# TITLE: Boosting School Readiness: Should Preschools Target Skills or the Whole Child?

Greg Duncan, Jade Marcus Jenkins
*University of California, Irvine*

Anamarie Auger
*RAND Corporation*

Marianne Bitler
*University of California, Davis and NBER*

Thurston Domina, Margaret Burchinal
*University of North Carolina at Chapel Hill*

**ABSTRACT**

Both federal and state governments regulate early childhood education programs in hopes of promoting the school readiness of disadvantaged children. We draw on data from the experimental Preschool Curriculum Evaluation Research Initiative Study to provide an aggregated look at the impacts of four types of published preschool curricula, which are packaged teaching materials aimed at boosting children's cognitive and noncognitive outcomes. The math curriculum included in the study increased both classroom math activities and children's math achievement relative to the two alternative curricula (HighScope and Creative Curriculum) found in most Head Start and pre-K classrooms. Also relative to HighScope and Creative Curriculum, the literacy curriculum increased early literacy achievement despite producing no statistically significant differences in classroom activities or teacher-child interactions. Although Creative Curriculum produced much more positive classroom processes than locally developed curricula, it failed to improve either the academic achievement or behavior of preschool children relative to children exposed to the local curricula. Implications for Head Start and pre-K curricula choice and the utility of widely used classroom rating scales are discussed.

**INTRODUCTION**

The math and literacy skills of low-income children are a full year behind those of high-income children at the time of kindergarten entry, and these gaps do not diminish by the time the children reach eighth grade (Reardon, 2011). Early childhood education programs attempt to remedy these gaps by helping children build foundational academic and noncognitive skills prior to the start of formal schooling.

Experimental and quasi-experimental research indicates that exposure to high quality early childhood education can have long-term positive impacts on earnings and health, with the most encouraging evidence coming from early childhood education programs that operated decades ago—Abecedarian and Perry Preschool (Anderson, 2008; Belfield, Nores, Barnett, & Schweinhart, 2006; Campbell et al., 2014; Campbell et al., 2008; Heckman, Moon, Pinto, Savelyev, & Yavitz, 2010). Few of today's public preschool programs generate shorter-run impacts that can compare in magnitude to those of Abecedarian and Perry (Yoshikawa et al., 2013). Furthermore, recent evaluations of the federal Head Start program and state-funded public pre-K programs suggest that the effects of early childhood education vary considerably between sites (Bloom & Weiland, 2015; Walters, 2014; Wong, Cook, Barnett, & Jung, 2008).

In this paper, we focus on one potentially important source of variation in the impacts of early childhood programs on students' academic skill development: The content and style of instruction (known in schools and the education literature as the curriculum). Most preschool classrooms in the United States use an instructional approach that is typically described as "whole-child" or "global," an approach that federally-funded Head Start programs are mandated to use. Rather than having teachers provide explicitly academic instruction, this model seeks to promote learning by encouraging children to engage independently in a classroom stocked with toys and materials designed to promote child development. This approach is grounded in a rich body of research on child development (DeVries & Kohlberg, 1987; Piaget, 1976; Weikart & Schweinhart, 1987). Compellingly, it is also the approach that was utilized in the highly-effective Perry Preschools of the 1970s (Belfield et al., 2006; Schweinhart, 2005). Federal law requires Head Start programs to purchase and utilize instructional materials that adopt the whole-child approach, and many state-funded pre-K programs use similar instructional materials. However, despite this mandate no convincing causal evidence supports the efficacy of this whole child approach to early education relative to other curricula that provide more explicitly targeted instruction in specific topics, such as literacy and numeracy (U.S. Department of Education, 2013).

Using data from the Preschool Curriculum Evaluation Research Initiative Study (PCER, 2008), a multi-site experimental study in which early childhood education centers in 9 states were randomly assigned to receive various instructional materials, we investigate the relative effectiveness of whole-child and skills-based instruction. Using test scores as our measures, we find that children learn more in early childhood programs that provide explicit academic instruction in mathematics and literacy for a small portion of the day, compared with programs that take an exclusive whole-child approach.

**BACKGROUND**

Over the past 40 years, evidence of the long-term individual and societal benefits of early childhood programs has shifted U.S. public opinion and policy toward investments in public preschool programs (William S. Barnett, 1995; Warner, 2007). Federal spending on Head Start and the Child Care Development Fund, the federal government's two largest child development programs, totaled $12.8 billion in 2014 (Isaacs, Edelstein, Hahn, Steele, & Steuerle, 2015), with states spending an additional $5.5 billion on programs like universal pre-K (W. S. Barnett, Carolan, Squires, Brown, & Horowitz, 2015). Research has shown highly variable impacts for these programs, with Head Start appearing to produce both short and long-run gains in sibling-based studies (Deming, 2009) but small overall and quickly disappearing impacts in the National Head Start Impact Study (Puma et al., 2013).[1] In the case of pre-K studies, modified regression discontinuity studies of pre-K programs in five states show a diverse set of impacts (Wong et al., 2008). A recent study of Boston's pre-K finds uniformly large impacts on academic skills and noteworthy impacts on an important set of other key life skills – executive function (Weiland & Yoshikawa, 2013).[2]

Perhaps the most intuitive strategy for boosting the consistency and effectiveness of early education programs is improving the curricula they use to organize instruction. Curricula provide teachers with teaching materials to enable them to cultivate their students' academic and non-cognitive skills. Curricula set goals for the knowledge and skills that children should acquire in an educational setting, and support educators' plans for providing the day-to-day learning experiences to cultivate those skills with items such as such as daily lesson plans, materials, and other pedagogical tools (Gormley, 2007; Ritchie & Willer, 2008). Most preschool curricula are created by educational researchers and practitioners and then sold to preschool programs by publishers, yet there exists little or no evidence about which curricula are best for whom.

[1] Bitler, Hoynes, and Domina (2015) find that these small average effects after the first year of the experiment mask large impacts at the bottom of the child skill distribution. Interestingly, Kline and Walters (Forthcoming), find that

[2] It is worth noting that state pre-K programs vary considerably across states with respect to their availability, quality, and program standards. Most of the evaluations of pre-K programs have focused on the states with the highest quality programs. Cascio and Schanzenbach (2013) examine the effects of two universal public pre-K programs in Oklahoma and Georgia and find they lead to larger enrollments among low-income children but crowd out some private use by higher income children, and that they lead to higher test scores for low-income children.

Published curricula and teaching materials differ across a number of dimensions; including but not limited to philosophies, materials, the role of the teacher, small or large group settings, classroom design, and the need for child assessment. Preschool programs typically choose their own curriculum, but their curricular choices are often constrained by a pre-approved list developed by state agencies and accrediting bodies (Clifford & Crawford, 2009). Implementing a curriculum can be challenging, and programs often must train and mentor teachers to implement the chosen curriculum faithfully. Nonetheless, curricular guidance and restrictions may be an important and relatively efficient policy lever through which states can influence the quality and effectiveness of their preschool programs. We propose to compare the effectiveness of "whole-child" curricula and more targeted, skill-specific curricula in altering children's reading and mathematics test scores.

*Whole-Child Curricula*

Whole-child (sometimes termed "global") curricula emphasize "child-centered active learning," cultivated by strategically arranging the classroom environment (DeVries & Kohlberg, 1987; Piaget, 1976). Rather than explicitly targeting specific academic skills (e.g., math, reading), they seek to promote learning by encouraging children to interact independently with the equipment, materials, and other children in the classroom environment. The most famous example of a program based on a whole-child curriculum is the Perry Preschool study, which used a version of the HighScope curriculum that was very similar to the one evaluated here (Belfield et al., 2006; Schweinhart, 2005). Whole-child curricula dominate preschool programs, in part because Head Start program standards require centers to adopt them (Advisory Committee on Head Start Research and Evaluation, 2012). In addition, whole-child curricula reflect the standards for early childhood education put forth by the National Association for the

Education of Young Children—the leading professional and accrediting organization for early educators (Copple & Bredekamp, 2009).

We focus our empirical work on the two most common whole-child curricula used by Head Start grantees and other preschool programs, Creative Curriculum and HighScope (Clifford et al., 2005). Some 46% of the teachers responding to the national Head Start Family and Child Experiences Survey utilized Creative Curriculum; 19% utilized HighScope (Hulsey et al., 2011). Since these two whole-child curricula have a similar focus and approach, in this paper we consider them to be functionally equivalent.[3]

Despite the widespread adoption of these whole-child curricula in preschools, virtually no rigorous evaluation studies have estimated the impacts of whole-child curricula on children's school readiness. In fact, the Institute for Education Science's (IES) What Works Clearinghouse concludes that only one of the evaluation studies of Creative Curriculum meets minimal standards of empirical rigor (U.S. Department of Education, 2013). That study is based on the same PCER data we use in the current study, and its site-specific results reveal no statistically or substantively significant differences across a host of test scores—including children's oral language, print knowledge, phonological processing, or math skills—between Creative Curriculum and the control group's curricula that were developed by individual preschool programs ("locally developed curricula"). Evidence supports the effectiveness of the earliest version of the HighScope curriculum from the 1960s Perry Preschool studies on children's

[3] The Department of Education's IES What Works Clearinghouse (WWC) describes Creative Curriculum as "designed to foster development of the whole child through teacher-led, small and large group activities centered around 11 interest areas (blocks, dramatic play, toys and games, art, library, discovery, sand and water, music and movement, cooking, computers, and outdoors). The curriculum provides teachers with details on child development, classroom organization, teaching strategies, and engaging families in the learning process" (U.S. Department of Education, 2013, p. 1). Creative Curriculum also allows children a large proportion of free-choice time (Fuligni, Howes, Huang, Hong, & Lara-Cinisomo, 2012). HighScope is similar and emphasizes, "active participatory learning," where students have direct, hands-on experiences and the teacher's role is to expand children's thinking through scaffolding (Schweinhart & Weikart, 1981).

cognitive scores and outcomes like crime. However, there are three reasons to wonder about the applicability of these findings to the current setting. First, there exist no methodologically strong evaluations of recent versions of the curriculum. Second, the modern-day counterfactual conditions for many children are quite different than they were in the 1960s, with a large share of 3 and 4 year olds attending some out of home care today. Thirdly, the children in the Perry study were extremely disadvantaged, with IQ tests 1-2 standard deviations below the population mean (Schweinhart & Weikart, 1981).

*Content-Specific Curricula*

Supporters of curricula that target specific academic or behavioral skills argue that preschool children benefit most from sequenced, explicit instruction focused on those skills but that this should be implemented through "free play" and exploration activities that are age-appropriate (Wasik & Hindman, 2011). These curricula often supplement a classroom's regular curriculum (e.g., Creative Curriculum or a teacher or locally developed curriculum). Some random-assignment evaluations of content-specific curricula focusing on language, mathematics, and socio-emotional skills find positive impacts on their targeted sets of skills (Bierman, Domitrovich, et al., 2008; Bierman, Nix, Greenberg, Blair, & Domitrovich, 2008; Clements & Sarama, 2008; Diamond, Barnett, Thomas, & Munro, 2007; Fantuzzo, Gadsden, & McDermott, 2011; Klein, Starkey, Clements, Sarama, & Iyer, 2008; Morris et al., 2014). For example, children who received a literacy-targeted curriculum showed improvements in their literacy and language skills (Justice et al., 2010; Lonigan, Farver, Phillips, & Clancy-Menchetti, 2011). Corresponding gains are also observed in the case of a preschool mathematics curriculum, with children in classrooms implementing this curriculum showing larger gains in their math skills compared with children who received business as usual only (i.e., the regular curriculum) (Clements & Sarama, 2007, 2008).

*Locally Developed Curricula*

Many states allow early childhood education providers not otherwise subjected to curriculum requirements to develop their own lesson plans or curricula rather than purchasing a published curriculum. These are designed by local districts or teachers themselves, but may incorporate components of various commercial curricula. While not as common as whole-child curricula in public preschool programs, locally or teacher-developed preschool curricula constitute the business as usual conditions in some of the control classrooms in the data we analyze (see below).

Given the large negative gaps in achievement and behavior between low- and higher-income children at school entry, coupled with the widespread adoption of global curricula such as Creative Curriculum and HighScope, it is crucial for policy to evaluate whether achievement-focused or locally developed curricula systematically outperform the most commonly used preschool curricula – Creative Curriculum and HighScope – across various domains of school readiness including both cognitive and social-emotional test scores. Our article undertakes such a comparison.

**DATA**

We draw on data from the Preschool Curriculum Evaluation Research (PCER) Initiative Study (2008). The PCER study, funded by the Institute of Education Sciences, began in 2003 and provided evaluations of 14 early childhood education curricula. A total of 12 grantees were selected to independently evaluate one or more curricula; all used common measures of child outcomes, classroom processes, and implementation quality. The 14 curricula were evaluated at 18 different research locations, and 2,911 children were included in the evaluation. Each of the grantees independently selected their early childhood education centers, conducted random assignment, and managed their own evaluation with assistance from Mathematica and RTI. The level of random assignment differed across grantees, with the majority of grantees randomly assigning whole preschools to the treatment or business as usual comparison conditions and the rest randomly assigning classrooms within preschool centers to treatment or business as usual curricula. The centers included in the PCER study were public preschools, Head Start programs, and private child care; all primarily served children from low-income families.

The analyses in the PCER final report (2008) provide only grantee-specific estimates of the standardized outcome differences between the treatment curricula and the counterfactual control "business as usual" curricula. Our study pools data from all grantees that implemented: i) a math or literacy curriculum where the comparison control condition was Creative Curriculum or HighScope; ii) a literacy curriculum where the comparison control condition was a locally developed curriculum (not enough math sites included a locally developed comparison); or iii) the Creative Curriculum where the comparison control condition was a locally developed curriculum.[4]

Our inclusion criteria (described below) led us to drop four grantees and a total of 1,070 children from the study: three grantees were omitted because they evaluated a whole-child curriculum other than Creative Curriculum or HighScope (the Wisconsin, Missouri and three Success For All sites), and the other (New Hampshire) evaluated a literacy enhanced version of

[4] Note that while for the first 2 comparisons, Creative Curriculum is among the business as usual control group curricula, for two of the PCER grantees, the Creative Curriculum was the assigned treatment curriculum, with locally-developed curricula as the control. This third comparison provides us the experimental estimate of the impacts of the Creative Curriculum relative to the locally-developed ones.

Creative Curriculum with Creative Curriculum as the comparison condition. Thus, the whole-child curricula included in our study represent the curricula as they are typically implemented in large-scale preschool programs.

**Curricula Categories: Literacy, Mathematics, Whole Child, and Locally-Developed**

We coded each of the treatment curricula in the PCER study into one of four mutually exclusive categories: literacy, mathematics, whole-child, and locally developed. All literacy curricula focused on a so-called literacy domain, which could have included phonological skills (e.g., sounds that letters make), prewriting skills, or any other early literacy skill, and which differed widely. By contrast, the PCER study included only one math-focused preschool curriculum.

Each of the included PCER curricula and its designated category are described in Table 1. Eight curricula that targeted language/literacy were included that varied in terms of content and focus. [5] Despite these differences and in hopes of attaining some degree of generalizability, we included all of these in our "literacy" group. The one math curriculum combined Pre-K Mathematics with software from the DLM Early Childhood Express Math to focus on sequenced instruction in numeracy and geometry.

Our "whole-child" category included HighScope and Creative Curriculum, both of which share a broad focus on developing children's social and academic skills and promoting health and nutrition. Our final category, "locally-developed curricula," included curricula that were

[5] One curriculum focused solely on language – the Language-Focused Curriculum, and sought to improve language skills through enhancing the language stimulation techniques used in the classroom. The other seven focused primarily on literacy instruction, but varied in terms of structure and sequence. The least structured literacy curriculum appeared to be Bright Beginnings, which focused on child-centered curriculum units. In the middle are Ladders to Literacy and Doors to Discovery, which provided skill-building activities designed to improve language and basic literacy skills. The remaining four curricula were the most structured; explicitly focusing on sequenced instruction in oral language, phonological and phonemic awareness and letter knowledge.

developed either by teachers in the classrooms or by the local school district, or were a combination of several of these types of curricula. We lack information on the general content of the locally-developed curricula used in some of the PCER study control classrooms and suspect they likely vary widely. Nonetheless, they characterize the kinds of settings experienced by a substantial share of preschoolers and serve as a useful counterfactual in some of our comparisons. Our data also provide some measures detailing classroom processes associated with each curriculum with the classroom outcome models presented in the next section. (Classroom processes are teacher-student interactions, overall instructional quality, and the total number of academic activities.) Table 1 and Figure 1 summarize all of the experimental contrasts and sites included in our study, along with other study information.

**Fidelity of Implementation**

The results of most program evaluations depend on the fidelity of program implementation, which, in our case, means the fidelity with which the treatment and "business as usual" control curricula were implemented. Classroom ratings of fidelity of implementation were reported in the PCER (2008) report and are reproduced in Table 1. Table 1 shows that fidelity was typically medium (2 or medium on a 1 (not at all) to 3 (high) scale). Importantly there were only small differences in average fidelity across the treatment and control groups, ranging from 0.15 for literacy vs. whole child (on a mean of 2-2.5) to 0.5 for math vs. whole child (on the same mean).[6] Treatment sites also received additional training and professional support to implement the curricula, whereas control conditions implemented the curricula as usual. But this training and support failed to generate substantial differences in fidelity.

**Classroom Process Measures**

[6] Some sites had a pilot year and we test for whether this affected outcomes, finding no significant differences.

One drawback to using cognitive test scores to assess the quality of instruction is that this approach provides no information about what aspect of teaching is leading to improvements in child outcomes. By contrast, the goal of classroom observations is to assess what teachers do and how they interact with their students to unpack this black box. In the teacher effectiveness/value added literature, researchers have incorporated classroom observations to assess the processes and learning activities occurring in classrooms (Kane, Taylor, Tyler, & Wooten, 2011). Our data also provide a look into the black box of classroom processes that may underlie preschool effectiveness. We use several classroom-level observational measures assessing the quality of the preschool classrooms that were included in the PCER study, which enables us to assess whether the type of approach used by the teacher impacts the nature of classroom activities and the warmth of teacher-child interactions. We convert each measure to standard deviation units so the estimates can be interpreted as effect sizes. Reliability, citations, and additional information for each of the process quality measures we use are available in Appendix Table 1.

The most widely known is the *Early Childhood Environment Rating Scale – Revised* (ECERS-R)(Harms, Clifford, & Cryer, 1998). The ECERS–R is an observational tool used by trained observers who conduct interviews with the staff at the center and observe the classroom during a recommended time period of three hours. Classrooms are observed for safety features, teacher-child interactions, and classroom materials, and program staff are interviewed to assess teacher qualifications, ratio of children to adults, and program characteristics, spread across 7 subscales. Each item is rated on a scale of 1 – 7, with 1 indicating inadequate quality and 7 excellent quality. Previous analyses show that two key factors come out of these items – an Interactions scale, which focuses on teacher-child interactions, and a Provisions scale, which contains items related to classroom materials and the safety features of the setting (Pianta et al., 2005). ECERS-R observations were conducted in the fall and spring of the 2003-04 preschool year; the spring measure serves as one of our classroom quality outcomes; the fall score is used as one of the control variables in our impact regressions.

The *Teacher Behavior Rating Scale* (TBRS)(Landry, Crawford, Gunnewig, & Swank, 2002) includes four scales that capture the quantity and quality of math and literacy activities conducted in the classroom. Classrooms were observed and assessed by trained observers on the number of math (5 items) and literacy activities present in the classroom (25 items; 4 categories – book reading, print and letter knowledge, oral language use, and written expression). We combined the quality and quantity scales for literacy to form a literacy activity composite, and combined the math quality and quantity scales to form a math composite, which became our primary outcome measures. (We also control for TBRS observation time to account for variation in time spent observing each classroom.) The TBRS was administered only in the spring of 2004.

The *Arnett Caregiver Interaction Scale* (Arnett, 1989) was designed to measure the caregivers' positive interactions, warmth, sensitivity, and punishment style. It is also used in some state quality ratings. Observers rate interactions between the caregivers and the children on 30 items using a four-point scale. Our analyses use the total score, which is the average of the 30 items, with the negative items reversed. A higher score indicates a more supportive, positive classroom environment. As with the ECERS-R, Arnett observations were conducted in the fall and spring of the 2003-04 preschool year; the spring measure serves as one of our classroom outcomes, and the fall score is used as a control.

Because the time between the fall and spring assessments varies across classrooms and sites, we control for elapsed time between fall and spring assessments to ensure that these differences do not confound the length of the curricular implementation period with classroom quality assessments. [7]

**Children's Outcomes**

Children's academic achievement and social-emotional skills were assessed using nationally normed measures that are developmentally appropriate for preschool children and used frequently in developmental research. Children were assessed or rated on each of the academic and socioemotional outcomes in the fall and spring of the 2003-04 preschool year. We focus on aggregated measures of math, literacy, and socioemotional skills. Appendix Tables 2 and 3 present the means, standard deviations, and observation counts for all outcomes and covariates by treatment status for all four curricula comparison groups in Table 1. Observation counts are rounded to the nearest ten in accordance with NCES data policies.

*Literacy Outcomes*

We draw upon three commonly utilized literacy outcomes. The *Peabody Picture Vocabulary Test* (PPVT)(Dunn & Dunn, 1997) assesses children's receptive vocabulary. It takes approximately 5-10 minutes to complete, is administered by a trained researcher, and requires the child to point to the picture that represents the word spoken to them by the researcher. Words increase in difficulty and scores are standardized for the age of the child. The measure is nationally normed, with a mean of 100 and a standard deviation of 15.

The second and third literacy measures – Letter Word and Spelling – come from the *Woodcock-Johnson III (WJ-III) Tests of Achievement* (Woodcock, McGrew, & Mather, 2001). The *Letter Word* subtest is similar to the PPVT in that it asks children to identify the letter or

[7] In the fall, the classroom quality assessments were conducted between 2 and 8 weeks after the start of the preschool year, and in the spring 2-15 weeks before the end of the preschool year.

word spoken to them, and the test gradually increases in difficulty to require the child to read words out of context. The *Spelling* subtest requires children to write and spell words presented to them. Both of these assessments from the WJ-III were administered by trained researchers and each took approximately 10 minutes to administer. As with the PPVT, scores are standardized by the age of the child and nationally normed to have a mean of 100 and a standard deviation of 15. The assessments were standardized for the sample to have a mean of 0 and a standard deviation of 1, and averaged together. We then restandardized the composite to have a mean of zero and a standard deviation of 1.[8]

*Math Outcomes*

To measure student mathematics skills, we combine data from two measures into a summary composite. The *Applied Problems* subtest comes from the WJ-III and requires children to solve increasingly difficult math problems. This instrument also assesses basic skills such as number recognition. Like the literacy measures from the WJ-III, the Applied Problems subtest is standardized for a child's age, and nationally normed to have a mean of 100 and a standard deviation of 15. The assessment takes approximately 10 minutes to administer. The second math assessment, the Child Math Assessment-Abbreviated (CMAA)(Klein & Starkey, 2002) is less well known, and was designed specifically for the PCER study (Preschool Curriculum Evaluation Research Consortium, 2008). It assesses young children's math ability in the domains of numbers, operations, geometry, patterns, and nonstandard measurement. Our analyses use the composite score from the CMAA. To create an overall math outcome composite, both math measures were standardized for the sample to have a mean of zero and a

[8] We also consider both these tests and the math tests discussed below taken one at a time and these results are presented in Appendix Table 5. The advantage of combining them as we do here is that it addresses concerns about multiple testing implicit with using more than 1 measure, and additionally might capture an overall significant effect where the individual measures do not.

standard deviation of one. The measures were then averaged together and restandardized (mean 0, SD 1). We also constructed an academic composite score that combined the math and literacy composites and then restandarized the sum.

*Social-Emotional Outcomes*

Teachers rated children's social skills and behavior problems using the *Social Skills Rating System* (SSRS)(Gresham & Elliott, 1990). The SSRS preschool edition contains 30 items related to social skills and 10 items related to problem behaviors. Each item is rated on a three-point scale, ranging from never to very often). The social skills and problem behaviors scales are nationally normed to have a mean of 100 and a standard deviation of 15. To form a social-skills composite score, we standardized (within the sample) both scales to have a mean of zero and a standard deviation of one, reverse coded the problem behaviors scale, averaged the two scores together and restandardized.

**Baseline Controls**

To increase the precision of our experimental impact estimates, we include a host of baseline covariates in all analyses. At baseline the primary caregiver reported on child, personal, and family demographics and background characteristics. Child-level characteristics included gender, race (white as the omitted category, dummies for black, Asian, Hispanic, and other), and age in months. Maternal/Primary caregiver and family characteristics included education level in years, a dummy variable for working or not, age in years, annual household income in thousands of dollars, and a dummy for receiving welfare support. We also control for children's fall preschool academic and social skills composites, along with classroom measures as appropriate.

**ANALYTIC APPROACH**

We conducted two sets of analyses; the first focusing on classroom process outcomes and the second on child achievement and noncognitive outcomes. Both are based on the following regression model:

(1) $O_{icj} = \alpha + \beta_1 T_{cj} + \beta_2 Cov_{icj} + \mu_j + e_{icj}$,

where $O_{icj}$ is the classroom or child outcomes observed for child *i* in classroom *c* in research site *j;* $T_{cj}$ is a dichotomous indicator of assignment to the treatment or control curriculum (this varies by classroom or site); $Cov_{icj}$ are classroom, child, and family covariates for child *i*, $\mu_j$ are research site fixed effects; and $e_{icj}$ is an error term. For each classroom[9] or child outcome, we estimate four versions of equation (1), one for each of the four treatment/control comparisons shown in Figure 1. The results illustrated in Figure 2 show the magnitude and significance of $\beta_1$ for our four primary outcomes (ECERS-R, literacy skills, math skills, and social skills).

All analyses use Ordinary Least Squares with standard errors clustered at the classroom level (c) and including fixed effects for the unit within which random assignment is made (school or research site, denoted by "j" in equation (1)).[10] Including the research site fixed effects produces random-assignment variation in our treatment/control contrasts. Thus, the parameter $\beta_1$ in equation (1) provides an average effect size based on all of the treatment/control differences across all of the research sites evaluating a given curriculum contrast. We handled missing data in independent variables using dummy variables.

**Samples**

[9] Note that the classroom observations do not vary across children within classroom. However, we run these regressions at the child level in part because we are also controlling for individual covariates.
[10] Results are robust to alternative approaches to conducting inference, including other clustering schemes and various bootstrapping approaches; these results are discussed below.

Our sample for the classroom process analyses included children in classrooms in one of the curricula comparison sites listed in Table 1 for whom at least one of the classroom observational composite measures (ECERS-R, TBRS Math, TBRS Literacy, Arnett) and one of the academic outcome composite measures at the end of preschool were available**.** The sample for our child outcomes analyses consisted of children who had at least one school readiness outcome at the end of preschool and were enrolled in one of the curricula comparison sites listed in Table 1.

**RESULTS**

Appendix Tables 2 and 3 present descriptive statistics for the four curriculum comparison samples outlined in Table 1 separately for children in the treatment and control groups. We compared balance in the covariates at baseline between each treatment and control group using a clustered t-test (accounting for nonindependence within experimental site) to assess whether the randomization was successful. P-values from t-tests show that child and family characteristics, including children's baseline school readiness scores, were statistically indistinguishable across literacy vs. whole child or math vs. whole child comparisons. There were also no differences in the classroom observational measures for these comparisons.

Baseline differences emerged in the classroom observational measures in the locally developed versus Creative Curriculum experimental comparisons, and the literacy versus locally-developed comparison, however. This difference was also noted in PCER by study investigators and may reflect the fact that classroom processes in the Creative Curriculum treatment schools may have changed prior to the time that the baseline measurements were conducted (2008). The PCER report also noted that at the Vanderbilt site (Creative Curriculum compared with locally-developed curricula) there was a possible early treatment effect on an ECERS-R scale and in the

Texas site (literacy compared with locally-developed curricula) the investigators note baseline on an Arnett subscale. Of somewhat less concern, some of the baseline Xs were also significantly different individually in comparison III (gender (5% level); parent's education and household income (at the 10%% level), but the joint test of significance across baseline measures was insignificant and baseline cognitive tests were not significantly different. We address these issues by controlling for classroom assessment scores at baseline and for child and family covariates. It is possible that baseline controls and controls for covariates may not completely restore equivalence. We view the more troubling comparison as that between Creative and the locally-developed curricula and we regard this comparison as less rigorously causal than the others and place less weight on this in our conclusions and discussion. Still, even though the joint test of baseline controls for the Math versus whole child curricula is not statistically significant, one might worry about the fact that two SES measures look marginally different.

**Classroom Outcomes**

Table 2 shows impact estimates for the classroom outcomes, which are also displayed in Figures 2-4 for Contrasts I, III, and IV. All dependent variables were converted into standard deviation units (variables were standardized within the sample), with a mean of zero and standard deviation of one so that the coefficients can be interpreted as effect sizes. Our main results used the four composite classroom measures as the dependent variables. We show the same models using the composite components as dependent variables in Appendix Table 4.

*I. Literacy Curricula Compared with Creative/HighScope*

The ECERS score was 0.25 standard deviations (sd) higher in classrooms with the Literacy Curricula, but the difference was only significant at the 10% level. There were no other statistically significant differences at the end of the preschool year between classrooms using a literacy curriculum and classrooms using the Creative/HighScope curriculum on the 3 remaining classroom observational measures.

*II. Literacy Curricula Compared with Locally-Developed Curricula*

Classrooms using a literacy curriculum scored one-half of a sd higher on the ECERS-R total score (significant at the 5% level), and 0.83 sd higher on the TBRS Literacy activities composite (significant at the 1% level) at the end of the preschool year than classrooms using a locally-developed curriculum.

*III. Math Curricula Compared with Creative/HighScope*

Classrooms using the math curriculum scored more than one standard deviation higher on the TBRS Math activities scale (significant at the 5% level) than control classrooms using Creative/HighScope at the end of the preschool year. There were no other significant differences between Math treatment and Creative/HighScope classrooms.

*IV. Creative Curriculum Compared with Locally-Developed Curricula*

Classrooms using Creative Curriculum had consistently higher ECERS-R, TBRS Math, TBRS Literacy, and Arnett scores (effect sizes = 0.61 sd, 0.51 sd, 0.71 sd, 0.99 sd, respectively, and all significant at the 5% level) at the end of the preschool year than classrooms using a locally-developed curriculum.

In sum, conventional measures of classroom instruction and teacher-child interactions were uniformly better with the whole-child Creative Curriculum than with the assortment of locally-developed curricula comprising the control condition. If better processes in the classroom translate into better test scores and behavior, then we would expect positive effects on child outcomes for Creative Curriculum vs. business as usual curricula. Further classroom improvements from supplementing or replacing whole-child curricula with skill-focused curricula were more selective.

**Child Cognitive and Social-Emotional Outcomes**

Table 3 show impacts of the various curricula contrasts on children's school readiness outcomes; results for the literacy, math, and social skills composites are also illustrated in Figures 2-4 for Contrasts I, III, and IV. As with the process measures, outcomes were standardized within the sample so that coefficients can be interpreted as effect sizes. Our main models used the four composite child outcome measures as the dependent variables. We show the same models using the composite components as dependent variables in Appendix Table 5.

*I. Literacy Curricula Compared with Creative/HighScope: Literacy Curricula Raise Composite Literacy Scores*

Children in classrooms randomly assigned to a Literacy curriculum had modestly but significantly higher literacy composite scores (0.15 sd) at the end of preschool than did classrooms using Creative/HighScope. Appendix Table 5 shows that this marginally significant difference in literacy scores is driven in part by a increase in the WJ Spelling test of 0.18 standard deviations (SE of 0.07, significant at the 5% level), and that the point estimates for the WJ Letter Word are also positive but insignificant. There were no other statistically significant differences between children exposed to literacy curricula and Creative/HighScope, although Appendix Table 6 shows significant detrimental impacts of the literacy curricula on one of the two components of the social skills composite.

*II. Literacy Curricula Compared with Locally-Developed Curricula: Literacy Curricula Lead to Higher Math and Composite Scores*

Children in classrooms randomly assigned to a literacy curriculum had marginally significantly ($p<.10$) higher math (0.14 sd) and academic composite scores (0.15 sd) at the end of preschool than children who received a locally-developed curriculum. These stem from an in increase of 0.18 sd in the CMAA math component (significant at the 1% level) and an increase in the WJ spelling literacy component of 0.16 sd (significant at the 10% level). The effect size for the literacy composite was similar (0.15 sd), but not statistically significant at conventional levels.

*III. Math Curricula Compared with Creative/HighScope: Math Curricula Raises Math and Academic Composite Scores*

Children in classrooms randomly assigned to the Math curriculum had substantially higher math (0.35 sd) and academic composite scores (0.25 sd) at the end of preschool compared with children who received Creative/HighScope. The WJ Applied Problems and CMAA math scores are also both significantly higher for children who were in classrooms with the Math Curriculum. Children did not have significantly different literacy or social skills composite scores.

*IV. Creative Curriculum Compared with Locally-Developed Curricula: No Effects on School Readiness*

Despite the consistently positive impacts of the Creative Curriculum on all composite measures of classroom process, there were substantively small and statistically insignificant differences between the school readiness skills of children exposed to Creative Curriculum and locally-developed curricula. When looking at the components, some of the coefficients are negative but insignificant (WJ Letter Word, WJ Spelling, CMAA), while others are positive but insignificant (WJ Applied Problems) and only one is even marginally significant (PPVT).

In sum, despite the uniformly better process measures for Creative compared with the locally-developed curricula, there were no significant differences in school readiness (and the differences there were small in magnitude). This is true despite the large and significant differences in the classroom process measures described above. By contrast, despite mixed differences across the whole child and targeted math and literacy curricula in the process outcomes, both the literacy and math scores outperformed the alternatives in the skill they were targeted at, with the math vs. Creative differences being quite large.

**Robustness Checks**

*Classroom Outcomes*

One might be concerned that PCER provided too few classrooms to generate unbiased cluster-adjusted standard errors and that clustering would instead lead to over-rejection (e.g., Bertrand, Duflo, & Mullainathan, 2004). Another concern is the assumption that classroom clustering treats classrooms within site as independent. To address these concerns, we repeated the above analyses using the wild bootstrap based on random assignment site for the literacy versus whole child comparison where there are 5 sites (Cameron, Gelbach, & Miller, 2008).[11] Wild bootstrap inference leads to conclusions about significance that are very similar to those presented above. We also ran classroom outcome models that omitted the Fall 2003 baseline scores because some study sites participated in the pilot year (see Table 1), and therefore the Fall 2003 classroom process measure were not a true baseline score. The coefficients were generally similar, and for several comparisons, larger than those presented in Table 2.

*Pilot Year Interactions*

[11] This also adjusts for the fact that there may be more than one classroom within specific random assignment sites.

We wanted to test for differences in effects between sites that participated in a pilot implementation year and those that did not. As indicated in Table 1, all sites in comparisons II, III, and IV were pilot sites, so we were only able to test for differences between pilot and non-pilot sites for comparison I (Literacy vs. HighScope and Creative Curriculum). We found no significant differences in the effects of literacy curricula on the classroom or child outcomes by pilot site status.

*Pooling HighScope and Creative Curriculum*

In the Literacy vs. Creative Curriculum/HighScope comparison, four sites used HighScope and one site used Creative Curriculum. We tested whether removing the Creative Curriculum site from this analysis would alter the results. The coefficients from these analyses were very similar to those presented in Table 3, with the exception of the ECERS-R scores, which increased from 0.25 sd to 0.34 sd and reached statistical significance.

*New York Control Group*

The Math curriculum was randomly assigned to classrooms at two sites: New York and California. The original PCER study control group for New York consisted of state prekindergarten (pre-K) classrooms using a locally-developed curriculum (excluded from above analyses) and Head Start classrooms using Creative Curriculum/HighScope (included). Because our analyses effectively split the New York control group by both curricula and program type, we tested whether different constructions of the Math curriculum control group would affect our results. Appendix Table 6 shows results from the model presented in our main results, a model that excludes all of the New York control group children, and one that excludes the New York Math site entirely. The magnitude and significance of the Math curriculum effect on the math composite is robust to different constructions of the control group, but the statistical significance of the effect on the academic composite is sensitive to changing the control group, most likely because of the small sample size.

*Creative Curriculum and HighScope Comparison with the Head Start Impact Study*

One concern with the Creative Curriculum/HighScope comparison groups is that these specific sites may not be representative of the way other programs use these curricula. To address this concern, we compared the ECERS-R and Arnett scores from the Head Start classrooms that used Creative Curriculum or HighScope in the Head Start Impact Study (HSIS) with those of classrooms in the PCER study using these curricula (pooled across all research sites). The overall average ECERS-R scores in the PCER and HSIS samples were 4.21 and 5.22, respectively. On the Arnett the respective PCER and HSIS sample averages were 3.12 and 2.55, respectively. These differences suggest some limitations on external validity; PCER sites using whole child curricula that chose to participate were ones where their overall quality was subpar. An additional critique that might be raised is that our study does not properly address whether the fully and properly implemented whole child curricula do as well as do the experimental targeted curricula. We respond to this argument by noting that this is at least a test of one feasible policy alternative, replacing the current set of business as usual curricula (improperly implemented) with fully implemented targeted approaches.

We also compared baseline academic scores for children in the 4-year old cohort in the HSIS with children in the PCER study who received the Creative Curriculum or HighScope curriculum. Scores were compared at the beginning of the preschool year to determine if the samples were similar. Children in the HSIS who received one of the comparison curricula had very similar scores to those of children in the PCER study, with no significant differences across the two groups.[12]

*Child Outcomes*

As might be expected, children in our PCER-based analysis sample were not representative of the national distribution of children for which the nationally normed outcome measures (PPVT, Woodcock-Johnson Letter-word, Spelling, and Applied Problems) are calibrated. Thus, the effect sizes here may not capture the effect size in the national population if these comparisons were examined at-scale. We used the same comparisons and specifications presented to estimate treatment effects on raw outcome scores and calculated effect sizes by dividing by the standard deviation for the population (15). These coefficients and effect sizes are presented in Appendix Table 7, and are virtually identical to those presented in Table 3.

We also estimated the same specifications as our main analyses but excluding the set of child and family control variables, some of which may be endogenous. The results were very similar to those presented above.

*Child Outcomes at Kindergarten*

The PCER study included a follow-up data collection of children's outcomes at the end of their kindergarten year, one year after the outcomes we report in Figure 2. Using the same comparisons and specifications presented above, we tested whether curricular effects were sustained until the spring of kindergarten. For composite outcomes, none of the statistically significant content-focused curricular effects shown in Table 3 remained statistically significant

[12] The PPVT scores averaged 92.18 in the HSIS and 86.68 in the PCER; WJ Applied Problems means were HSIS: 90.28, PCER: 92.80; WJ Letter Word means were HSIS: 95.12, PCER: 99.82; and WJ Spelling means were HSIS: 92.74, PCER: 94.27).

at the end of kindergarten. Fadeout is all too common in early childhood program evaluations and perhaps points to the need to coordinate curricula and instruction between preschool and early elementary school grade so that preschool intervention gains might be sustained (e.g., Clements, Sarama, Wolfe, & Spitler, 2013)

*Training and Professional Development for Curricular Implementation*

An alternative explanation for the observed effects is that the professional development and training provided to treatment classrooms are driving our results, and not the curricula per se. The argument here is that treatment classrooms my have obtained much more intensive implementation than "business as usual" curricula users. But if the training associated with these programs alone accounted for the differences, we should have seen significant differences in child outcomes in the Creative Curriculum treatment condition compared with the teacher developed control (comparison IV). Training and professional development are important components of any preschool program, but they do not explain the pattern of results we see here.

## DISCUSSION

Given the large, persistent, and consequential gaps in literacy and numeracy skills between high- and low-income children when they enter kindergarten, the most important policy goal of publicly supported early childhood education programs should be to boost early achievement skills and promote the socioemotional behaviors that support these skills. Federal, state, and local policy can influence the effectiveness of preschool programs by prescribing curricula, as well as by regulating and monitoring early care settings. Our evidence speaks most directly to curriculum policies. Considering that curricula cost between $1100-$4100 per classroom, with 50,000 classrooms in the Head Start program alone, the costs of such policies are nontrivial (Office of Head Start, 2010).

We find that curricular supplements focused on academic skills are indeed more successful at boosting literacy and math skills than are widely used whole-child curricula. What about the whole-child curricula themselves, which programs like Head Start require their classrooms to use? Our data showed no advantages for Creative Curriculum compared with locally-developed curricula in improving academic skills, nor in promoting positive behavior. Here it is important to bear in mind that none of the curricula were implemented with high fidelity under the developer's recommended conditions. On the other hand, the classrooms in the PCER study are likely to reflect a degree of implementation found in many actual classrooms.

Our results, coupled with the absence of other high-quality evaluation evidence demonstrating the effectiveness of the Creative Curriculum, HighScope or any other whole-child curricula lead us to question the policy wisdom of prioritizing either "whole-child" curricula as a whole, or Creative Curriculum and HighScope in particular. While it is conceivable that some kind of "whole-child" curriculum may ultimately be found to be particularly effective at promoting a valued conception of school readiness, there is currently no evidence to support that conclusion. In the absence of such evidence, we conclude that policy efforts should focus more attention on assessing and implementing proven academically focused curricula and move away from the comparatively ineffective whole child approach. While curricula developers may protest that this study is not a valid test of how the curricula would perform if implemented perfectly as designed, it is a test of the de facto experience of many low-income children in preschool programs. Just as some clinical trials lead to larger differences between new drugs and the previous standard treatments than is found when the new drugs are widely adopted, so might it be for the ideal implementation of curricula versus what is happening on the ground.

Our findings further suggest that some commonly used child care quality instruments (i.e., classroom observations) may be too global to provide useful measurement of children's experiences in those settings that drive the acquisition of academic and social skills (Burchinal et al., 2015). State and federal policies have focused on measures of classroom quality, with the assumption that higher classroom quality will lead to larger gains in academic and social skills among young children. As with prior studies, our study finds no consistency between curricular impacts on classroom quality and impacts on children's school readiness. The most striking example is the contrast between classrooms adopting Creative Curriculum and classrooms with an assortment of locally-developed curricula. Almost all of our measures of the quality and quantity of academic content, the sensitivity of teacher-pupil interactions, and the global rating scale of classroom quality (the ECERS-R) currently used by most states were significantly more favorable in classrooms that had implemented Creative Curriculum than in classrooms using locally-developed curricula. And yet these classroom process advantages failed to translate into better academic or socioemotional outcomes for children. Nevertheless, these findings provide further evidence that evaluations may need to include assessments of child outcomes as well as classroom quality if the goal of the program is impact children's school readiness skills. Another possibility is to reserve the highest tier of ratings for programs that successfully implement proven academically focused, content-specific curricula. However, our study suggests a more direct approach: encouraging or mandating the use of academically focused curricula to enhance the school readiness of low-income children. [13]

A number of considerations suggest caution in drawing strong policy conclusions from our analysis. First, the results are specific to the skill-focused curricula included in the PCER study. In the case of math, only one curriculum was tested, and it is one of the few preschool math curricula to have proved its effectiveness in other random-assignment evaluation studies (Clements & Sarama, 2011). Eight different literacy curricula were tested in the PCER study, and, although effects are imprecisely estimated, the PCER evaluation showed that the impacts of those curricula on literacy achievement were quite heterogeneous. Our analyses, which combine these heterogeneous programs into a single category thus provide an estimate of the average effects of these eight literature curricula. Our estimates would likely be larger had we limited the sample to literacy curricula with strong evidence of effectiveness. While the collection of skill-focused curricula used in our analyses outperformed the widely used global curricula in boosting academic skills, future research should focus on specific curricula to aid policy choices in this area. It is also important to note that curricula targeting children's socioemotional skills or executive functioning (e.g., the REDI program or Tools of the Mind) were not included in the PCER study; these should be compared in future research.

A second and enduring feature of most evaluation studies is that their comparisons involve real-world classrooms in which curricula implementation may fall short of what curricula designers judge to be adequate. Implementation assessment scores in the PCER were

[13] It is possible that state quality rating systems for early education programs are not targeting the elements of program quality that matter the most for boosting children's school readiness skills. Some evidence suggests that newer classroom quality measures that capture the nature of teacher-child relationships or quality of domain-specific instruction may provide stronger linkages between classroom process and children's outcomes (Burchinal, Vandergrift, Pianta, & Mashburn, 2010).

fairly high, but in many cases, teachers received less training prior to implementing curricula than designers recommend. Teachers in the control conditions did not receive any additional training on their curricula, representing de facto real-world curricular implementation in scaled-up public preschool programs. In the case of HighScope, for example, recommended training lasts four weeks, which was considerably longer than the training times in the PCER study. HighScope also recommends a curriculum implementation protocol that was more sophisticated than the PCER protocol. Of course, there may have been similar problems in the implementation of the academic and even locally-developed curricula. The policy infrastructure surrounding curricular requirements would therefore also need to involve on-site assistance and/or extensive training opportunities for child care providers if proven curricula are to be effective at scale.

**Integrated Curricular Approaches: Boston's Pre-K Program**

Looking beyond individual curricula and quality rating systems, a third policy approach to promoting school readiness is to develop a completely integrated academic and behavioral curriculum and then focus on ensuring that it is implemented in classrooms as faithfully as possible. Classroom "quality" in this case amounts to the fidelity of the implementation of the curriculum. This has been the approach taken over the past decade by Boston Public Schools (Duncan & Murnane, 2014, Chapter 5; Weiland & Yoshikawa, 2013).

System leaders developed a highly-scripted play-based curriculum by combining proven literacy, math and social skills interventions. The academic components focused on concept development, the use of multiple methods and materials to promote children's learning, and a variety of activities to encourage analysis, reasoning, and problem-solving (Weiland and Yoshikawa, forthcoming). Pre-K classrooms were embedded in existing public schools and taught by credentialed teachers who received extensive professional development training and on-going coaching to ensure that they understood the curriculum and were able to implement it effectively in their classrooms.

A regression-discontinuity evaluation of the Boston pre-K system showed much larger impacts on vocabulary, math, and reading (effect sizes ranging from .45 to .62 standard deviations) than the PCER curricula were able to generate, although some of these differences might be attributed to the differences between the regression discontinuity and the RCT evaluation designs and the external validity of the estimates they generate (regression discontinuity vs. RCT; Weiland & Yoshikawa, 2013). Interestingly, the evaluation also found smaller, but still noteworthy impacts on working memory and inhibitory control (effect sizes ranging from .21 to .28 standard deviations). Given its all-or-nothing nature, the evaluation could not identify which subset of the many program components were the "active ingredients" leading to the school readiness impacts. It is obviously premature to view integrated curricular approaches to preschool quality as the gold standard approach. The Boston model needs to be evaluated using a strong design that can track impacts on child outcomes during and beyond elementary school. And it needs to be shown to be replicable at scale in other school systems serving predominantly low-income children.

Stepping back, our results from the PCER preschool experiments provide a number of reasons to question the wisdom of current school readiness policies. Our study highlights the importance of curricula as a policy lever to influence the school readiness skills of low-income children, based on good, experimentally-based evidence. We find no such support for policies targeting preschool process quality alone. The entire policy debate would benefit from a stronger culture of telling program evaluations.

**Acknowledgments**

We are grateful to the Institute of Education Sciences (IES) for supporting this work through grant R305B120013 awarded to Principal Investigator Greg Duncan and Co-Principal Investigators Farkas, Vandell, Bitler, and Carpenter, and to the Eunice Kennedy Shriver National Institute of Child Health & Human Development of the National Institutes of Health under award number P01-HD065704. The content is solely the responsibility of the authors and does not necessarily represent the official views of IES, the U.S. Department of Education, or the National Institutes of Health. We would also like to thank Douglas Clements, Dale Farran, Rachel Gordon, Susanna Loeb, and Aaron Sojourner for helpful comments on prior drafts.

Data used in this article are restricted-use because they contain individually identifiable information that are confidential and protected by law. IES requires a license for their use, which assures that the users of the data abide by the set of laws that provide for the security and privacy of personal data maintained by the Federal Government to protect the confidentiality of respondents' individually identifiable information. The IES releases restricted-use data only to qualified organizations (universities, research organizations and other government agencies) in the United States. Ph.D. level researchers can apply as a Principal Project Officer through their organization to access these data and replicate our work. The IES received IRB approval through the respective states of each participant in the study.

While we are unable to share the data with other researchers because of our agreement with the IES, we can offer guidance (including sharing of programming code if necessary) to researchers who do receive the data from IES and wish to replicate and/or extend our results.

The authors have no conflicts of interest to report.

**References**

Advisory Committee on Head Start Research and Evaluation. (2012). *Final Report*. Retrieved from Washington, DC:

Anderson, M. L. (2008). Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects. *Journal of the American statistical Association, 103*(484), 1481-1495. doi:10.1198/016214508000000841

Arnett, J. (1989). *Caregiver Interaction Scale*. Princeton, NJ: Educational Testing Service.

Barnett, W. S. (1995). Long-term effects of early childhood programs on cognitive and school outcomes. *Future of Children, 5*, 25-50.

Barnett, W. S., Carolan, M. E., Squires, J. H., Brown, K. C., & Horowitz, M. (2015). *The State of Preschool 2014*. New Brunswick, NJ.

Belfield, C. R., Nores, M., Barnett, W. S., & Schweinhart, L. J. (2006). The High/Scope Perry Preschool Program. *Journal of Human Resources, XLI*(1), 162-190. doi:10.3368/jhr.XLI.1.162

Bertrand, M., Duflo, E., & Mullainathan, S. (2004). How Much Should We Trust Differences-In-Differences Estimates? *The Quarterly Journal of Economics, 119*(1), 249-275. doi:10.1162/003355304772839588

Bierman, K. L., Domitrovich, C. E., Nix, R. L., Gest, S. D., Welsh, J. A., Greenberg, M. T., . . . Gill, S. (2008). Promoting Academic and Social-Emotional School Readiness: The Head Start REDI Program. *Child Development, 79*(6), 1802-1817. doi:10.1111/j.1467-8624.2008.01227.x

Bierman, K. L., Nix, R. L., Greenberg, M. T., Blair, C., & Domitrovich, C. E. (2008). Executive functions and school readiness intervention: Impact, moderation, and mediation in the Head Start REDI program. *Development and psychopathology, 20*(03), 821-843. doi:doi:10.1017/S0954579408000394

Bitler, M. P., Hoynes, H. W., & Domina, T. (2014). Experimental Evidence on Distributional Effects of Head Start. *National Bureau of Economic Research Working Paper Series, No. 20434*. doi:10.3386/w20434

Bloom, H. S., & Weiland, C. (2015, Nov 7). *Quantifying Variation in Head Start Effects on Young Children's Cognitive and Socio-emotional Skills using data from the National Head Start Impact Study*. Working Paper. New York, NY.

Burchinal, M. R., Vandergrift, N., Pianta, R., & Mashburn, A. J. (2010). Threshold analysis of association between child care quality and child outcomes for low-income children in pre-kindergarten programs. *Early Childhood Research Quarterly, 25*(2), 166-176. doi:http://dx.doi.org/10.1016/j.ecresq.2009.10.004

Cameron, C., Gelbach, J. B., & Miller, D. L. (2008). Bootstrap-based improvements for inference with clustered errors. *Review of Economics and Statistics, 90*(3), 414-427.

Campbell, F. A., Conti, G., Heckman, J. J., Moon, S. H., Pinto, R., Pungello, E., & Pan, Y. (2014). Early Childhood Investments Substantially Boost Adult Health. *Science, 343*(6178), 1478-1485. doi:10.1126/science.1248429

Campbell, F. A., Wasik, B. H., Pungello, E., Burchinal, M. R., Barbarin, O., Kainz, K., . . . Ramey, C. T. (2008). Young adult outcomes of the Abecedarian and CARE early childhood educational interventions. *Early Childhood Research Quarterly, 23*(4), 452-466. doi:10.1016/j.ecresq.2008.03.003

Cascio, E. U., & Schanzenbach, D. W. (2013). The Impacts of Expanding Access to High-Quality Preschool Education. *National Bureau of Economic Research Working Paper Series, No. 19735*.

Clements, D. H., & Sarama, J. (2007). Effects of a preschool mathematics curriculum: Summative research on the Building Blocks project. *Journal for Research in Mathematics Education*, 136-163.

Clements, D. H., & Sarama, J. (2008). Experimental Evaluation of the Effects of a Research-Based Preschool Mathematics Curriculum. *American Educational Research Journal, 45*(2), 443-494. doi:10.3102/0002831207312908

Clements, D. H., & Sarama, J. (2011). Early Childhood Mathematics Intervention. *Science, 333*(6045), 968-970. doi:10.1126/science.1204537

Clements, D. H., Sarama, J., Wolfe, C. B., & Spitler, M. E. (2013). Longitudinal evaluation of a scale-up model for teaching mathematics with trajectories and technologies persistence of effects in the third year. *American Educational Research Journal, 50*(4), 812-850.

Clifford, R. M., & Crawford, G. M. (Eds.). (2009). *Beginning School: U.S. policies in international perspective*. New York: Teachers College Press.

Copple, C., & Bredekamp, S. (2009). *Developmentally appropriate practice in early childhood programs serving children from birth through age 8* (3rd ed.). Washington, DC: National Association for the Education of Young Children.

Deming, D. (2009). Early Childhood Intervention and Life-Cycle Skill Development: Evidence from Head Start. *American Economic Journal: Applied Economics, 1*(3), 111-134. doi:10.2307/25760174

DeVries, R., & Kohlberg, L. (1987). *Programs of early education: The constructivist view*: Longman White Plains, NY.

Diamond, A., Barnett, W. S., Thomas, J., & Munro, S. (2007). Preschool Program Improves Cognitive Control. *Science, 318*(5855), 1387-1388. doi:10.1126/science.1151148

Duncan, G. J., & Murnane, R. J. (2014). *Restoring Opportunity: The crisis of inequality and the challenge for American education*. Cambridge, MA, and New York, NY: Harvard Education Press and Russell Sage Foundation.

Dunn, L. M., & Dunn, L. M. (1997). *Peabody Picture Vocabulary Test--Third Edition (PPVT-III)*. Upper Saddle River, NJ: Pearson Publishing.

Fantuzzo, J. W., Gadsden, V. L., & McDermott, P. A. (2011). An Integrated Curriculum to Improve Mathematics, Language, and Literacy for Head Start Children. *American Educational Research Journal, 48*(3), 763-793. doi:10.3102/0002831210385446

Fuligni, A. S., Howes, C., Huang, Y., Hong, S. S., & Lara-Cinisomo, S. (2012). Activity settings and daily routines in preschool classrooms: Diverse experiences in early learning settings for low-income children. *Early Childhood Research Quarterly, 27*(2), 198-209.

Gormley, W. T. (2007). Early childhood care and education: Lessons and puzzles. *Journal of Policy Analysis and Management, 26*(3), 633-671.

Gresham, F. M., & Elliott, S. N. (1990). *Social skills rating system (SSRS)*. Circle Pines, MN: American Guidance Service.

Harms, T., Clifford, R. M., & Cryer, D. (1998). *Early childhood environment rating scale*. New York: Teachers College Press.

Heckman, J. J., Moon, S. H., Pinto, R., Savelyev, P. A., & Yavitz, A. (2010). The rate of return to the HighScope Perry Preschool Program. *Journal of Public Economics, 94*(1), 114-128.

Hulsey, L. K., Aikens, N., Kopack, A., West, J., Moiduddin, E., & Tarullo, L. B. (2011). *Head Start children, families, and programs: Present and past data from FACES*. Retrieved from Washington, DC:

Isaacs, J., Edelstein, S., Hahn, H., Steele, E., & Steuerle, C. E. (2015). *Kids Share 2015: Report on Federal Expenditures on Children in 2014 and Future Projections*.

Justice, L. M., McGinty, A. S., Cabell, S. Q., Kilday, C. R., Knighton, K., & Huffman, G. (2010). Language and literacy curriculum supplement for preschoolers who are academically at risk: a feasibility study. *Language, Speech & Hearing Services in Schools, 41*(2), 161-178. doi:10.1044/0161-1461(2009/08-0058)

Kane, T. J., Taylor, E. S., Tyler, J. H., & Wooten, A. L. (2011). Identifying effective classroom practices using student achievement data. *Journal of Human Resources, 46*(3), 587-613.

Klein, A., & Starkey, P. (2002). Child Math Asssessment-Abbreviated.

Klein, A., Starkey, P., Clements, D. H., Sarama, J., & Iyer, R. (2008). Effects of a Pre-Kindergarten Mathematics Intervention: A Randomized Experiment. *Journal of Research on Educational Effectiveness, 1*(3), 155-178. doi:10.1080/19345740802114533

Kline, P., & Walters, C. (forthcoming). Evaluating public programs with close substitutes: The case of Head Start. *Quarterly Journal of Economics*.

Landry, S. H., Crawford, A., Gunnewig, S. B., & Swank, P. R. (2002). Teacher Behavior Rating Scale (TBRS). *Center for Improving the Readiness of Children for Learning and Education, unpublished research instrument*.

Lonigan, C. J., Farver, J. M., Phillips, B. M., & Clancy-Menchetti, J. (2011). Promoting the development of preschool children's emergent literacy skills: A randomized evaluation of a literacy-focused curriculum and two professional development models. *Reading and Writing, 24*(3), 305-337. doi:10.1007/s11145-009-9214-6

Morris, P. A., Mattera, S. K., Castells, N., Bangser, M., Bierman, K. L., & Raver, C. C. (2014). *Impact findings from the Head Start CARES Demonstration: National evauation of three approaches to improving preschoolers' social and emotional competence.* Retrieved from Washington, DC:

Office of Head Start. (2010). Head Start Program Fact Sheet Fiscal Year 2010. Retrieved from http://eclkc.ohs.acf.hhs.gov/hslc/data/factsheets/fHeadStartProgr.htm

Piaget, J. (1976). *Piaget's theory*. New York, NY: Springer.

Pianta, R., Howes, C., Burchinal, M. R., Bryant, D., Clifford, R., Early, D. M., & Barbarin, O. A. (2005). Features of Pre-Kindergarten Programs, Classrooms, and Teachers: Do They Predict Observed Classroom Quality and Child-Teacher Interactions? *Applied Developmental Science, 9*(3), 144-159. doi:10.1207/s1532480xads0903_2

Preschool Curriculum Evaluation Research Consortium. (2008). *Effects of preschool curriculum programs on school readiness*. Retrieved from Washington, DC:

Reardon, S. F. (2011). The widening academic achievement gap between the rich and the poor: New evidence and possible explanations. In G. J. Duncan & R. J. Murnane (Eds.), *Whither opportunity* (pp. 91-116). New York, NY: Russell Sage Foundation.

Ritchie, S., & Willer, B. (2008). *Curriculum: A guide to the NAEYC early childhood program standard and related accreditation criteria*. Washington, D.C.: National Association for the Education of Young Children (NAEYC).

Schweinhart, L. J. (2005). *Lifetime Effects: The High/Scope Perry Preschool Study through Age 40*. High/Scope Educational Research Foundation, Ypsilanti, M.I.

Schweinhart, L. J., & Weikart, D. P. (1981). Effects of the Perry Preschool Program on Youths Through Age 15. *Journal of Early Intervention, 4*(1), 29-39. doi:10.1177/105381518100400105

U.S. Department of Education. (2013). *Early childhood education intervention report: The Creative Curriculum for Preschool, Fourth Edition.* Retrieved from Washington, DC:

Walters, C. (2014). Inputs in the Production of Early Childhood Human Capital: Evidence from Head Start. *National Bureau of Economic Research Working Paper Series, No. 20639*. doi:10.3386/w20639

Warner, M. (2007). Child care and economic development: Markets, households and public policy. *International Journal of Economic Development, 9*(3), 111-121.

Wasik, B. A., & Hindman, A. H. (2011). Improving vocabulary and pre-literacy skills of at-risk preschoolers through teacher professional development. *Journal of Educational Psychology, 103*(2), 455-469. doi:10.1037/a0023067

Weikart, D. P., & Schweinhart, L. (1987). The High/Scope cognitively oriented curriculum in early education. In J. L. Roopnarine & J. E. Johnson (Eds.), *Approaches to early childhood education* (pp. 253-268). New York: Merill/Macmillan.

Weiland, C., & Yoshikawa, H. (2013). Impacts of a Prekindergarten Program on Children's Mathematics, Language, Literacy, Executive Function, and Emotional Skills. *Child Development, 84*(6), 2112-2130. doi:10.1111/cdev.12099

Wong, V. C., Cook, T. D., Barnett, W. S., & Jung, K. (2008). An effectiveness-based evaluation of five state pre-kindergarten programs. *Journal of Policy Analysis and Management, 27*(1), 122-154.

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). Woodcock-Johnson-III Tests of Achievement. Itasca, IL: Riverside.

Yoshikawa, H., Weiland, C., Brooks-Gunn, J., Burchinal, M. R., Espinosa, L. M., Gormley, W., . . . Zaslow, M. J. (2013). *Investing in our future: The evidence base on preschool education*. Retrieved from New York, NY:

**Table 1: Description of curricula comparisons.**

| Grantee and sample size | Site | Treatment Curriculum | Control Curriculum(a) | Project-reported impacts | | | Pilot Year ? | Fidelity of Implementation | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | Literacy | Math | Socio-emotional | | Treatment | Control |
| **I. Literacy vs. HighScope and Creative Curriculum** | | | | | | | | | |
| University of North Florida n=250 | FL | Early Literacy and Learning Model | Creative Curriculum | ns, ns, ns | ns, ns | ns, ns | Y | 2.5 | Not Provided |
| Florida State University n=200 | FL | Literacy Express | HighScope | ns, ns, ns | ns, ns | ns, ns | | 2.5 | 2.0 |
| Florida State University n=200 | FL | DLM Early Childhood Express supplemented with Open Court Reading Pre-K | HighScope | +,+,+ | +, ns | ns, ns | | 2.3 | 2.0 |
| University of California-Berkeley n=290 | NJ | Ready Set Leap | HighScope | ns, ns, ns | ns, - | ns, ns | | 1.9 | 2.0 |
| University of Virginia n=200 | VA | Language Focused | HighScope | ns, ns, ns | ns, ns | ns, ns | | 2.0 | 2.0 |
| **II. Literacy vs. Locally-Developed Curriculum** | | | | | | | | | |
| University of Texas Health Science Center at Houston n=200 | TX | Doors to Discovery | Locally Developed | ns, ns, ns | ns, ns | ns, ns | Y | 2.1 | 1.0 |
| University of Texas Health Science Center at Houston n=200 | TX | Let's Begin with the Letter People | Locally Developed | ns, ns, ns | ns, ns | ns, ns | Y | 1.9 | 1.0 |
| Vanderbilt University n=210 | TN | Bright Beginnings | Locally Developed | ns, ns, ns | ns, ns | ns, ns | Y | 1.9 | 2.0 |
| **III. Math vs. HighScope and Creative Curriculum** | | | | | | | | | |
| University of California-Berkeley and SUNY University of Buffalo n=320 | CA and NY | Pre-K Mathematics supplemented with DLM Early Childhood Express (Math Software only) | Creative Curriculum or HighScope | ns, ns, ns | ns, + | ns, ns | Y | CA (2.7); NY (2.3) | CA (2.0); NY (2.0) |
| **IV. Creative Curriculum vs. Locally-Developed Curriculum** | | | | | | | | | |
| University of North Carolina at Charlotte n=310 | NC and GA | Creative Curriculum | Locally Developed | ns, ns, ns | ns, ns | ns, ns | Y | 2.1 | 1.5 |
| Vanderbilt n=210 | TN | Creative Curriculum | Locally Developed | ns, ns, ns | ns, ns | ns, ns | Y | 2.1 | 2.0 |

*Note*: "Literacy" outcomes include the PPVT, WJ Letter-Word and WJ Spelling. "Math" outcomes include WJ Applied problems and CMAA. "Socioemotional" outcomes include social skills and problem behaviors."+" indicates beneficial impact with $p<.05$; "-" indicates detrimental impact with $p<.05$; "ns" indicates $p>=.05$. Fidelity of implementation was rated on a 4-point scale (0 = Not at all; 3 = High). Ns are rounded to the nearest 10 in accordance with NCES data policies.

**Table 2. Effects of treatment curricula on classroom observational measures at the end of preschool**

| | ECERS total score | TBRS Math | TBRS Literacy | Arnett total score |
|---|---|---|---|---|
| **I. Literacy vs. HighScope and Creative Curriculum** | 0.25* | -0.14 | 0.07 | 0.18 |
| | (0.15) | (0.16) | (0.16) | (0.16) |
| N | 890 | 880 | 880 | 890 |
| *Classroom N= 100* | | | | |
| **II. Literacy vs. Locally-Developed Curricula** | 0.51** | 0.46 | 0.83** | 0.38 |
| | (0.23) | (0.32) | (0.37) | (0.25) |
| N | 460 | 440 | 440 | 440 |
| *Classroom N=60* | | | | |
| **III. Math vs. HighScope and Creative Curriculum** | 0.15 | 1.16** | 0.34 | 0.63 |
| | (0.32) | (0.52) | (0.31) | (0.52) |
| N | 210 | 200 | 200 | 200 |
| *Classroom N=30* | | | | |
| **IV. Creative Curriculum vs. Locally-Developed Curricula** | 0.61** | 0.51** | 0.71*** | 0.99** |
| | (0.23) | (0.23) | (0.17) | (0.36) |
| N | 350 | 320 | 320 | 330 |
| *Classroom N=30* | | | | |

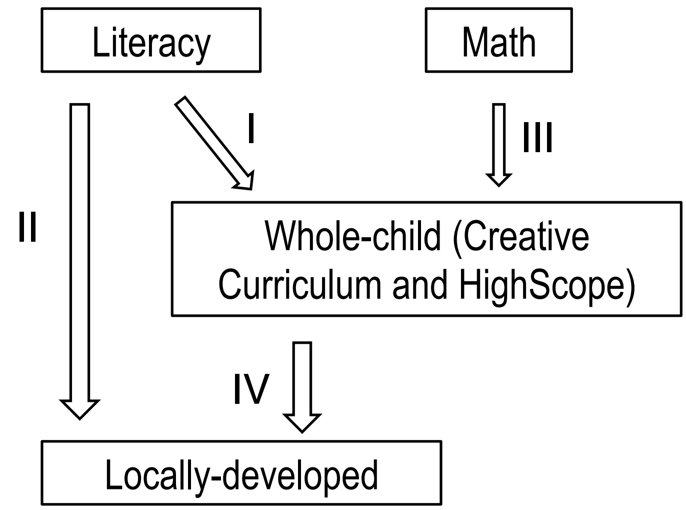
*Note.* Each entry represents results from a separate regression. Standard errors clustered at the classroom level are in parentheses. Fixed effects at the random assignment site level are included in all analyses. Child and family controls included for child gender, race, age (months), baseline achievement and social skills; parent/primary caregiver education (years), whether working, age (years), annual household income (thousands), and whether receiving welfare. Classroom observational measures at baseline, time in days from the start of the preschool year and the date of the observational assessment, a quadratic version of this time in days, and the time in days between a classroom's fall and spring observational assessment were also included for the estimates for the Arnett and ECERS. Duration of TBRS observation in minutes was included in TBRS Math and Literacy models. Missing dummy variables were included in the analyses to account for missing independent variables. Outcomes were standardized to have a mean of 0 and standard deviation of 1. Ns are rounded to the nearest 10 in accordance with NCES data policies.
* $p < .10$, ** $p < .05$, *** $p < .01$.

**Table 3. Effects of treatment curricula on child school readiness skills at the end of preschool**

| | Literacy composite | Math composite | Academic composite | Social skills composite |
|---|---|---|---|---|
| **I. Literacy vs. HighScope and Creative Curriculum** | 0.15*** | -0.01 | 0.06 | -0.13 |
| | (0.05) | (0.05) | (0.05) | (0.10) |
| N | 890 | 890 | 880 | 860 |
| **II. Literacy vs. Locally-Developed Curricula** | 0.15 | 0.14* | 0.15* | -0.18 |
| | (0.09) | (0.07) | (0.08) | (0.19) |
| N | 480 | 480 | 480 | 440 |
| **III. Math vs, HighScope and Creative Curriculum** | 0.05 | 0.35*** | 0.25** | 0.14 |
| | (0.10) | (0.11) | (0.11) | (0.17) |
| N | 220 | 220 | 220 | 210 |
| **IV. Creative Curriculum vs. Locally-Developed Curricula** | 0.02 | 0.02 | 0.02 | -0.03 |
| | (0.08) | (0.08) | (0.08) | (0.22) |
| N | 360 | 360 | 360 | 350 |

*Note.* Each entry represents results from a separate regression. Standard errors clustered at the classroom level are in parentheses. The literacy composite included PPVT, WJ Letter Word and WJ Spelling. The math composite included WJ Applied Problems, and CMAA. The academic composite weights the math and literacy composite scores equally. The social skills composite included teacher rated social skills and a reverse-coded teacher rated behavior problems (higher means fewer problems). Models include fixed effects for the unit of random assignment. Child and family controls included for child gender, race, age (months), baseline achievement and social skills; parent/primary caregiver education (years), whether working, age (years), annual household income (thousands), and whether receiving welfare. Missing dummy variables were included in the analyses to account for missing independent variables. Outcomes were standardized to have a mean of 0 and standard deviation of 1. Ns are rounded to the nearest 10 in accordance with NCES data policies.
$^{*}p < .10$, $^{**}p < .05$, $^{***}p < .01$.

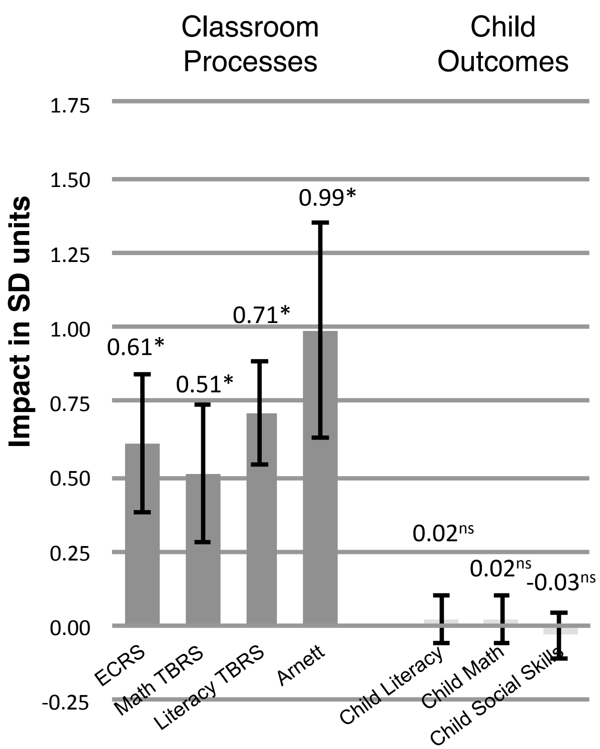
**Figure 1: Curricula comparisons in study sample**




Note: Comparison IV only involves the Creative Curriculum

*Notes*: All curricula comparisons are within-site comparisons of randomly assigned treatment-control conditions. Curricula and site-specific information are available in Table 1.

Figure 2. Classroom and Child Impacts of Creative Curriculum vs. Locally-Developed




Source: Tables 2 and 3, *p<.05

Figure 3. Classroom and Child Impacts of Math vs. Creative and HighScope Curricula

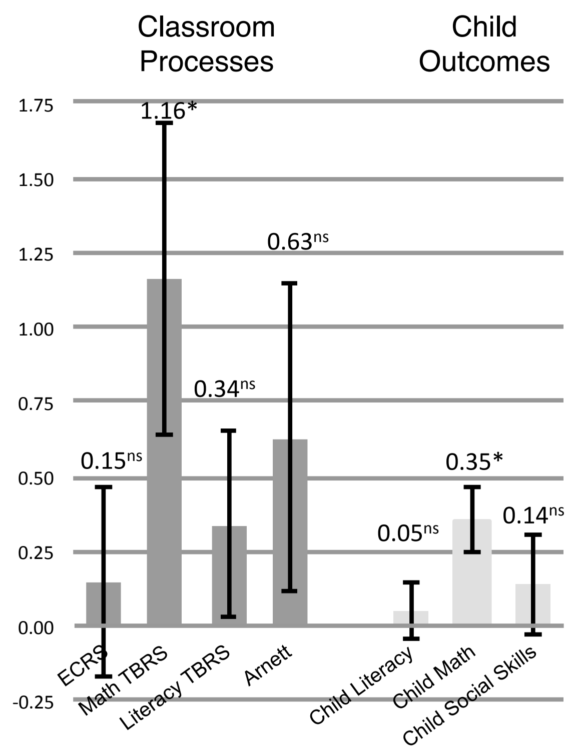
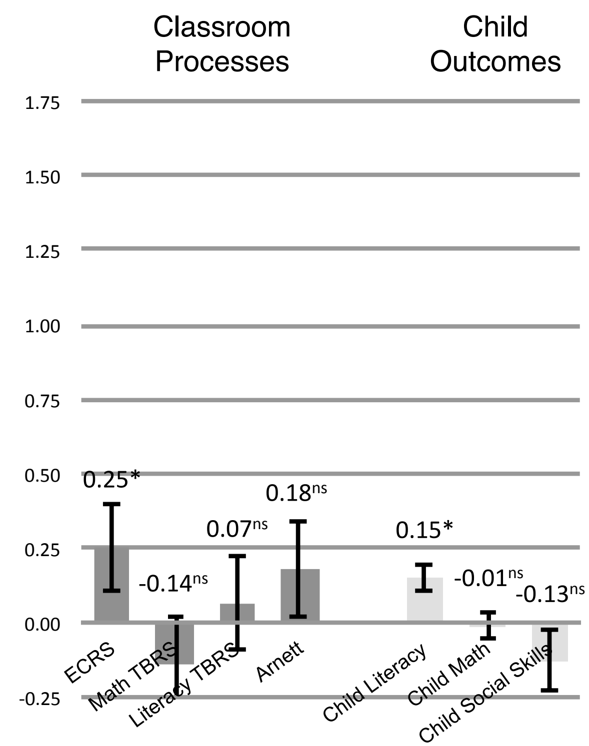



Figure 4. Classroom and Child Impacts of Literacy vs. Creative and HighScope Curricula




*Notes*: Bars show estimated impacts of various curricula comparisons on classroom process quality and child outcomes as measured by composite standardized scores of literacy skills, math skills and socioemotional skills. Each figure is from one of the curricula comparisons described in Figure 1, and each bar is from a separate regression (Comparison II is omitted). Standard error bars are shown for each estimate. *$p$<.05

APPENDIX

**Appendix Table 1. Description of classroom observational measures (process measures).**

| **Name of Measure** | **Abbreviation** | **Description of Measure** | **Items and Rating Scale** |
|---|---|---|---|
| Teacher Behavior Rating Scale (Landry, Crawford, Gunnewig, & Swank, 2002) | TBRS | Using the TBRS, trained observers rate the amount and quantity of academic activities present in a classroom. There are two content areas measured by the TBRS - math and literacy. | Quality of the activities were rated from 0-3 (0 = activity not present; 3 = activity high quality). Quantity of activities was similarly rated from 0-3 (0 = activity not present; 3 = activity happened often or many times). Reliability: Math scale, .94; Literacy scale, .87 |
| Early Childhood Environment Rating Scale - Revised (Harms, Clifford, & Cryer, 1998) | ECERS-R | This instrument measures the overall quality of the classroom including structural features (such as the availability of developmental materials in the classroom), and teacher-child interactions (including the use of language in the classroom). | Total score - 43 items; Provisions factor - 12 items; Interaction factor - 11 items. All items were rated by a trained observer on a scale from 1-7 (1 = inadequate quality; 7 = excellent quality. Reliability: Total score, .92; Provisions factor, .89; Interactions factor, .91 |
| Arnett Caregiver Interaction Scale (Arnett, 1989) | Arnett CIS | The Arnett CIS examines the positive interactions, harshness, detachment, and permissiveness between the teacher and children. | Total number of items - 26. Trained observers rated each item from 1-4 (1 = not true at all; 4 = very much true). Reliability: .95 |

**Appendix Table 2. Literacy curricula and control group summary statistics (Comparisons 1 and 2) for child and family background and demographic characteristics, classroom observations, and child school readiness skills**

| | I. Literacy Curricula Compared With HighScope and Creative Curricula | | | | | | | II.Literacy Curricula Compared With Locally Developed Curricula | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Literacy Curricula (Treat) | | | HighScope-Creative Curricula Comparison | | | | Literacy Curricula (Treat) | | | Locally Developed Curricula | | | |
| | N | Mean | SD | N | Mean | SD | P-value | N | Mean | SD | N | Mean | SD | P-value |
| **Covariates at Baseline (Fall 2003)** | | | | | | | | | | | | | | |
| Child Gender - Female | 550 | 0.48 | | 340 | 0.44 | | 0.30 | 290 | 0.45 | 0.50 | 190 | 0.48 | 0.50 | 0.48 |
| Child Race - Black | 550 | 0.55 | | 340 | 0.52 | | 0.96 | 290 | 0.08 | 0.27 | 190 | 0.10 | 0.30 | 0.43 |
| Child Race - Asian | 550 | 0.00 | | 340 | 0.00 | | 1.00 | 290 | 0.01 | 0.12 | 190 | 0.03 | 0.18 | 0.27 |
| Child Race - Hispanic | 550 | 0.11 | | 340 | 0.10 | | 0.54 | 290 | 0.28 | 0.45 | 190 | 0.23 | 0.42 | 0.92 |
| Child Race - Other | 550 | 0.04 | | 340 | 0.03 | | 0.60 | 290 | 0.05 | 0.22 | 190 | 0.05 | 0.22 | 0.67 |
| Child Age (months) | 550 | 54.66 | 3.78 | 340 | 54.82 | 3.92 | 0.39 | 290 | 54.73 | 3.74 | 190 | 54.93 | 3.72 | 0.24 |
| Parent Education (years) | 550 | 12.90 | 1.59 | 340 | 12.53 | 1.68 | 0.02 | 290 | 13.23 | 2.06 | 190 | 12.92 | 1.64 | 0.48 |
| Parent Working | 550 | 0.62 | | 340 | 0.58 | | 0.31 | 290 | 0.51 | 0.50 | 190 | 0.48 | 0.50 | 0.58 |
| Parent Age (years) | 550 | 30.95 | 7.18 | 340 | 30.68 | 6.81 | 0.69 | 290 | 32.71 | 6.31 | 190 | 32.59 | 6.80 | 0.58 |
| Annual Household Income (thousands) | 550 | 27.08 | 17.19 | 340 | 26.08 | 17.87 | 0.27 | 290 | 39.85 | 27.17 | 190 | 32.57 | 20.82 | 0.24 |
| Receiving Welfare | 550 | 0.12 | | 340 | 0.18 | | 0.12 | 290 | 0.07 | 0.26 | 190 | 0.08 | 0.27 | 0.77 |
| **Classroom Observations - Fall 2003** | | | | | | | | | | | | | | |
| CIS Arnett Total | 550 | 3.22 | 0.41 | 340 | 3.41 | 0.33 | 0.37 | 280 | 3.33 | 0.44 | 180 | 3.00 | 0.58 | 0.02 |
| ECERS Total | 550 | 4.59 | 1.17 | 340 | 5.02 | 1.17 | 0.15 | 280 | 3.63 | 0.65 | 190 | 3.14 | 0.53 | 0.00 |
| ECERS Provisions | 550 | 4.55 | 1.42 | 340 | 5.15 | 1.26 | 0.59 | 280 | 3.47 | 0.69 | 190 | 2.97 | 0.46 | 0.00 |
| ECERS Interaction | 550 | 5.15 | 1.36 | 340 | 5.57 | 1.29 | 0.46 | 280 | 4.54 | 1.23 | 190 | 3.60 | 0.90 | 0.00 |
| **Classroom Observations - Spring 2004** | | | | | | | | | | | | | | |
| TBRS Math Quality | 550 | 0.99 | 0.64 | 340 | 1.14 | 0.66 | 0.65 | 260 | 1.34 | 0.76 | 190 | 0.98 | 0.50 | 0.13 |
| TBRS Math Quantity | 550 | 1.11 | 0.48 | 340 | 1.20 | 0.57 | 0.87 | 260 | 1.33 | 0.56 | 190 | 1.07 | 0.40 | 0.15 |
| TBRS Literacy Quality | 550 | 1.58 | 0.44 | 340 | 1.59 | 0.40 | 0.86 | 260 | 1.75 | 0.52 | 190 | 1.31 | 0.35 | 0.01 |
| TBRS Literacy Quantity | 550 | 1.50 | 0.59 | 340 | 1.53 | 0.55 | 0.49 | 260 | 1.69 | 0.68 | 190 | 1.12 | 0.45 | 0.01 |
| CIS Arnett Total | 550 | 3.16 | 0.47 | 340 | 3.17 | 0.37 | 0.17 | 260 | 3.34 | 0.49 | 190 | 3.09 | 0.55 | 0.2 |
| ECERS Total | 550 | 4.36 | 1.14 | 340 | 4.32 | 0.97 | 0.16 | 280 | 3.99 | 0.80 | 190 | 3.53 | 0.78 | 0.06 |
| ECERS Provisions | 550 | 4.37 | 1.19 | 340 | 4.40 | 1.03 | 0.57 | 280 | 3.92 | 0.87 | 190 | 3.36 | 0.64 | 0.01 |
| ECERS Interaction | 550 | 4.89 | 1.45 | 340 | 4.94 | 1.22 | 0.09 | 280 | 4.95 | 1.28 | 190 | 4.32 | 1.37 | 0.12 |
| **Child Outcomes - Fall 2003** | | | | | | | | | | | | | | |
| PPVT | 550 | 87.81 | 13.23 | 340 | 86.48 | 14.93 | 0.04 | 290 | 87.89 | 18.76 | 190 | 89.52 | 19.15 | 0.39 |
| WJ Letter Word | 550 | 100.78 | 15.76 | 340 | 99.18 | 15.20 | 0.30 | 290 | 94.85 | 16.17 | 190 | 98.30 | 15.58 | 0.42 |
| WJ Spelling | 550 | 94.97 | 13.85 | 340 | 94.61 | 14.57 | 0.35 | 290 | 90.69 | 13.16 | 190 | 91.81 | 12.62 | 0.81 |
| WJ Applied Problems | 550 | 92.44 | 13.70 | 340 | 92.14 | 13.28 | 0.61 | 290 | 94.79 | 16.66 | 190 | 95.17 | 16.58 | 0.94 |
| CMAA Composite | 550 | 0.42 | 0.25 | 340 | 0.43 | 0.23 | 0.89 | 290 | 0.39 | 0.25 | 190 | 0.42 | 0.25 | 0.90 |
| Social Skills (teacher report) | 500 | 100.42 | 16.10 | 310 | 100.69 | 15.23 | 0.98 | 280 | 101.69 | 21.15 | 180 | 98.55 | 15.11 | 0.28 |
| Behavior Problems (teacher report) | 530 | 100.31 | 13.54 | 310 | 100.91 | 12.74 | 0.08 | 280 | 102.58 | 15.56 | 180 | 99.58 | 13.12 | 0.55 |
| **Child Outcomes - Spring 2004** | | | | | | | | | | | | | | |
| PPVT | 550 | 91.89 | 13.92 | 340 | 90.29 | 14.84 | 0.08 | 290 | 92.70 | 16.73 | 190 | 94.95 | 17.50 | 0.29 |
| WJ Letter Word | 540 | 105.15 | 13.60 | 340 | 102.85 | 14.07 | 0.00 | 290 | 101.37 | 14.08 | 190 | 105.64 | 14.26 | 0.16 |
| WJ Spelling | 520 | 96.95 | 14.59 | 320 | 93.60 | 14.86 | 0.00 | 290 | 93.97 | 13.18 | 190 | 97.91 | 12.92 | 0.13 |
| WJ Applied Problems | 540 | 93.78 | 13.28 | 330 | 91.73 | 13.74 | 0.02 | 290 | 97.80 | 16.66 | 190 | 99.53 | 13.73 | 0.51 |
| CMAA Composite | 550 | 0.58 | 0.23 | 340 | 0.59 | 0.22 | 0.96 | 290 | 0.59 | 0.26 | 190 | 0.65 | 0.23 | 0.26 |
| Social Skills (teacher report) | 530 | 103.91 | 15.65 | 320 | 107.57 | 15.56 | 0.13 | 270 | 110.74 | 13.80 | 170 | 106.74 | 14.77 | 0.09 |
| Behavior Problems (teacher report) | 530 | 101.64 | 13.59 | 330 | 101.22 | 13.75 | 0.79 | 270 | 99.36 | 12.71 | 170 | 99.73 | 13.71 | 0.96 |
| Literacy composite score | 550 | 0.11 | 0.95 | 340 | -0.10 | 1.01 | 0.00 | 290 | -0.07 | 1.06 | 190 | 0.23 | 1.05 | 0.14 |
| Math composite score | 550 | -0.06 | 0.95 | 340 | -0.11 | 0.94 | 0.17 | 290 | 0.12 | 1.19 | 190 | 0.34 | 0.98 | 0.36 |
| Academic composite score | 540 | 0.01 | 0.94 | 340 | -0.11 | 0.97 | 0.02 | 290 | 0.04 | 1.17 | 190 | 0.32 | 1.02 | 0.24 |
| Social skills composite score | 530 | -0.16 | 0.99 | 330 | -0.00 | 1.03 | 0.30 | 270 | 0.19 | 0.91 | 170 | 0.02 | 1.01 | 0.36 |

*Note.* TBRS = Teacher Behavior Rating Scale. TBRS Literacy variables are composites of oral language, book reading, written expression, and print and letter knowledge. Further detail on classroom observational measures is available in Table S2. *p*-values account for clustering by random assignment site and date of classroom observational assessment (for classroom observation t-tests only). Ns are rounded to the nearest 10 in accordance with NCES data policies.

**Appendix Table 3. Math and Creative Curriculum and control group summary statistics (Comparisons III and IV) on child and family background and demographic characteristics, classroom observations, and child school readiness skills**

| | III. Math Curriculum Compared With HighScope and Creative Curricula | | | | | | | IV. Creative Curriculum Compared with Locally Developed Curricula | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Math Curriculum (Treat) | | | HighScope-Creative Curricula Comparison | | | | Locally Developed Curricula | | | Creative Curriculum (Treat) | | | |
| | N | Mean | SD | N | Mean | SD | P-value | N | Mean | SD | N | Mean | SD | P-value |
| **Covariates at Baseline (Fall 2003)** | | | | | | | | | | | | | | |
| Child Gender - Female | 110 | 0.58 | | 110 | 0.45 | | 0.04 | 270 | 0.53 | | 260 | 0.52 | | 0.62 |
| Child Race - Black | 110 | 0.40 | | 110 | 0.34 | | 0.58 | 270 | 0.51 | | 260 | 0.54 | | 0.97 |
| Child Race - Asian | 110 | 0.06 | | 110 | 0.01 | | 0.32 | 270 | 0.00 | | 260 | 0.00 | | 0.33 |
| Child Race - Hispanic | 110 | 0.24 | | 110 | 0.27 | | 0.77 | 270 | 0.08 | | 260 | 0.09 | | 0.65 |
| Child Race - Other | 110 | 0.05 | | 110 | 0.15 | | 0.10 | 270 | 0.03 | | 260 | 0.03 | | 0.75 |
| Child Age (months) | 110 | 53.20 | 3.26 | 110 | 52.64 | 3.31 | 0.26 | 270 | 53.98 | 3.58 | 260 | 54.22 | 3.63 | 0.47 |
| Parent Education (years) | 110 | 13.08 | 1.78 | 110 | 12.49 | 1.81 | 0.08 | 270 | 12.65 | 1.49 | 260 | 12.60 | 1.60 | 0.36 |
| Parent Working | 110 | 0.57 | | 110 | 0.43 | | 0.10 | 270 | 0.49 | | 260 | 0.47 | | 0.52 |
| Parent Age (years) | 110 | 32.94 | 9.12 | 110 | 31.96 | 7.60 | 0.42 | 270 | 31.86 | 7.67 | 260 | 31.36 | 7.51 | 0.04 |
| Annual Household Income (thousands) | 110 | 29.51 | 17.51 | 110 | 24.66 | 13.80 | 0.06 | 270 | 23.07 | 14.98 | 260 | 21.80 | 14.47 | 0.22 |
| Receiving Welfare | 110 | 0.16 | | 110 | 0.18 | | 0.72 | 270 | 0.13 | | 260 | 0.12 | | 0.54 |
| **Classroom Observations - Fall 2003** | | | | | | | | | | | | | | |
| CIS Arnett Total | 100 | 3.15 | 0.37 | 110 | 3.18 | 0.59 | 0.23 | 260 | 2.90 | 0.66 | 250 | 3.07 | 0.71 | 0.04 |
| ECERS Total | 100 | 3.48 | 0.67 | 110 | 3.79 | 0.84 | 0.26 | 270 | 3.40 | 0.94 | 260 | 3.92 | 1.03 | 0.03 |
| ECERS Provisions | 100 | 3.50 | 0.62 | 110 | 3.81 | 0.85 | 0.34 | 270 | 3.36 | 0.94 | 260 | 4.01 | 1.07 | 0.04 |
| ECERS Interaction | 100 | 4.04 | 1.06 | 110 | 4.45 | 1.41 | 0.27 | 270 | 3.97 | 1.50 | 260 | 4.70 | 1.59 | 0.01 |
| **Classroom Observations - Spring 2004** | | | | | | | | | | | | | | |
| TBRS Math Quality | 100 | 1.21 | 0.94 | 110 | 0.73 | 0.48 | 0.31 | 260 | 1.00 | 0.71 | 230 | 1.18 | 0.77 | 0.02 |
| TBRS Math Quantity | 100 | 1.26 | 0.69 | 110 | 0.95 | 0.33 | 0.86 | 260 | 1.13 | 0.60 | 230 | 1.23 | 0.63 | 0.01 |
| TBRS Literacy Quality | 100 | 1.12 | 0.33 | 110 | 1.14 | 0.39 | 0.66 | 260 | 1.32 | 0.34 | 230 | 1.39 | 0.40 | 0.00 |
| TBRS Literacy Quantity | 100 | 1.01 | 0.35 | 110 | 0.93 | 0.41 | 0.87 | 260 | 1.07 | 0.40 | 230 | 1.27 | 0.48 | 0.01 |
| CIS Arnett Total | 100 | 3.06 | 0.63 | 110 | 2.97 | 0.61 | 0.92 | 270 | 3.10 | 0.53 | 240 | 3.14 | 0.53 | 0.01 |
| ECERS Total | 110 | 3.81 | 0.95 | 110 | 3.66 | 0.87 | 0.83 | 270 | 3.77 | 0.79 | 260 | 3.96 | 0.78 | 0.01 |
| ECERS Provisions | 110 | 3.67 | 1.03 | 110 | 3.55 | 0.74 | 0.91 | 270 | 3.64 | 0.81 | 260 | 3.97 | 0.89 | 0.00 |
| ECERS Interaction | 110 | 4.66 | 1.44 | 110 | 4.39 | 1.40 | 0.25 | 270 | 4.40 | 1.21 | 260 | 4.79 | 1.30 | 0.91 |
| **Child Outcomes - Fall 2003** | | | | | | | | | | | | | | |
| PPVT | 110 | 89.28 | 12.60 | 110 | 92.44 | 14.29 | 0.25 | 270 | 85.50 | 15.94 | 260 | 85.35 | 16.01 | 0.91 |
| WJ Letter Word | 110 | 102.86 | 17.40 | 110 | 101.65 | 13.92 | 0.74 | 270 | 93.95 | 16.65 | 260 | 95.75 | 16.44 | 0.69 |
| WJ Spelling | 110 | 95.30 | 14.09 | 110 | 91.99 | 12.01 | 0.27 | 270 | 89.73 | 13.13 | 260 | 89.64 | 12.84 | 0.86 |
| WJ Applied Problems | 110 | 99.78 | 12.84 | 110 | 96.44 | 14.34 | 0.17 | 270 | 90.67 | 15.63 | 260 | 91.17 | 14.73 | 0.44 |
| CMAA Composite | 110 | 0.44 | 0.24 | 110 | 0.44 | 0.24 | 0.97 | 270 | 0.31 | 0.22 | 260 | 0.32 | 0.22 | 0.73 |
| Social Skills (teacher report) | 110 | 106.06 | 13.46 | 110 | 106.61 | 15.68 | 0.86 | 270 | 101.29 | 19.12 | 260 | 100.29 | 16.72 | 0.98 |
| Behavior Problems (teacher report) | 110 | 96.00 | 12.10 | 110 | 96.75 | 14.29 | 0.80 | 270 | 101.34 | 14.89 | 260 | 100.97 | 14.17 | 0.77 |
| **Child Outcomes - Spring 2004** | | | | | | | | | | | | | | |
| PPVT | 110 | 94.84 | 13.02 | 110 | 93.67 | 14.95 | 0.68 | 260 | 88.81 | 14.98 | 260 | 90.13 | 14.86 | 0.15 |
| WJ Letter Word | 110 | 101.46 | 14.15 | 110 | 100.80 | 14.33 | 0.83 | 270 | 99.20 | 12.83 | 260 | 100.00 | 12.07 | 0.97 |
| WJ Spelling | 110 | 95.90 | 13.27 | 110 | 92.95 | 11.99 | 0.25 | 270 | 90.58 | 13.53 | 260 | 91.44 | 13.02 | 0.96 |
| WJ Applied Problems | 110 | 98.81 | 13.43 | 110 | 94.69 | 13.07 | 0.17 | 270 | 92.70 | 14.73 | 260 | 93.36 | 13.23 | 0.27 |
| CMAA Composite | 110 | 0.66 | 0.21 | 110 | 0.54 | 0.21 | 0.00 | 270 | 0.46 | 0.28 | 260 | 0.48 | 0.27 | 0.86 |
| Social Skills (teacher report) | 110 | 113.37 | 12.08 | 100 | 108.93 | 14.77 | 0.19 | 260 | 107.86 | 13.76 | 250 | 107.18 | 14.27 | 0.99 |
| Behavior Problems (teacher report) | 110 | 96.15 | 12.77 | 100 | 98.98 | 13.92 | 0.40 | 260 | 99.30 | 12.96 | 250 | 99.87 | 13.49 | 0.89 |
| Literacy composite score | 110 | 0.05 | 0.96 | 110 | -0.09 | 0.95 | 0.49 | 270 | -0.36 | 1.00 | 260 | -0.27 | 0.94 | 0.70 |
| Math composite score | 110 | 0.34 | 0.89 | 110 | -0.11 | 0.85 | 0.02 | 270 | -0.37 | 1.13 | 260 | -0.32 | 1.04 | 0.48 |
| Academic composite score | 110 | 0.24 | 0.91 | 110 | -0.11 | 0.88 | 0.07 | 270 | -0.40 | 1.08 | 260 | -0.32 | 0.99 | 0.55 |
| Social skills composite score | 110 | 0.41 | 0.91 | 100 | 0.13 | 1.02 | 0.26 | 260 | 0.08 | 0.91 | 250 | 0.03 | 0.96 | 0.93 |

*Note.* TBRS = Teacher Behavior Rating Scale. TBRS Literacy variables are composites of oral language, book reading, written expression, and print and letter knowledge. Further detail on classroom observational measures is available in Table S2. *p*-values account for clustering by random assignment site and date of classroom observational assessment (for classroom observation t-tests only). Ns are rounded to the nearest 10 in accordance with NCES data policies.

**Appendix Table 4. Effects of treatment curricula on classroom observational measures at the end of preschool**

| | ECERS total score | ECERS Provisions | ECERS Interactions | TBRS Math Quality | TBRS Math Quantity | TBRS Literacy Quality | TBRS Literacy Quantity | Arnett total score |
|---|---|---|---|---|---|---|---|---|
| **I. Literacy vs. HighScope and Creative Curriculum** | 0.25* | 0.22 | 0.16 | -0.16 | -0.12 | 0.09 | 0.05 | 0.18 |
| | (0.15) | (0.15) | (0.16) | (0.16) | (0.17) | (0.16) | (0.17) | (0.16) |
| N | 890 | 890 | 890 | 890 | 890 | 890 | 890 | 890 |
| *Classroom N= 100* | | | | | | | | |
| **II. Literacy vs. Locally-Developed Curricula** | 0.51** | 0.53** | 0.47* | 0.53 | 0.40 | 0.79** | 0.88** | 0.38 |
| | (0.23) | (0.27) | (0.23) | (0.33) | (0.33) | (0.38) | (0.37) | (0.25) |
| N | 210 | 210 | 210 | 200 | 200 | 200 | 200 | 200 |
| *Classroom N=60* | | | | | | | | |
| **III. Math vs. HighScope and Creative Curriculum** | 0.15 | 0.20 | 0.26 | 1.24** | 1.09* | 0.39 | 0.30 | 0.63 |
| | (0.32) | (0.26) | (0.40) | (0.51) | (0.53) | (0.28) | (0.35) | (0.52) |
| N | 350 | 350 | 350 | 320 | 320 | 320 | 320 | 330 |
| *Classroom N=30* | | | | | | | | |
| **IV. Creative vs. Locally-Developed Curricula** | 0.61** | 0.43* | 0.83*** | 0.51** | 0.51* | 0.74*** | 0.67*** | 0.99** |
| | (0.23) | (0.22) | (0.24) | (0.21) | (0.27) | (0.17) | (0.21) | (0.36) |
| | 320 | 320 | 330 | 350 | 350 | 350 | 320 | 320 |

*Note.* Each entry represents results from a separate regression. Standard errors clustered at the classroom level are in parentheses. Fixed effects at the random assignment site level are included in all analyses. Child and family controls included child gender, race, age (months), baseline achievement and social skills; parent/primary caregiver education (years), whether working, age (years), annual household income (thousands), and whether receiving welfare. Classroom observational measures at baseline, time in days from the start of the preschool year and the date of the observational assessment, a quadratic version of this time in days, and the time in days between a classroom's fall and spring observational assessment were also included in all models (Arnett and ECERS). Duration of TBRS observation in minutes was included in TBRS Math and Literacy models. TBRS Math is composite of quantity and quality of math activities, and TBRS Literacy is a composite of literacy (oral language, book reading, written expression, and print and letter knowledge) quantity and quality activities. TBRS = Teacher Behavior Rating Scale. Further detail on classroom observational measures is available in Table S2. Missing dummy variables were included in the analyses to account for missing independent variables. Outcomes were standardized to have a mean of 0 and standard deviation of 1. Ns are rounded to the nearest 10 in accordance with NCES data policies. *$p < .10$, **$p < .05$, ***$p < .01$.

**Appendix Table 5. Effect of treatment curricula on child school readiness skills at the end of preschool, by outcome component measures**

| | PPVT | WJ Letter-Word | WJ Spelling | WJ Applied Problems | CMAA | Social Skills | Problem Behaviors |
|---|---|---|---|---|---|---|---|
| | 0.06 | 0.10 | 0.18** | 0.09 | -0.09 | -0.25** | -0.00 |
| **I. Literacy vs. HighScope and Creative Curriculum** | (0.04) | (0.06) | (0.07) | (0.06) | (0.06) | (0.10) | (0.10) |
| N | 890 | 880 | 830 | 870 | 890 | 850 | 860 |
| **II. Literacy vs. Locally-Developed Curricula** | 0.06 | 0.14 | 0.16* | 0.06 | 0.18*** | -0.27 | 0.18 |
| | (0.08) | (0.11) | (0.08) | (0.09) | (0.07) | (0.19) | (0.19) |
| N | 470 | 480 | 480 | 480 | 480 | 440 | 440 |
| | 0.16* | -0.09 | 0.07 | 0.27** | 0.35*** | 0.29 | -0.15 |
| **III. Math vs. HighScope and Creative Curriculum** | (0.08) | (0.10) | (0.10) | (0.13) | (0.10) | (0.18) | (0.16) |
| N | 220 | 220 | 220 | 220 | 220 | 210 | 210 |
| **IV. Creative Curriculum vs. Locally-Developed Curricula** | 0.12* | -0.04 | -0.05 | 0.10 | -0.07 | -0.05 | 0.05 |
| | (0.07) | (0.09) | (0.09) | (0.08) | (0.09) | (0.21) | (0.19) |
| N | 360 | 360 | 360 | 360 | 360 | 350 | 350 |

*Note.* Each entry represents results from a separate regression. Standard errors clustered at the classroom level are in parentheses. Fixed effects at the random assignment site level are included in all analyses. Standard errors are clustered at the classroom level. Models include fixed effects for the unit of random assignment (i.e. grantee, school). Child and family controls included child gender, race, age (months), baseline achievement and social skills; parent/primary caregiver education (years), whether working, age (years), annual household income (thousands), and whether receiving welfare. Missing dummy variables were included in the analyses to account for missing independent variables. Outcomes were standardized to have a mean of 0 and standard deviation of 1. Ns are rounded to the nearest 10 in accordance with NCES data policies.

*p < .10, **p < .05, ***p < .01.

**Appendix Table 6. Alternate constructions of the math control group in the New York site: effects on composite outcomes**

| | | Literacy composite | Math composite | Academic composite | Social skills composite |
|---|---|---|---|---|---|
| **NY Math treatment group with NY control group that includes Head Start classrooms implementing High/Scope and Creative Curriculum, excluding NY Pre-k control classrooms (Same as second row in Table 4)** | | 0.05<br>(0.10) | 0.35***<br>(0.11) | 0.25**<br>(0.11) | 0.21<br>(0.24) |
| | N | 220 | 220 | 220 | 210 |
| **NY Math treatment group included, all NY control classrooms excluded** | | 0.11<br>(0.13) | 0.35*<br>(0.18) | 0.27<br>(0.16) | -0.04<br>(0.38) |
| | N | 210 | 210 | 210 | 200 |
| **Only CA math site** | | 0.06<br>(0.12) | 0.30*<br>(0.17) | 0.23<br>(0.15) | -0.01<br>(0.31) |
| | N | 150 | 150 | 150 | 150 |

*Note.* Each entry represents results from a separate regression. Standard errors clustered at the classroom level are in parentheses. Fixed effects at the random assignment site level are included in all analyses. Reference group is Creative Curriculum or High/Scope. Standard errors are clustered at the classroom level. Literacy composite included PPVT, WJ Letter Word and WJ Spelling. Math composite included WJ Applied Problems, and CMAA. Academic composite weights the math and literacy composites equally. The social skills composite included teacher rated social skills and behavior problems (reverse coded). Models include fixed effects for the unit of random assignment (i.e. grantee, school). Child and family controls included child gender, race, age (months), baseline achievement and social skills; parent/primary caregiver education (years), whether working, age (years), annual household income (thousands), and whether receiving welfare. Missing dummy variables were included in the analyses to account for missing independent variables. Outcomes were standardized to have a mean of 0 and standard deviation of 1. Ns are rounded to the nearest 10 in accordance with NCES data policies. *p < .10, **p < .05, ***p < .01.

**Appendix Table 7. Effects of PCER treatment curricula on raw outcome scores: Effect sizes calculated based on national standard deviation**

| | | PPVT | WJ Letter-Word | WJ Spelling | WJ Applied Problems | CMAA | Social Skills | Problem Behaviors |
|---|---|---|---|---|---|---|---|---|
| **I. Literacy vs. HighScope and Creative Curriculum** | | 0.06 | 0.10 | 0.18** | 0.09 | -0.09 | -0.25** | -0.001 |
| | | (0.04) | (0.07) | (0.07) | (0.06) | (0.06) | (0.10) | (0.10) |
| | N | 890 | 880 | 830 | 870 | 890 | 850 | 860 |
| **II. Literacy vs. Locally-Developed Curricula** | | 0.06 | 0.14 | 0.16* | 0.06 | 0.18*** | -0.27 | 0.18 |
| | | (0.08) | (0.11) | (0.08) | (0.09) | (0.07) | (0.19) | (0.19) |
| | N | 480 | 480 | 480 | 480 | 480 | 450 | 450 |
| **III. Math vs. HighScope and Creative Curriculum** | | 0.16* | -0.09 | 0.07 | 0.27** | 0.35*** | 0.29 | -0.15 |
| | | (0.08) | (0.10) | (0.10) | (0.13) | (0.10) | (0.18) | (0.16) |
| | N | 220 | 220 | 220 | 220 | 220 | 210 | 210 |
| **IV. Creative Curriculum vs. Locally-Developed Curricula** | | 0.12* | -0.04 | -0.05 | 0.10 | -0.07 | -0.05 | 0.05 |
| | | (0.07) | (0.09) | (0.09) | (0.08) | (0.09) | (0.21) | (0.19) |
| | N | 360 | 360 | 360 | 360 | 360 | 350 | 350 |

*Note.* Each entry represents results from a separate regression. Standard errors clustered at the classroom level are in parentheses. Models include fixed effects for the unit of random assignment (i.e. grantee, school). Child and family controls included child gender, race, age (months), baseline achievement and social skills; parent/primary caregiver education (years), whether working, age (years), annual household income (thousands), and whether receiving welfare. Missing dummy variables were included in the analyses to account for missing independent variables. Outcomes were standardized to have a mean of 0 and standard deviation of 1. Ns are rounded to the nearest 10 in accordance with NCES data policies. *$p < .10$, **$p < .05$, ***$p < .01$.