grade. Longitudinal research, including follow through interventions in these grades, is needed to determine if these early gains truly "fade," or if, as we posit, the problem is that primary grade curricula and teachers do not build upon them.
Appendix A. References


Presentation #2

Title: The Mechanisms behind the Results: Moderators of Building Blocks Curricular Effects

Author(s):
Dale Farran, PhD, Vanderbilt University (PRESENTER) (dale.farran@vanderbilt.edu)
Mark Lipsey, PhD, Vanderbilt University (mark.lipsey@vanderbilt.edu)
Douglas H. Clements, PhD, University at Buffalo, The State University of New York (clements@buffalo.edu)
Julie Sarama, PhD, University at Buffalo, The State University of New York (jsarama@buffalo.edu)
Kerry Hofer, PhD, Vanderbilt University (kerry.g.hofer@vanderbilt.edu)
Carol Bilbrey, PhD, Vanderbilt University (carol.bilbrey@vanderbilt.edu)
Elizabeth Vorhaus, Vanderbilt University (elizabeth.vorhause@vanderbilt.edu)

Background / Context:

In intervention research, it is critical to determine not just if an intervention is effective, but for whom it is effective and under what circumstances those effects occur. Moderators can be the key to answering those questions. A moderator is a variable that affects either the direction or the strength of the relationship between the predictor (curriculum condition, in this case) and the dependent variable (here, child outcomes) (Baron & Kenny, 1986). Identifying those variables that help specify the conditions under which interventions are most effective is central to social science research (Cohen et al. 2003). Moderators of curricular effects may be particularly important to scale-up studies. There may be no more challenging educational and theoretical issue than scaling up educational programs across a large number of diverse populations and contexts in the early childhood system in the U.S. The paucity of high-quality instruments, or the use of any measure of the fidelity of implementation, is one of the most important deficits in the field of scaling up educational innovations (Borman et al., 2003). This paper examines possible moderators, using fidelity as the primary variable of interest, of the effects of an early mathematics curriculum used in a scale-up study across three states.

Purpose / Objective / Research Question / Focus of Study:

The original project scaled up the implementation of a prekindergarten mathematics intervention that had been demonstrated in several randomized trials of increasing scope to increase foundational mathematics skills (Clements & Sarama, 2008). The key question for the scale-up project was whether the curriculum could be effective when provided under circumstances of routine practice on realistic scale—critical if it is to have any potential to help preschools across country improve math instruction. Many early childhood programs developed in universities or specialized research centers have proved initially effective but, when scaled up to be used by a “second-generation” (Farran, 1990, pg. 508), the effects have been diluted or proved non-existent [e.g., the Infant Health and Development Program (Brooks-Gunn, et al., 1994), Even Start (St. Pierre & Swartz, 1996), Head Start (U. S. GAO, 1997), and the Comprehensive Child Development Program (St. Pierre, Layzer, Goodson, & Bernstein, 1997)].
In contrast to that reported dilution, we found significant effects across outcomes in the children who participated in the original scale-up project. However, throughout the course of the study, we noticed that teachers implemented the curriculum to varying degrees, some displaying much greater fidelity of implementation than others. Additionally, we noticed that some children seemed to respond better to the intervention than their peers. This paper addresses those differences through a moderator approach, examining those variables like fidelity and child demographics that might explain some of the relationship between curriculum intervention and students’ academic outcomes.

Setting:

This scale-up intervention took place in preschool classrooms in three urban school districts: the Buffalo Public School system in Buffalo, NY, the Boston Public School system in Boston, MA, and a combination of the Metropolitan Nashville Public School system and the Metropolitan Action Council Head Start system in Nashville, TN. A total of 62 sites (26 in Buffalo, 16 in Boston, and 20 in Nashville including 16 public schools and 4 Head Start centers) were randomly assigned to one of two conditions. This process yielded 103 classrooms that participated in the new math curriculum training and 60 classrooms that conducted business as usual. The original study sample included over 2000 children who had at least partial pretest information collected on them, whether by direct assessment, teacher ratings, or observations.

Population / Participants / Subjects:

Participants in this study included primarily at-risk preschoolers between the ages of 4 and 6 from low-income households. The analysis sample, defined as those students who had at least partial pretest information, comprised of 2076 students. The sample was roughly half male and half female, and predominately African American. Attrition rates were low throughout the study.

Intervention / Program / Practice:

Building Blocks was based on a comprehensive Curriculum Research Framework (Clements, 2007) and its efficacy validated by two Cluster Randomized Trial (CRT) evaluations, yielding effect sizes ranging from .5 to over 2 (Clements & Sarama, 2007, 2008). TRIAD’s Professional Development includes multiple forms of training (15 full days over two years, the first year a “gentle introduction” with no data collection) and support (coaching and mentoring). Each of these uses the software application, Building Blocks Learning Trajectories (BBLT), which presents and connects all components of the innovation. BBLT provides scalable access to the learning trajectories via descriptions, videos, and commentaries. The two main aspects of each learning trajectory—the developmental progressions of children’s thinking and connected instruction—are linked to each other. Building Blocks is a supplemental mathematics curriculum designed to develop preschool children's early mathematical knowledge through various individual and small- and large-group activities. The curriculum embeds mathematical learning in children's daily activities, ranging from designated math activities to circle and story time, with the goal of helping children relate their informal math knowledge to more formal mathematical concepts.
The Building Blocks curricular intervention in this scale-up study was implemented during the preschool year after teachers had a year of training and practice. Children from both treatment and control classrooms were followed through their first grade years. Also during the full implementation preschool year, multiple classroom observations were conducted. Observations focused on a myriad of activities in the classroom, including implementation fidelity, teacher strategies, the general classroom environment, the specific math environment, and the behaviors of individual children. This is one of the few longitudinal studies to have detailed descriptions of classroom experiences in the prekindergarten year and to have descriptions that are directly related to the quantity and quality of mathematics instruction the children received.

Research Design:

This scale-up study was a randomized field trial in which schools/centers were randomly assigned to experimental conditions.

Data Collection and Analysis:

Child outcomes were measured with the Research-based Early Mathematics Assessment (REMA), which uses an individual interview format, with explicit protocol, coding, and scoring procedures. It assesses children’s thinking and learning along research-based developmental progressions within areas of mathematics considered significant for preschoolers, as determined by a consensus of participants in a national conference on early childhood mathematics (Clements & Conference Working Group, 2004), rather than mirroring objectives or activities from any curriculum or state. Topics in number include verbal counting, number recognition and subitizing, object counting and counting strategies, number comparison and sequencing, number composition and decomposition, and adding and subtracting; geometry topics include shape identification, shape composition and decomposition, congruence, construction of shapes, and turns; and finally there are items on measurement and patterning. Content validity was assessed via expert panel review; concurrent validity was established with a .86 correlation with another instrument (Klein, Starkey, & Wakeley, 2000). For this study, Rasch scores for the total instrument were computed on correctness scores and logits transformed to T-scores (M = 50, SD = 10) for ease of interpretation. These T-scores were used for all statistical analyses. The REMA was given to children at the beginning and end of preschool, the end of kindergarten, and the end of first grade. In Nashville, children’s math and literacy achievement were also assessed using two math subtests and one literacy subtest from the Woodcock Johnson III. Each of these tests was given to children twice during the preschool year, once at the end of the kindergarten year, and again at the end of the first grade year. Demographic information such as age, gender, ethnicity, IEP status, and ELL status were also collected.

Teacher implementation fidelity was measured with two instruments, one completed in both treatment and control classrooms, and the other completed only in treatment classrooms. The Classroom Observation of Early Mathematics-Environment and Teaching (COEMET) was used by observers in all participating classrooms. The COEMET is an instrument that measures the quality of the mathematics environment and activities, not connected to any curriculum. Thus, it allows for treatment-control group contrasts, no matter what the source of the enacted curriculum. Observers completed the COEMET during three separate observations across the full
implementation year for all teachers. The Near Fidelity measure was used by observers in the treatment classrooms only. This measure was focused on the general quality of the mathematics environment in the classroom, as well as detailed information concerning specific math activities. This instrument evaluated the degree to which teachers were teaching the specific intervention curriculum. This data from the COEMET and the Near Fidelity were collected from three four-hour observations in classrooms during the preschool implementation year: fall, winter, and spring.

To examine the possible existence of moderator variables that altered the effectiveness of the curriculum in enhancing children’s mathematics skills, a series of linear mixed models was conducted. Independent models predicted children’s skills on each of the assessments at the end of prekindergarten, the end of kindergarten, and the end of first grade from the main effect of curriculum condition, the main effect of each moderator, and the interaction of each moderator with curriculum condition. In analyses of the Near Fidelity measure, condition was not included. Children were nested in their prekindergarten classrooms, schools, and sites. Due to the nested nature of the design, the effective sample size for analyses is decreased by a factor related to the Intraclass Correlation Coefficient (ICC), or the degree to which classroom and school units are non-independent. Because of this reduction in analytical sample size, a p-value of .10 was held as the significance marker rather than the more conservative .05.

Findings / Results:

Teachers in the Building Blocks classrooms had higher COEMET scores across all time points than did control teachers. On a 5-point scale with 5.0 being the highest implementation score, treatment teachers scored an average of 3.84 across time while control teachers scored an average of 3.18. Regarding the number of math activities, teachers in Building Blocks classrooms were observed with an average of 2.4 structured, planned math activities and 9.3 less formal, miniature math activities per day, while teachers in control classrooms were observed with an average of 1.5 structured math activities and 5.7 miniature math activities per day. Additionally, the quality with which math activities were taught and managed was higher for treatment teachers than control teachers (3.44 and 3.09, respectively, on a 1-5 scale). Though there was a range of implementation observed in the treatment classrooms, teachers on average implemented over half of the curriculum’s suggestions in each instructional component (mean scores: General Curriculum: 12 out of 17; Hands-On Centers: 13 out of 22; Whole Group Activities: 26 out of 35; Small Group Activities: 54 out of 93; Computer Activities: 32 out of 49). While the differences in mathematics activities between treatment and control teachers were both significant and meaningful, the relatively low scores for the treatment teachers is evidence of the difficulty of changing the mathematics environment in early childhood classrooms.

Preliminary results regarding fidelity as a moderator of curriculum effects indicate that teachers with higher implementation scores had students who made greater gains on their mathematics measures (See Figures 1 and 2). Additional analyses are being conducted to examine the moderator effect of treatment-only fidelity and child demographics.

Conclusions:

Classrooms are complicated environments; educational interventions being scaled up rely on being able to change fundamental aspects of that environment, including teacher instructional
behaviors and child learning behaviors. Data from this study demonstrated teacher adoption of many aspects of the curriculum, and the degree of adoption was related modestly to the amount children gained in math knowledge over the preschool year. While teachers enacted more math instructional activities, changing the quality of their instruction was much more difficult. Moreover, the overall instructional classroom environment in treatment and control classrooms remained very similar, as did children’s learning behaviors. TRIAD is a multi-dimensional intervention that includes extensive professional development and in-classroom mentoring. The results from this scale up investigation demonstrate how difficult immediate and sustained change is to effect.
Appendix A. References


Appendix B. Tables and Figures

Figure 1. Mean Preschool NUMBER Student Gain by Curriculum/COEMET Grouping

Figure 2. Mean Preschool GEOMETRY Student Gain by Curriculum/COEMET Grouping
Presentation #3

Title: The Effects of Two Language-Focused Preschool Curricula on Children’s Achievement through First Grade

Authors: 
Ann Kaiser, PhD, Vanderbilt University (PRESENTER) 
David Dickinson, PhD, Vanderbilt University 
Megan Roberts, Vanderbilt University 
Catherine Darrow, PhD, University of North Carolina Chapel Hill 
Jill Freiberg, Vanderbilt University 
Kerry Hofer, PhD, Vanderbilt University

Background / Context:

Effective early language and literacy instruction to remediate language deficits and to prevent problems in learning to read is an important area for intervention research. Children with early language deficits who are growing up in poverty are dually at risk. Early deficits in language development predict both continued delays in language development and problems in acquisition of reading-related skills (Dickinson, McCabe, Anastasopoulos, Peisner-Feinberg & Poe, 2003; Snow, Burns & Griffen, 1998; Spira, Storch & Fischel, 2005; Storch & Whitehurst, 2002; Walker, Greenwood, Hart & Carta, 1994). Early emergent problem behavior, particularly in children with low language skills, also predicts difficulties in reading, academic performance and peer relationships (Hester & Kaiser, 1998). Without effective early intervention to teach language and emergent literacy skills, many of these children will require intensive, long-term special education. This paper reports findings from a large randomized field trial examining the effects of three variations of early literacy curricula implemented in Head Start classrooms on children with low language and matched children with average language skills. This project involved the comparison of three conditions: Opening the World of Learning (OWL), OWL curriculum combined with Enhanced Milieu Teaching (EMT) for low language children, and an enhanced version of Creative Curriculum (CC), the existing literacy program used by the Head Start center. Effects of the preschool curricula on children’s end of preschool, end of kindergarten, and end of first grade outcomes were examined.

Purpose / Objective / Research Question / Focus of Study:

This study was conducted as part a randomized field trial examining the effects of three variations of early literacy curricula implemented in Head Start classrooms located in an urban city in the southeast United States. This larger project involved the comparison of three conditions: Opening the World of Learning (OWL; Schickedanz & Dickinson, 2005), OWL curriculum combined with Enhanced Milieu Teaching (EMT; Kaiser, 1993) for low language children, and an enhanced version of Creative Curriculum (CC; Dodge, Colker, & Heoman, 2001), the existing literacy program which had been used by the Head Start centers for several years prior to the study. Head Start representatives requested that core elements of the existing curriculum be retained across all conditions; thus, the experimental conditions shared a common
classroom schedule, center choices, and certain thematic activities and materials. This paper focused on the following primary hypotheses:

- Children in the combined intervention (OWL+EMT) will show more growth in language than children in the language-rich curriculum (OWL) or the standard curriculum (CC).

- Children in the language-rich curriculum (OWL) will show more growth in language than children in the standard curriculum (CC).

- Children in the combined intervention (OWL+EMT) and language-rich curriculum (OWL) will show more growth in literacy skills than children in the standard curriculum (CC).

Setting:

This intervention took place in Head Start preschool classrooms in Birmingham, Alabama. Six clusters of Head Start centers that included 52 classrooms were randomly assigned to one of three conditions. This process yielded 19 classrooms that participated in the OWL+EMT curriculum, 17 classrooms that participated in the OWL curriculum, and 16 classrooms that conducted business as usual. The original study sample included over 450 children who had at least partial pretest information.

Population / Participants / Subjects:

After assignment of clusters to condition, 699 children preparing to enter a preschool classroom of three- and four-year-olds within those centers were screened for early expressive and auditory language skills using the Preschool Language Scale III (PLS-III; Zimmerman, Steiner & Pond, 1992). In order to be selected for screening, children had to be four years old by September of the upcoming school year. Based on the PLS-III total score, children were designated as low language (PLS score < 75; more than 1.5 standard deviations below the normative mean) or typical language (PLS score > 75). The Head Start agency assigned children to classrooms. Following classroom assignments, the research team selected four low language children and four typical language children based on their PLS-III total scores from each classroom to target for the project sample. Typical children were matched to low language children based on gender and age. In the OWL + EMT condition some children were moved among classrooms within centers so that four children with low language skills based on the PLS-III were included in each classroom. All children in every classroom, regardless of their PLS-III scores, received the treatment in the OWL and in the CC condition. In the OWL + EMT condition, all children received the OWL curriculum, but only children with low-language skills received the EMT component.

The analysis sample included 489 children for whom pretest information was available from direct assessments of their language and literacy skills or teacher ratings of social skills and behavior. The majority of this sample was African American (97.3%; 2.5% Caucasian and 0.2% Hispanic) and from primarily low-income households. Less than 5% of the sample had Individualized Education Plans at the start of the study. This sample consisted of 247 children (140 boys, 107 girls) who were low language (48.8%) and 242 children (124 boys, 118 girls)
who were typical language (51.2%). The average age of children at the PLS-III screening assessment was 52.9 months.

**Intervention / Program / Practice:**

**OWL:** All treatment programs in the study used *Opening the World of Learning (OWL)*, a curriculum for 4-year old preschool children, with adaptations suggested when 3-year olds used with. It addresses all curriculum areas, but has a particular focus on language and literacy skills. OWL is designed to systematically build language and associated conceptual knowledge by first introducing words when story books are read four times, then reinforcing key concepts and language in a brief group lesson, and then helping children to gain ownership of concepts and language through hands-on small groups followed in centers time when children explore concepts in child-initiated activities. The curriculum is sequenced to ensure that teachers increase the level of expectations and instruction during the year.

**OWL+EMT:** Four low language children in each OWL + EMT classroom received 60 individual 10-minute sessions of language instruction following the principles of the EMT model. Each teacher and assistant teacher was assigned two low language children in his/her classroom and worked with the same child throughout the implementation year. EMT sessions were conducted three times each week with individual children during centers time or nap time at tables away from other classroom activities. Teachers were trained to use EMT strategies while playing with children using thematic toys sets.

**CC:** The Birmingham Head Start program had used an enhanced version of Creative Curriculum for a number of years. For them CC is a framework that provides basic ways of interacting with children. To provide teachers more concrete guidance the program developed a set of approximately 10 themes that teachers used each year. As a result, the classroom environment and the ways of interacting with children were guided by CC, but the lesson plans were written based on the themes and theme activities provided by the agency.

The first year of the project was devoted to introducing the experimental curricula to the educational specialists and the teachers. In year two classroom supports continued, but an entirely new group of coaches worked with the teachers. It was during year two that all outcomes of interest were measured.

**Research Design:**

This study was a randomized field trial in which centers were randomly assigned to experimental conditions. Head Start clusters were randomly assigned to one of the three conditions. Four centers with 17 classrooms participated in the OWL curriculum alone; 4 centers with 19 classrooms participated in OWL + EMT for low language children; and 5 centers with 16 classrooms conducted business as usual and implemented the existing curriculum (CC).

**Data Collection and Analysis:**

Children were individually directly assessed with a number of early literacy and language measures at the beginning and end of preschool, the end of kindergarten, and the end of first grade. The following is a list of the direct assessments with a brief description of each measure:
• **Peabody Picture Vocabulary Test (PPVT):** The PPVT is a standard assessment of children’s receptive vocabulary skills and can be used with people ages 2 to 90+. During the assessment, a subject is read vocabulary words and asked to point to one of four pictures that the word represents. It was administered at the beginning of Pre-K, at the end of Pre-K, and at the end of Kindergarten.

• **Expressive Vocabulary Test (EVT):** The EVT is a standard assessment of children’s expressive vocabulary skills and can be used with people ages 2 to 90+. During the assessment, a subject is shown pictures and asked to verbally label the illustration with the correct vocabulary word. It was administered at the beginning of Pre-K, at the end of Pre-K, and at the end of Kindergarten.

• **Woodcock Johnson III Tests of Achievement:** The Woodcock Johnson is a standard assessment of a range of skills, designed to be used with people ages 2 to 90+. It was administered at the beginning of Pre-K, at the end of Pre-K, and at the end of Kindergarten.
  - **Letter-Word Identification:** This subtest measures literacy skills by requiring subjects to name letters and read words.
  - **Understanding Directions:** This subtest assesses the subject’s ability to follow oral directions.
  - **Sound Awareness:** This subtest measures phonological awareness through the four components of rhyming, deletion, substitutions, and reversals. However, only rhyming and deletion were administered in this study.

• **Preschool Language Scale (PLS-III):** This is a standardized assessment of language development that yields scores for expressive, receptive and total language for children ages 12-60 months. It was administered at screening and the end of Pre-K.

• **Language Samples:** These 30-minute interactions with a trained examiner were designed to provide a sample of children’s expressive language in a standardized context. Approximately 10 minutes was spent in three different contexts (play, narrative, and book). Language samples occurred at each assessment point. A standard set of linguistic measures is derived from the language sample (total words, number of different words, mean length of utterance, language complexity, sentence length, information scores, etc.)

• **Test of Language Development (TOLD):** This is a standardized assessment of oral language development that yields scores for semantics and grammar; listening, organizing, and speaking; and overall language ages 4-0 years to 8-11 years. It was administered at screening and the end of Kindergarten and First Grade.

Composite outcome measures were created from the individual instruments using conceptual groupings. Composite scores included Vocabulary (PPVT, EVT, Number of Different Words), Grammatical Sophistication (Sentence Length, Complexity, Information, Understanding Directions, MLU), Complex Language (PLS/TOLD), and Print Knowledge (Letter-Word Identification, Sound Awareness when applicable). In order to examine the effectiveness of the curricula in enhancing children’s language and literacy skills, a series of linear mixed models were conducted. Independent models predicted children’s skills on each of the composite measures at the end of preschool, the end of kindergarten, and the end of first grade from curriculum condition, controlling for children’s pretest scores and a host of demographic covariates. Children were nested in their preschool classrooms, centers, and clusters. Due to the nested nature of the design, the effective sample size for analyses is
decreased by a factor related to the Intraclass Correlation Coefficient (ICC), or the degree to which classroom and school units are non-independent. Hypotheses were tested within the language groupings (low and matched). Because of this reduction in analytical sample size, a p-value of .10 was held as the significance marker rather than the more conservative .05.

**Findings / Results:**

Children in all conditions gained significantly (~.5 SD), suggesting a beneficial effect of Head Start in general. Results from the multilevel analyses can be seen in Table 1. There were no significant overall curriculum differences on any of the composite measures at the end of preschool. At the end of kindergarten, there were significant curriculum effects for low-language children on Vocabulary (OWL only was significantly lower than the other two conditions) and Grammatical Sophistication (CC was significantly higher than the other two conditions), and a significant curriculum effect for matched-language children on Print Knowledge (OWL+EMT was significantly higher than the other two conditions). At the end of first grade, there was a significant curriculum effect for low-language children on Vocabulary (OWL only was significantly lower than the other two conditions). There were no significant curricular effects for matched-language children at the end of first grade.

**Conclusions:**

Overall, there were few systematic differences in language and literacy outcomes across time for low language and matched typical language children. Outcomes varied by child language status (low vs. matched language) and assessment time (end of preschool vs. end of K vs. end of 1st). The results were not those anticipated. Four potential explanations are being considered and examined in follow up analyses of predictors of child outcomes, teacher implementation fidelity, and “active ingredients” (specific teacher behaviors associated with positive language and literacy outcomes). We have not yet completed the full set of moderator analysis and we anticipate these analyses will aid in understanding the treatment outcomes. First, all classrooms across the three conditions were of relatively high quality as indicated by their scores on the direct observations of the classrooms. Second, the control and treatment classrooms shared a common base curriculum (Creative Curriculum) with the intervention curricula (OWL and OWL+EMT) overlaid on this base in the two treatment conditions. This arrangement (which was deemed the only acceptable arrangement by the Head Start administration) may have served to minimize differences in the treatment and control classrooms. Because teachers had many years experience with Creative Curriculum and the OWL and OWL+EMT interventions were newly trained and required significant changes in practice, they may have maintained a common high quality implementation of Creative Curriculum and that drove the similarity of findings across conditions. Fidelity coding (in progress) suggests that the implementation of the Creative Curriculum features was similar across the three treatment conditions. Third, the active ingredients in the intervention (teacher behavior related to language and literacy) did vary by teacher, but may not have varied sufficiently by condition. We have preliminary analyses to suggest that at least one key feature, teacher use of specific vocabulary, was related to child outcomes, varied across teachers, but not across conditions (Dickinson et al, 2009; IES poster).
Appendix A. References


Appendix B. Tables and Figures

Table 1. Curricular Effects on Child Outcomes over Time

<table>
<thead>
<tr>
<th>Source</th>
<th>N</th>
<th>F</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low-Language Sample</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>End of Pre-Kindergarten Outcomes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vocabulary Composite</td>
<td>194</td>
<td>1.796</td>
<td>.327</td>
</tr>
<tr>
<td>Grammatical Sophistication Composite</td>
<td>191</td>
<td>0.232</td>
<td>.807</td>
</tr>
<tr>
<td>Complex Language</td>
<td>195</td>
<td>1.430</td>
<td>.370</td>
</tr>
<tr>
<td>Print Knowledge</td>
<td>197</td>
<td>3.165</td>
<td>.119</td>
</tr>
<tr>
<td>End of Kindergarten Outcomes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vocabulary Composite</td>
<td>193</td>
<td>6.990</td>
<td>.001</td>
</tr>
<tr>
<td>Grammatical Sophistication Composite</td>
<td>190</td>
<td>3.818</td>
<td>.024</td>
</tr>
<tr>
<td>Complex Language</td>
<td>196</td>
<td>0.156</td>
<td>.865</td>
</tr>
<tr>
<td>Print Knowledge</td>
<td>196</td>
<td>0.380</td>
<td>.684</td>
</tr>
<tr>
<td>End of First Grade Outcomes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vocabulary Composite</td>
<td>191</td>
<td>4.474</td>
<td>.013</td>
</tr>
<tr>
<td>Grammatical Sophistication Composite</td>
<td>188</td>
<td>0.457</td>
<td>.666</td>
</tr>
<tr>
<td>Complex Language</td>
<td>193</td>
<td>1.563</td>
<td>.350</td>
</tr>
<tr>
<td>Print Knowledge</td>
<td>181</td>
<td>1.015</td>
<td>.475</td>
</tr>
<tr>
<td>End of Pre-Kindergarten Outcomes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vocabulary Composite</td>
<td>207</td>
<td>2.315</td>
<td>.112</td>
</tr>
<tr>
<td>Grammatical Sophistication Composite</td>
<td>200</td>
<td>0.610</td>
<td>.544</td>
</tr>
<tr>
<td>Complex Language</td>
<td>213</td>
<td>0.440</td>
<td>.660</td>
</tr>
<tr>
<td>Print Knowledge</td>
<td>211</td>
<td>0.613</td>
<td>.582</td>
</tr>
<tr>
<td>End of Kindergarten Outcomes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vocabulary Composite</td>
<td>203</td>
<td>0.465</td>
<td>.629</td>
</tr>
<tr>
<td>Grammatical Sophistication Composite</td>
<td>197</td>
<td>0.469</td>
<td>.629</td>
</tr>
<tr>
<td>Complex Language</td>
<td>208</td>
<td>1.356</td>
<td>.267</td>
</tr>
<tr>
<td>Print Knowledge</td>
<td>207</td>
<td>4.289</td>
<td>.015</td>
</tr>
<tr>
<td>End of First Grade Outcomes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vocabulary Composite</td>
<td>198</td>
<td>0.564</td>
<td>.572</td>
</tr>
<tr>
<td>Grammatical Sophistication Composite</td>
<td>192</td>
<td>0.006</td>
<td>.994</td>
</tr>
<tr>
<td>Complex Language</td>
<td>202</td>
<td>0.123</td>
<td>.884</td>
</tr>
<tr>
<td>Print Knowledge</td>
<td>187</td>
<td>1.589</td>
<td>.329</td>
</tr>
</tbody>
</table>
Presentation #4

Title: The Effects of the Tennessee Voluntary Pre-Kindergarten Program: A Randomized Field Experiment

Authors:
Kerry Hofer, PhD, Vanderbilt University (PRESENTER) (kerry.g.hofer@vanderbilt.edu)
Mark Lipsey, PhD, Vanderbilt University (mark.lipsey@vanderbilt.edu)
Dale Farran, PhD, Vanderbilt University (dale.farran@vanderbilt.edu)
Carol Bilbrey, PhD, Vanderbilt University (carol.bilbrey@vanderbilt.edu)

Background / Context:
Few randomized studies of the effectiveness of contemporary public pre-kindergarten programs have been conducted, and virtually none have followed students past the end of the pre-k year to investigate longer term effects. Despite widespread support and advocacy for the expansion of public pre-k, therefore, evidence of its benefits is sparse. The study reported here was launched in partnership with the Tennessee State Department of Education’s Office of Early Learning to provide a rigorous assessment of the effects of the statewide Tennessee Voluntary Pre-Kindergarten (TN-VPK) program on the readiness for kindergarten of the economically disadvantaged population it serves.

Purpose / Objective / Research Question / Focus of Study:
This study was designed to assess the effects of participation in the TN-VPK program, relative to the options parents would otherwise exercise (e.g., Head Start, community daycare, no preschool), on kindergarten readiness and later academic achievement.

Setting:
A range of urban and rural Tennessee School Districts distributed across the state.

Population / Participants / Subjects:
To be eligible for TN-VPK, children must be age four on or before September 30 of the respective school year. By statutory requirement, the program gives top priority to children who qualify for the Free or Reduced Price Lunch Program, and 86% of the children enrolled statewide meet that criterion. The sample participating in the study was restricted to children who met that criterion. Statewide, 28% of the children enrolled in TN-VPK are African-American with another 9% other minorities, 4% are English language learners, and 10% have special education designations.

Intervention / Program / Practice:
TN-VPK began as a pilot program in 1998 and expanded rapidly after 2005 with legislation that provided funding from state lottery monies. The program operates through competitive grants to local school systems who apply for approval and funding. Those grants support only a portion of the necessary classroom funding, the balance must come from other sources. This arrangement permits and encourages collaboration between school systems and other organizations. In this “collaboration model,” school districts may operate their pre-k programs through collaborative agreements with local non profit and for-profit child care
providers and Head Start programs so long as those agencies have attained the highest rating from the licensing system administered by the Tennessee Department of Human Services and meet the State pre-k standards. The TN-VPK program in each participating school district must meet standards set by the State Board of Education that, among other things, require the following:

- A state licensed teacher with an early childhood education endorsement in each classroom;
- A teacher assistant who holds or is actively working toward at least a CDA or associate degree in early childhood;
- Professional development support for teachers;
- An adult-student ratio no smaller than 1:10;
- A small class size (maximum of 20);
- An approved age-appropriate curriculum aligned with the Tennessee Early Childhood Education Developmental Standards;
- A family engagement component and a pre-k to kindergarten transition plan for each child;
- Vision, hearing, and health screening and referral services;
- A minimum of 5.5 hours per day, exclusive of nap time, for a minimum of 180 days per year within a calendar that includes 200 working days of 7.5 hours for teaching staff.

Research Design:
School districts with history or current circumstances that made it likely that they would receive more applications for TN-VPK than could be accommodated were recruited to participate in the study during the first year. 54 schools submitted lists of applicants and agreed to admit students in the order in which they were listed after they were randomized by the research team. Not all of the students on the respective list were admitted for 33 of the schools, leaving a control group. These 33 schools were thus blocks in a randomized block design with treatment (TN-VPK) and control (practice as usual) children in each block. A second year implementation of this same design is currently underway with a larger number of schools participating.

Data Collection and Analysis:
All of the approximately 1000 children in the schools that participated successfully in the randomization are being tracked through the Tennessee State Department of Education information system. Data is being extracted from this system on their attendance, retention in grade, and special education status. When they reach third grade, their scores on the statewide achievement tests (TCAP) will be available in reading/language arts, mathematics, science, and social studies. In addition, a subset of approximately 400 of these students has parental consent for individual assessments. These children are being assessed annually with a battery of Woodcock Johnson III scales and teacher ratings of academic performance and problem behaviors.

Findings / Results:
Data collection from the first wave of individual pre and post outcome measures has recently been completed and a second wave is underway. These have not been completely analyzed, but preliminary results will be available to report at the SREE meeting by March.