



Published in final edited form as:

Econ Educ Rev. 2018 August ; 65: 107–125. doi:10.1016/j.econedurev.2018.05.001.

Boosting School Readiness: Should Preschool Teachers Target Skills or the Whole Child?

Jade M. Jenkins^a, Greg J. Duncan^b, Anamarie Auger^c, Marianne Bitler^d, Thurston Domina^e, and Margaret Burchinal^f

^aUniversity of California, Irvine, 3200 Education, Irvine, CA 92697 (jvjenkin@uci.edu) ^bUniversity of California, Irvine, 3200 Education, Irvine, CA 92697 (gduncan@uci.edu) ^cRAND Corporation, 1776 Main St, Santa Monica, CA 90401 (aauger@rand.org) ^dUniversity of California, Davis & NBER, 1 Shields Avenue, Davis, CA 95616 (bitlter@ucdavis.edu) ^eUniversity of North Carolina at Chapel Hill, CB 3500, Chapel Hill, NC 27599 (tdomina@email.unc.edu) ^fUniversity of North Carolina at Chapel Hill, CB 8185, Chapel Hill, NC 27599 (burchinal@unc.edu)

Abstract

We use experimental data to estimate impacts on school readiness of different kinds of preschool curricula – a largely neglected preschool input and measure of preschool quality. We find that the widely-used “whole-child” curricula found in most Head Start and pre-K classrooms produced higher classroom process quality than did locally-developed curricula, but failed to improve children’s school readiness. A curriculum focused on building mathematics skills increased both classroom math activities and children’s math achievement relative to the whole-child curricula. Similarly, curricula focused on literacy skills increased literacy achievement relative to whole-child curricula, despite failing to boost measured classroom process quality.

Keywords

I28; J00; I20

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 in the 1960s and 1970s—Abecedarian and Perry Preschool (Campbell et al. 2008, Belfield et al. 2006, Campbell et al. 2014, Heckman et al. 2010, Anderson 2008, Conti, Heckman, and Pinto 2015). Growth in cognitive and noncognitive skills across a preschool academic year depends first and foremost on the amount and quality of the learning experiences in the classroom. Policy approaches to improving these learning experiences include reducing class size (in Kindergarten; Chetty et al. 2011), regulating the health, safety and, increasingly, the “process quality” of preschool classrooms through state Quality Rating Improvement Systems (QRIS; Sabol et al. 2013), and, less successfully for elementary and secondary

education, increasing teacher qualifications and/or pay (Jackson, Rockoff, and Staiger 2014).¹

We focus on a different and relatively neglected determinant of the quality of learning experiences: the content and style of instruction (known in schools and the education literature as the curriculum). Curricula provide teachers with day-to-day plans on what and how to teach. These include daily lesson plans, project materials, and other pedagogical tools. Instructional materials and the strategies promoted by curricula constitute some of the most direct and policy-relevant connections to learning activities in the classroom. The majority of early childhood education programs use curricula that are developed by publishers and marketed to teachers and schools to guide student learning. In most cases, educators choose among preselected curricular options based federal, state, or local policies, with little scientific guidance, a few popular selections, and substantial costs. Commonly-used preschool curricula range from \$1100-\$4100 per classroom, making curricula a \$100 million investment for the Head Start program alone.

Most preschool classrooms in the United States use what are typically described as “whole-child” or “global” curricula. Indeed, 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 whole-child instructional materials as well. Rather than directing teachers in their explicit academic instruction, this model seeks to promote learning by encouraging children to engage independently in a classroom stocked with prescribed toys and materials designed to promote noncognitive and, in some cases, cognitive skills. The whole-child approach, as embodied in an early version of the HighScope curriculum, was used in the very successful Perry Preschool program (Schweinhart and Weikart 1981).

The whole-child approach is grounded in a rich body of research from psychology on child development (Piaget 1976, DeVries and Kohlberg 1987, Weikart and Schweinhart 1987), but typically provides only very general guidance for teachers’ daily efforts. Given this lack of specific guidance, it requires a great deal of skill on the part of the teacher to translate child-chosen activities into cumulative growth in students’ cognitive and noncognitive (a.k.a. socioemotional) skills across the school year (Hart, Burts, and Charlesworth 1997). Descriptive studies of whole-child preschool classrooms find that children spend large portions of the day not engaged in learning activities (e.g., lining up, aimless wandering; Fuligni et al. 2012, Early et al. 2010), and this is likely true for all preschool classrooms. Skill-targeted curricula, by contrast with the whole-child curricula, lay out specific activities aimed at building up the targeted skills, while still allowing for child-directed activities.

In this paper, we evaluate the consequences of implemented whole-child versus skills-focused curricula for classroom environments and short-term achievement outcomes. Our analyses take advantage of a large-scale random-assignment evaluation of 14 preschool curricula that the U.S. Department of Education’s Institute for Education Sciences undertook beginning in 2003. We find that children gain more cognitive skills in early childhood

¹Correlational studies of early childhood education that resemble counterpart studies in K-12 find no significant associations between teacher qualifications and credentials and growth in child outcomes (see Early et al., 2007).

programs that provide supplemental academic instruction in mathematics and literacy content for a small portion of the day, compared with programs that take an exclusive whole-child approach. A math curriculum increased both classroom math activities and children's math achievement relative to the two whole-child curricula found in most Head Start and pre-K classrooms. Also relative to whole-child, literacy curricula increased literacy achievement despite producing no statistically significant gains in measured classroom quality. Whole-child curricula produced better classroom quality as measured by classroom observation than did locally-developed curricula, but failed to improve children's school readiness. This last point seems important for policy given that many states require use of the same classroom observation instrument (ECERS) to measure quality as was used in the study. To the extent our results generalize beyond the setting of our experiment, our results suggest that the curricular policies of Head Start and many state pre-K programs may be suboptimal, or at least deserve more study.

In the following section (Section I) we provide additional background information on curricula and preschool effectiveness, describe the data we use in Section II, present our analytic plan and results in Section III and IV. Section V includes tests for robustness, and we conclude in Section VI.

I. 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 (Warner 2007, Barnett 1995). 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 et al. 2015), with states spending an additional \$5.5 billion on programs like universal pre-K (Barnett et al. 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. 2012). Bitler, Hoynes, and Domina (2014) find that these small average effects after the first year of the experiment mask larger impacts at the bottom of the child skill distribution. Interestingly, Kline and Walters (Kline and Walters 2016), find that the counterfactual is important, with estimates suggesting that positive effects of the HS program are largest for those whose parents would otherwise have kept them at home and for those least likely to participate in the program. Gelber and Isen (2013) also find heterogeneous impacts. Evaluations of pre-K programs return generally positive impacts at the end of the pre-K year (Weiland and Yoshikawa 2013, Phillips, Dodge, and Pre-Kindergarten Task Force 2017, Wong et al. 2008).

Perhaps among the most useful strategies for boosting the consistency and effectiveness of early education programs is improving the curricula they use to organize instruction. As an important input into learning, 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 and Willer 2008). While social scientists have recently begun to consider the effects of curricula in other settings (Jackson and Makarin 2016, Koedel et al. 2017), there exists little or no evidence about which early childhood curricula are best for whom.

Published curricula and teaching materials differ across a number of dimensions – philosophies, materials, the role of the teacher, small or large group settings, classroom design, and the need for child assessment. In our analyses, we focus on three broad categories of early childhood curricular: Whole-child, content-specific, and locally-developed.

A. Whole-Child Curricula-the Most Common Business-as-Usual Curricula

Whole-child curricula emphasize “child-centered active learning,” cultivated by strategically arranging the classroom environment (Piaget 1976, DeVries and Kohlberg 1987). 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 (Coppie and 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 percent of the teachers responding to the national Head Start Family and Child Experiences Survey used Creative Curriculum; 19 used HighScope (Hulsey et al. 2011).

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, 1). Creative Curriculum also allows children a large proportion of free-choice time (Fuligni et al. 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 and Weikart 1981).

Despite the widespread adoption of these whole-child curricula in preschools, little evidence is available about the impacts of these curricula on children’s school readiness. Evidence from the 1960s Perry Preschool experiment suggests that HighScope boosts children’s early-

grade cognitive scores and reduces early adult outcomes like crime. But we lack methodologically strong, large-scale evaluations of recent versions of the curriculum as a stand-alone intervention. Since the children in the Perry study were extremely disadvantaged (Schweinhart & Weikart 1981), and the counterfactual in the Perry study was typically in-home care (Duncan and Magnuson 2013), the extent to which these results generalize to the present is unclear. Further, the only evaluation of Creative Curriculum that meets minimal standards of empirical rigor by the Institute for Education Sciences What Works Clearinghouse reveals that Creative Curriculum is no more effective than locally-developed curricula at improving children's oral language, print knowledge, phonological processing, or math skills (U.S. Department of Education 2013).

B. Content-Specific Curricula

Supporters of curricula that target specific academic or behavioral skills argue that preschool children benefit most from sequenced, explicit instruction, where instructional content is strategically focused on those skills. Content-specific curricula often supplement a classroom's regular curriculum (e.g., Creative Curriculum or a teacher or locally-developed curriculum) and provide instruction through developmentally-sound "free play" and exploration activities in small or large groups, or individually (Wasik and Hindman 2011). Random-assignment evaluations of content-specific curricula focusing on language, mathematics, and socioemotional skills often find positive impacts on their targeted sets of skills (Bierman, Nix, et al. 2008, Bierman, Domitrovich, et al. 2008, Clements and Sarama 2008, Fantuzzo, Gadsden, and McDermott 2011, Klein et al. 2008, Diamond et al. 2007, Morris et al. 2014). For example, children who received a curriculum targeting literacy showed improvements in their literacy and language skills (Justice et al. 2010, Lonigan et al. 2011). Clements & Sarama (2007; 2008) found large gains in math achievement relative to business-as-usual regular curricula from a targeted preschool mathematics curriculum. Such curricula range in cost, but effective packages like Clements & Sarama's *Building Blocks*, cost \$650 per classroom.

C. Locally-Developed Curricula-The Rest of Business-as-Usual

Many states allow early childhood education providers not otherwise subjected to curriculum requirements to develop their own lesson plans or curriculum rather than purchasing a published curriculum. Local districts or teachers design these themselves, but may incorporate components of various commercial curricula.

There are large negative gaps in achievement and behavior between low- and higher-income children at school entry. Because of these gaps, it is crucial for policy to be based on evaluations of whether children exposed to achievement-focused or locally-developed curricula systematically outperform children receiving the most commonly used preschool curricula – Creative Curriculum and HighScope – across cognitive and noncognitive domains of school readiness as well as the type of classroom observations that are increasingly mandated to measure preschool quality. Our article undertakes such a comparison.

II. 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 conduct independent evaluations of one or more curricula; all, however, used common measures of child outcomes, classroom processes, and implementation quality. The 14 curricula were evaluated at 18 different locations, and 2,911 children were included in the evaluations. Each of the grantees independently selected their early childhood education centers, randomly assigned whole classrooms to either treatment or control curricula and managed their own evaluation with assistance from Mathematica and RTI. 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 14 sets of grantee-specific estimates of the standardized outcome differences between the treatment curricula and the counterfactual control “business as usual” curricula. Our study is the first to pool data across grantees. Specifically, we pooled data across 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. Note that while for the first two 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.

Our inclusion criteria led us to drop four grantees and a total of 1,070 children from the study. Three of the four grantees were omitted because they evaluated a whole-child curriculum other than Creative Curriculum or HighScope (the Wisconsin, Missouri and three Success For All locations), while a fourth (New Hampshire) evaluated a literacy-enhanced version of Creative Curriculum with Creative Curriculum as the comparison condition. These sample deletions enable us to provide a focused evaluation of whole-child approaches that are most often found in large-scale preschool programs.

A. Randomization

We next describe the randomization implemented by the 11 grantees included in our curricula comparisons. Grantees are groups according to our 4 sets of curricula comparisons discussed below. Table 1 describes the grantee (column 1), geographic location of the classrooms (column 2), treatment (column 3) and control (column 4) curricula. Columns 5–7 describe the randomization. Columns 5 and 6 are mutually exclusive and describe whether all classrooms in the study within a given preschool were assigned to the same treatment status (Column 5 is yes, “Whole school randomized to same treatment”), or whether there was the potential for randomization of classroom within schools (Column 6 is yes, “Some within-school randomization to treatment”). Seven of the 11 grantee/curricula

comparisons used whole school randomization. Generally, for these comparisons, preschools were blocked based on characteristics of the neighboring elementary schools and the population they served, and then schools were randomly assigned within blocks. For the 4 remaining comparisons, there was at least some within-school randomization of classrooms. Importantly, a condition for participation in the experiment was that preschools and teachers had no say over which curricula they were assigned to. Column 7 reports whether classrooms were randomly assigned within schools. Column 8 reports the total number of schools in each aggregate comparison (school is the level at which we cluster the standard errors for the main results). Finally, Columns 9 to 12 report the number of schools (if relevant), classrooms, and children in the treatment and control groups. Columns 9 and 10 report the number of schools, classrooms, and children in the treatment and control conditions when randomization was at the school level, while Columns 11 and 12 report the number of treatment and control classrooms and children when there was within-school randomization.

B. 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 also described in Table 1. Eight curricula that targeted language/literacy but with diverse content and foci were included in our study.² Despite these differences and with a goal 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 only HighScope and Creative Curriculum, which we have already described.

Our final category, “locally-developed curricula,” included curricula that were 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

²One 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.

total number of academic activities.) Figure 1 summarizes the experimental contrasts and grantee-treatment curriculum comparisons included in our study, along with other study information.

1. 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 report (2008) and are reproduced in Table 2, as are grantee-based curricula impacts reported in the original IES funded evaluation. Table 2 shows that fidelity was typically medium (2 or medium on a 1 (not at all) to 3 (high) scale). Importantly there were relatively 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).³ 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 very large differences in fidelity.

C. Outcomes

1. Classroom Process Measures of Quality—One drawback to using cognitive test scores to assess the quality of instruction is that they provide 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, which can help us 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 et al. 2011). We use several classroom-level observational measures assessing the quality of the preschool classrooms that were included in the PCER study. These measures enable us to assess whether the approach used by the teacher differentially 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, correlations between measures, and additional information for each of the process quality measures we use are provided in Appendix Tables 1a and b.

The most widely known process quality instrument is the Early Childhood Environment Rating Scale – Revised (ECERS-R; Harms, Clifford, and 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. 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

³Some sites had a pilot year and we test for whether this affected outcomes, finding no significant differences.

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. We also note that this measure is increasingly being mandated to be collected by state pre-Ks (Barnett et al. 2017), so knowing more about how it correlates with learning is useful for policy.

The Teacher Behavior Rating Scale (TBRS; Landry et al. 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. 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.

The time between the fall (baseline) and spring assessments varies across classrooms and grantees. Thus, 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.⁴

2. Children's Achievement and Socioemotional Skills

Children's academic achievement and socioemotional skills were assessed using well-known nationally normed tests 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 as well as balance tests for all four curricula comparison groups in Tables 1 and 2. Observation counts are rounded to the nearest ten in accordance with NCES data policies.

⁴In 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 to 15 weeks before the end of the preschool year.

2a. Literacy Outcomes.—We draw upon three commonly utilized literacy outcomes. The Peabody Picture Vocabulary Test (PPVT; Dunn and 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. This test has been widely used, and is in the NLSY and Head Start Impact Study. The second and third literacy measures – Letter Word and Spelling – come from the Woodcock-Johnson III (WJ-III) Tests of Achievement (Woodcock, McGrew, and Mather 2001). The Letter Word subtest is similar to the PPVT in that it asks children to identify the letter or 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. 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 0 and a standard deviation of 1.⁵

2b. 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. The assessment takes approximately 10 minutes to administer. The second math assessment, the Child Math Assessment-Abbreviated (CMAA; Klein and Starkey 2002) is less well known, and was designed specifically for the PCER study (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 0 and a standard deviation of 1. 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 restandardized the sum.

2c. Socioemotional Outcomes.—Teachers rated children’s social skills and behavior problems using the Social Skills Rating System (SSRS; Gresham and 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). To form a social-skills composite score, we standardized (within the sample) both scales to have a mean of 0 and a standard deviation of 1, reverse coded the problem behaviors scale, averaged the two scores together and restandardized. The SSRS is a widely-used assessment, with good psychometric properties.

⁵We also analyze both these tests and the math tests discussed below separately, and these results are presented in Appendix Table 6. 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.

III. ANALYTIC APPROACH

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

$$O_{icsj} = \alpha + \beta_1 T_{cs} + \beta_2 Cov_{icj} + \mu_{js} + \varepsilon_{icsj}$$

where O_{icj} is the classroom or child outcomes observed for child i in classroom c in school s in grantee-treatment curricula comparison j ; T_{cs} is a dichotomous indicator of assignment to the treatment or control curriculum (this varies by classroom or school); Cov_{icj} are classroom, child, and family covariates for child i ; and ε_{icsj} is an error term. For each classroom⁶ 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 our estimate of β_1 for our four primary outcomes (ECERS-R, literacy skills, math skills, and socioemotional skills). All analyses use ordinary least squares with standard errors clustered at the school level (s). The regressions all include fixed effects - μ_{js} - for the unit within which random assignment is made (school or grantee by treatment curricula contrast, denoted by “ js ” in equation (1)).⁷ Including the fixed effects μ_{js} bases our estimates solely on random-assignment variation in our treatment/control contrasts.

We handled missing data in independent variables by imputing mean values for missing observations and used dummy variables to indicate the places of imputation. Because children were randomized after parental consent to participate was obtained, the PCER study had extremely low rates of missing data. Overall, missingness for child prescores ranged between 0–8%, being lowest for the cognitive tests. It was somewhat higher for parent characteristics from the parent survey, ranging between 9–25%. Importantly, rates of missingness did not differ by child treatment status for the covariates (see tests in Appendix Tables 2 and 3).

A. Baseline Controls

In the hopes of increasing the precision of our experimental impact estimates, we include a host of baseline covariates (Cov_{icj}) 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

⁶Note 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.

⁷Results are robust to alternative approaches to conducting inference, including other clustering schemes and various bootstrapping approaches; these results are discussed below.

composites, along with classroom measures as appropriate. (We test robustness to excluding these baseline measures as well.)

B. Samples

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.

IV. RESULTS

A. Balance

Given the experimental setting, we expect only trivial differences between the treatment and control groups across our four comparisons. 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 (Comparison I) or math vs. whole-child (III) comparisons. There were also no differences in the teacher characteristics or classroom observational measures for these comparisons.

Some mild baseline differences emerged in the classroom observational measures in the literacy vs. locally-developed comparison (II), and the locally-developed vs. Creative Curriculum experimental comparison (IV).⁸ A few baseline Xs were also significantly different individually in comparison III (gender, parent's education ($p < .05$); and household income ($p < .10$)), but the joint test of significance across baseline measures was insignificant and baseline cognitive tests were not significantly different from one another. We address these issues by controlling for classroom assessment scores at baseline and for child and family covariates.⁹ Also included in Appendix Tables 2 and 3 are indicators for the child not having a baseline cognitive or socioemotional test (Panel 5, Child Outcomes-Fall 2003).

This is rare for the cognitive tests, and even for the non-cognitive tests, is between 1 and 4%

⁸This 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.

⁹It 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.

for Comparisons II-IV, and 8% for the treatment group in comparison I and 11% for the control group. These differences are never statistically significant. The final outcomes (Panel 6, Child Outcomes-Spring 2004) are also very consistently reported, with little missingness.

B. Findings for Classroom Outcomes (Process Measures)

Table 3 shows impact estimates for the classroom outcomes, which are also displayed in Figures 2–5. All dependent variables were converted into standard deviation units 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.

1. Literacy Curricula vs. Creative/HighScope—As shown on the left-hand side of Figure 4, the ECERS classroom quality score was a marginally significant 0.25 standard deviations (sd) higher in classrooms with the Literacy Curricula ($p < .10$) compared with Creative/HighScope classrooms. Recall that the ECERS is an overall rating of the observed classroom quality that captures processes like teachers' interactions with children and the way a classroom is organized and maintained. There were no other statistically significant differences on the 3 remaining classroom observational measures.

2. Literacy Curricula vs. Locally-Developed Curricula—As when comparing literacy with the whole-child curricula, process measures look slightly higher compared to the locally-developed curricula. Classrooms using a literacy curriculum scored one-half of a sd higher on the ECERS-R ($p < .05$). Equally unsurprising but still informative, the targeted literacy curricula scored a marginally statistically 0.83 sd higher on the TBRS Literacy activities composite ($p < .10$) at the end of the preschool year than classrooms using a locally-developed curriculum.

3. Math Curricula vs. Creative/HighScope—Shown on the left-hand side of Figure 3, classrooms using the math curriculum scored more than one sd higher on the TBRS Math activities scale ($p < .05$) than control classrooms using Creative/HighScope at the end of the preschool year. There were no other significant differences.

4. Creative Curriculum vs. Locally-Developed Curricula—Unlike the previous comparisons, classrooms using Creative Curriculum had consistently higher ECERS-R, TBRS Math, TBRS Literacy, and Arnett scores (0.61 sd, 0.51 sd, 0.71 sd, 0.99 sd, respectively, all significant at the 5% level) at the end of the preschool year than classrooms using a locally-developed curriculum, as illustrated in Figure 2.

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. Classroom process impacts from using the skill-focused curricula were more varied. If better processes in the classroom translate into larger gains in skills and behavior, as if these better processes are indeed captured in our classroom measures, then we would expect positive effects on child outcomes for Creative Curriculum vs. business as usual curricula. The next section turns to these achievement and socioemotional results.

C. Findings for Child Cognitive and Socioemotional Outcomes

Table 4 shows 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 2–5. 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.

1. Literacy Curricula vs. Creative/HighScope: Literacy Curricula Raise Composite Literacy Scores—Children in classrooms randomized 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. This is a policy-relevant change in skills, matching the lower-bound estimate of early elementary achievement impacts from the Tennessee STAR class size reduction experiment (Nye, Hedges, and Konstantopoulos 2000). Appendix Table 5 shows that this marginally significant difference in literacy scores is driven in part by an increase in the WJ Spelling test of 0.18 sd (SE of 0.07, $p < .05$), 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 5 shows significant detrimental impacts of the literacy curricula on one of the two components of the social skills composite.

2. Literacy Curricula vs. Locally-Developed Curricula: Literacy Curricula Lead to Higher Math and Composite Scores—The literacy curricula generate larger impacts on achievement when compared with the locally-developed curricula. 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 increase of 0.18 sd in the CMAA math component ($p < .01$) and an increase in the WJ spelling literacy component of 0.16 sd ($p < .10$). The effect size for the literacy composite was similar (0.15 sd), but not statistically significant at conventional levels. While not overwhelmingly large, these are still important differences.

3. Math Curricula vs. Creative/HighScope: Math Curricula Raises Math and Academic Composite Scores—The differences between the targeted math curricula and the whole-child curricula are larger and more striking than those between the targeted literacy and whole-child curricula. 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. This difference is quite meaningful, matching those found by Angrist et al. in the evaluation of KIPP charter schools (2012), which would close one-third of the socioeconomic achievement gap in math skills present at school entry (Reardon and Portilla 2016). 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. Thus, importantly, while children gained

substantially in their early math achievement from being assigned to the targeted math curricula, this did not come at a cost to their literacy or social skills.

4. Creative Curriculum vs. 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 statistically insignificant differences between the school readiness skills of children exposed to Creative Curriculum and locally-developed curricula. Moreover, the point estimates for the differences are small and not economically meaningful. 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). By contrast, despite mixed differences across the whole-child and targeted math and literacy curricula in the process outcomes, children in the targeted math and literacy curricula had significantly higher scores in the skills targeted by the curricula, with the math vs. Creative/HighScope differences being quite large. The incongruity between impacts on classroom processes and impacts on child outcomes raises obvious questions about the ability of our widely-used process outcomes to identify classroom practices that best promote achievement. None of our curricular contrasts appears to affect noncognitive skills.

5. 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 4 remained statistically significant 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 et al. 2013). Still, evidence of longer-term impacts of early childhood educational experiences persist in spite of short-term fadeout (Chetty et al. forthcoming, Campbell et al. 2012).

V. ROBUSTNESS CHECKS

A. Cluster Robust Inference

One might be concerned that PCER contained too few schools to generate unbiased estimates of cluster-adjusted variance covariance matrices and that clustering would instead lead to over-rejection (e.g., Bertrand, Duflo, and Mullainathan 2004). As reported in Table 1, comparison I includes 72 schools, Comparison II, 41 schools, Comparison III, 36 Schools, and Comparison IV, 17 schools. Of these, all but comparison IV have enough clusters according to Cameron, Gelbach and Miller (2008). We have also explored block bootstrapping at the school level, using the wild-bootstrap at the school level, and even using

the wild bootstrap based on grantee by treatment curricula for the literacy versus whole-child comparison, which includes five such contrasts.¹⁰ The bootstrap and the various wild bootstrap inference approaches lead to conclusions about significance that are very similar to those presented above.

B. Pilot is Partial Treatment Year

In some study sites, our baseline scores are not true baselines, as there was a pilot year before our baseline year (one might worry the pilot year should be the first treated year). Unfortunately, we do not have data from the pilot years. A further concern is that the baseline scores are collected in early Fall, and one might worry that this means they reflect partial treatment. We also ran models that omitted the Fall 2003 baselines for both of these reasons. The coefficients were generally similar, and for several comparisons, larger than those presented in Table 3 (not shown). We also note that the baseline scores were largely balanced (Appendix Tables 2 and 3) suggesting any such early treatment effects were not substantial.

Additionally, we wanted to test for differences in effects between sites that participated in a pilot implementation year and those that did not. 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.

C. Did Pooling HighScope and Creative Curriculum Cause Misleading Estimates for HighScope?

One might worry (and the information about the curricula would suggest) that HighScope and Creative Curricula are quite different entities and that pooling them could be misleading. 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 Tables 3 and 4, with the exception of the ECERS-R scores, which increased from 0.25 sd to 0.34 sd and reached statistical significance. We also explored removing the High Scope controls from Comparison I and III, and found this also made no substantial difference.

D. Excluding the New York Control Group Makes No Important Difference

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

¹⁰This also adjusts for the fact that there may be more than one classroom within specific random assignment sites. We could not do the wild bootstrap in the other comparisons because there were not enough sites.

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.

E. Differential effects of teacher quality

One might be concerned that the differential quality of teachers in treatment and control classrooms may impact treatment effect estimates. Or one might simply want to see if the effects are larger where teachers are better. We estimated models where teacher's education (college degree or higher=1) and teacher's years of experience were separately interacted with treatment for both child and classroom outcomes. Results were mixed and largely null. For child outcomes, having a teacher with college degree or higher differentially benefitted the Creative Curriculum treatment group (Comparison IV) on their literacy outcome composites only, and no differential benefits of teacher's experience were found for any child outcome. For classroom process outcomes, teacher's education had a differential *negative* impact on math and literacy activities in the Creative Curriculum treatment comparison (IV), and a differential positive impact on math and literacy activities and the Arnett caregiver interaction scale for the literacy vs. locally-developed curricula (Comparison II). There were no differential benefits of teacher's years of experience for any of the classroom outcomes.

We also explored whether teachers with better process ratings (ECERS) or interactions with the children (Arnett) had better outcomes and found no important relationship between these classroom observation measures of quality and child outcomes.

These null findings aligns with the correlational literature in developmental psychology finding no associations between teacher's educational attainment and child outcomes (Early et al. 2007, Lin and Magnuson 2018, Burchinal et al. 2008). Taken together, we do not find strong evidence of teacher quality—however measured—moderating the impact of treatment curricula.

VI. DISCUSSION

We have shown—using randomized control trial data—that widely used whole-child curricula and locally-developed curricula appear to be inferior to targeted math and literacy curricula in producing achievement gains in math and literacy, respectively. By contrast, in the one case in which we can compare Creative Curricula to locally-developed curricula, Creative Curricula classrooms outperform comparison classrooms on a variety of classroom quality measures, but children in Creative classroom are no more ready for kindergarten at the end of the year than are children in comparison classrooms. Of course, it may be the case that our randomized control trial evidence, while strongly internally valid, might not be externally valid beyond these sites. Further, it is also true that we are comparing experimentally-assigned curricula to control curricula that have been able to be adjusted by

teachers to fit the local environment (this is always true when the counterfactual control condition is what exists in the real world). It is possible that the effects we have found might not be maintained were the schools to permanently adopt these new curricula. Nonetheless, our findings are a first step towards systematically assessing curricula.

Curricula developers may raise several additional issues that we cannot test with our data. One is that our study does not (and cannot with the data we have) address whether the fully and properly implemented whole-child curricula do as well as do the experimental targeted curricula. We respond by noting that complete implementation as defined by the developers is almost never attained in real-world settings, and ours is an analysis of one feasible policy alternative—replacing the current set of business-as-usual curricula (improperly implemented) with fully implemented targeted approaches.

Our preferred interpretation of our findings is that targeted math and literacy curricula are superior in our sample and setting to the dominant whole-child and locally-developed curricula in raising achievement, while at the same time not adversely affecting children's noncognitive skills. Critics may instead argue 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 may 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.¹¹

One valid concern with the Creative Curriculum/HighScope comparison groups is that the specific sites in the PCER study may not be representative of the way other programs use these curricula and thus that our study has limited external validity. 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 HSIS was an experimental evaluation of oversubscribed Head Start centers beginning in 2002, and represents the bulk of Head Start Centers in the country. The overall average ECERS-R scores in the PCER and HSIS samples were 4.21 and 5.22, respectively, and Arnett 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 (20th percentile of HSIS classrooms in quality).

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

¹¹Of course, like any experiment, it is possible that when implemented in real life at scale, the experimental targeted curricula would be carried out differently and not be as effective. Similarly, it is possible that with additional resources, the whole-child curricula in the field could be carried out more effectively. Yet, we point out that even when one of the two most prominent whole-child curricula, Creative Curriculum, is the experimental curricula, it is no more effective for cognitive outcomes than the locally-developed curricula that the schools otherwise would have used.

curriculum. Children in the HSIS had very similar scores to those of children in the PCER study, with no significant differences across the two groups.¹²

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. These coefficients and effect sizes are presented in Appendix Table 7, and are virtually identical to those presented in Table 4.

Conclusion

Given the large, persistent, and consequential gaps in literacy, numeracy, and socioemotional 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 noncognitive behaviors that support these skills. Federal, state, and local policy can and do 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 with a focus on specific school readiness 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 socioemotional or noncognitive skills. 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 call for more research to be done before mandating 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 developmentally-appropriate, proven skills-focused curricula and move away from the comparatively

¹²The 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).

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 superficial to provide useful measurement of children's experiences and interactions with teachers 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, broadly defined, will lead to larger gains in academic and social skills among young children. As with prior work, our study finds no consistency between curricular impacts on overall 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.

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 and 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 literacy 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 whole-child 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 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.

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.

Acknowledgements

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.

APPENDIX

“Boosting School Readiness: Should Preschool Teachers Target Skills or the Whole Child?”

Appendix Table 1a.

Description of classroom observational measures (process measures) and correlations of 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 (Hamm, 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 1b.

Correlation between ECERS-R and Arnett Caregiver Interaction Scale, overall and by curriculum Correlation

	Correlation
Overall	0.70
<i>Curriculum</i>	
Early Literacy and Learning Model	0.83
Literacy Express	0.75
DLM Early Childhood Express supplemented with Open Court Reading Pre-K	0.31
Ready Set Leap	0.80
Language Focused	0.87
Doors to Discovery	0.45
Let's Begin with the Letter People	0.53
Bright Beginnings	0.93
Pre-K Mathematics supplemented with DLM Early Childhood Express (Math Software only)	0.68
Creative Curriculum	0.73
HighScope	0.62

Appendix Table 2.

Literacy curricula and control group summary statistics (Comparisons I and II) 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	N	Mean	SD	P-value	N	Mean	SD	N	Mean	SD	N	Mean	SD	P-value
Covariates - Fall 2003																			
Child Gender - Female	530	0.47	330	0.44		330	0.46		0.31	270	0.46		170	0.46		170	0.46		0.98
Child Race - Black	530	0.56	330	0.52		330	0.08		0.90	270	0.08		170	0.08		170	0.08		0.76
Child Race - Asian	530	0.00	330	0.00		330	0.01		0.18	270	0.01		170	0.04		170	0.04		0.41
Child Race - Hispanic	530	0.11	330	0.10		330	0.27		0.42	270	0.27		170	0.24		170	0.24		0.54
Child Race - Other	530	0.04	330	0.03		330	0.06		0.65	270	0.06		170	0.05		170	0.05		0.86
Parent Education (years)	530	12.90	330	12.57	1.68	330	13.30	2.06	0.08	270	13.30		170	12.76	1.60	170	12.76		0.21
Parent Working	530	0.63	330	0.60		330	0.51		0.47	270	0.51		170	0.48		170	0.48		0.56
Parent Age (years)	530	30.96	330	30.76	6.68	330	32.82	6.41	0.75	270	32.82		170	32.39	6.84	170	32.39		0.78
Child Age (months)	530	54.70	330	54.84	3.97	330	54.82	3.73	0.41	270	54.82		170	54.79	3.76	170	54.79		0.35
Annual Household Income (thousands)	530	27.13	330	26.43	18.10	330	40.53	27.64	0.41	270	40.53		170	29.52	16.91	170	29.52		0.11
Receiving Welfare	530	0.12	330	0.18		330	0.06		0.08	270	0.06		170	0.09		170	0.09		0.86
Missing any baseline covariates	530	0.12	330	0.12		330	0.15		0.53	270	0.15		170	0.22		170	0.22		0.15
Classroom Observations - Fall 2003																			
CIS Arnett Total	530	3.22	330	3.41	0.33	330	3.33	0.44	0.05	270	3.33		170	3.04	0.55	170	3.04		0.06
ECERS Total	530	4.59	330	5.02	1.18	330	3.61	0.64	0.28	270	3.61		170	3.15	0.51	170	3.15		0.01
ECERS Provisions	530	4.55	330	5.15	1.26	330	3.44	0.69	0.12	270	3.44		170	2.96	0.43	170	2.96		0.02
ECERS Interaction	530	5.14	330	5.56	1.31	330	4.52	1.23	0.55	270	4.52		170	3.63	0.88	170	3.63		0.00
Teacher Characteristics																			
No college (T)	530	0.26	330	0.14		330	0.09			270	0.09		170	0.00		170	0.00		
Associates (T)	530	0.08	330	0.08		330	0.05		0.32	270	0.05		170	0.07		170	0.07		0.40
College + (T)	530	0.66	330	0.78		330	0.86			270	0.86		170	0.93		170	0.93		
Annual salary (T)	504	36262.22	310	39952.84	12682.25	310	42001.90	12700.16	0.52	270	42001.90		170	42981.45	8842.69	170	42981.45		0.77

	I. Literacy Curricula Compared With HighScope and Creative Curricula					II. Literacy Curricula Compared With Locally Developed Curricula								
	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD	P-value	Diff
Yrs. teaching experience	530	12.35	7.94	330	11.21	9.69	270	14.84	7.82	170	9.81	5.99	0.82	0.02
Classroom Observations - Spring 2004														
TBRS Math Quality	530	0.99	0.64	330	1.14	0.65	250	1.37	0.75	170	1.04	0.50	0.46	0.21
TBRS Math Quantity	530	1.11	0.49	330	1.19	0.56	250	1.36	0.55	170	1.11	0.40	0.69	0.23
TBRS Literacy Quality	530	1.58	0.44	330	1.58	0.41	250	1.77	0.52	170	1.34	0.35	0.90	0.02
TBRS Literacy Quantity	530	1.49	0.59	330	1.52	0.56	250	1.73	0.68	170	1.15	0.47	0.88	0.03
CIS Arnett Total	530	3.15	0.47	330	3.17	0.36	250	3.36	0.47	170	3.15	0.49	0.50	0.34
ECERS Total	530	4.36	1.14	330	4.32	0.98	270	4.01	0.80	170	3.57	0.77	0.12	0.10
ECERS Provisions	530	4.36	1.19	330	4.39	1.04	270	3.94	0.89	170	3.39	0.66	0.10	0.03
ECERS Interaction	530	4.88	1.45	330	4.94	1.22	270	5.00	1.27	170	4.42	1.31	0.52	0.19
Child Outcomes - Fall 2003														
PPVT	530	87.97	13.16	330	86.43	15.08	270	89.85	19.22	170	87.02	18.56	0.12	0.34
WJ Letter Word	530	100.85	15.87	330	99.39	15.41	270	98.40	15.61	170	93.56	15.63	0.15	0.39
WJ Spelling	530	94.95	13.94	330	94.91	14.62	270	91.88	12.76	170	90.03	13.12	0.38	0.65
WJ Applied Problems	530	92.50	13.51	330	92.10	13.53	270	95.31	16.67	170	93.83	16.52	0.19	0.81
CMAA Composite	530	0.42	0.25	330	0.43	0.23	270	0.42	0.26	170	0.38	0.25	0.66	0.78
Social Skills (teacher report)	490	100.44	16.10	300	100.78	15.27	260	99.02	15.21	170	101.19	21.13	0.92	0.15
Behavior Problems (teacher report)	520	100.25	13.51	300	100.78	12.71	270	99.10	12.89	170	102.65	15.45	0.91	0.28
Missing any academic prescore	530	0.00		330	0.00		270	0.00		170	0.00		-	-
Missing any socio-emotional prescore	530	0.08		330	0.11		270	0.03		170	0.02		0.66	0.77
Child Outcomes - Spring 2004														
PPVT	530	91.89	13.99	330	90.43	14.76	270	95.61	17.51	170	91.86	16.55	0.11	0.24
WJ Letter Word	530	105.31	13.48	320	103.47	13.63	270	105.84	14.41	170	100.95	14.28	0.01	0.28
WJ Spelling	500	97.01	14.66	300	93.83	14.91	270	98.06	12.89	170	93.32	13.30	0.00	0.17
WJ Applied Problems	520	93.73	13.41	320	92.00	13.63	270	99.85	13.72	170	97.28	16.35	0.04	0.50
CMAA Composite	530	0.58	0.24	330	0.60	0.22	270	0.66	0.23	170	0.57	0.26	0.84	0.31
Social Skills (teacher report)	530	103.94	15.65	320	107.55	15.58	270	107.00	14.68	170	110.74	13.80	0.08	0.20

	I. Literacy Curricula Compared With HighScope and Creative Curricula						II. Literacy Curricula Compared With Locally Developed Curricula							
	Literacy Curricula (Treat)			HighScope-Creative Curricula Comparison			Diff	Literacy Curricula (Treat)			Locally Developed Curricula			
	N	Mean	SD	N	Mean	SD		N	Mean	SD	N	Mean	SD	P-value
Behavior Problems (teacher report)	530	101.64	13.59	330	101.22	13.75	0.78	270	99.69	13.61	170	99.36	12.71	0.97
Literacy composite score	530	0.11	0.95	330	-0.07	1.00	0.00	270	0.26	1.06	170	-0.13	1.06	0.21
Math composite score	530	-0.06	0.96	330	-0.09	0.94	0.27	270	0.37	0.97	170	0.06	1.16	0.39
Academic composite score	530	0.01	0.95	330	-0.08	0.95	0.03	270	0.35	1.02	170	-0.02	1.16	0.30
Social skills composite score	530	-0.16	0.99	330	-0.01	1.03	0.24	270	0.04	1.00	170	0.19	0.91	0.44
Missing any academic outcome	530	0.06		330	0.11		0.29	270	0.01		170	0.01		0.61
Missing any socio-emotional outcome	530	0.00		330	0.02		0.25	270	0.00		170	0.00		0.33

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 Appendix Tables 1a & b. *p*-values account for clustering by random assignment site and date of classroom observational assessment (for classroom observation t-tests only). *N*s 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								
	N	Mean	SD	HighScope-Creative Curricula Comparison	Diff	N	Mean	SD	Locally Developed Curricula	Diff	N	Mean	SD	P-value
Covariates at Baseline (Fall 2003)														
Child Gender - Female	110	0.58		100	0.46		170	0.50	0.50	0.05	180	0.52	0.50	0.50
Child Race - Black	110	0.40		100	0.34		170	0.41	0.49	0.61	180	0.38	0.49	0.69
Child Race - Asian	110	0.06		100	0.01		170	0.01	0.08	0.33	180	0.00	0.00	0.33
Child Race - Hispanic	110	0.24		100	0.29		170	0.10	0.31	0.58	180	0.09	0.28	0.54
Child Race - Other	110	0.05		100	0.14		170	0.03	0.18	0.15	180	0.02	0.15	0.44
Parent Education (years)	110	13.08	1.78	100	12.46	1.85	170	12.54	1.54	0.05	180	12.70	1.45	0.30
Parent Working	110	0.57		100	0.43		170	0.45	0.50	0.13	180	0.48	0.50	0.58
Parent Age (years)	110	32.94	9.12	100	31.79	7.18	170	30.80	6.56	0.31	180	32.06	7.93	0.15
Child Age (months)	110	53.20	3.26	100	52.73	3.30	170	54.26	3.79	0.35	180	54.02	3.48	0.48
Annual Household Income (thousands)	110	29.51	17.51	100	24.54	13.60	170	22.27	14.47	0.06	180	25.14	16.06	0.30
Receiving Welfare	110	0.16		100	0.17		170	0.10	0.31	0.81	180	0.13	0.33	0.47
Missing any baseline covariates	108	0.10		100	0.11		174	0.20		0.92	180	0.18		0.56
Classroom Observations - Fall 2003														
CIS Arnett Total	100	3.15	0.37	100	3.18	0.59	170	3.20	0.66	0.66	180	2.88	0.66	0.19
ECERS Total	100	3.48	0.67	100	3.80	0.84	170	4.06	0.88	0.22	180	3.29	0.96	0.06
ECERS Provisions	100	3.50	0.62	100	3.83	0.87	170	4.14	0.99	0.25	180	3.21	0.94	0.04
ECERS Interaction	100	4.04	1.06	100	4.45	1.41	170	4.94	1.42	0.34	180	3.78	1.40	0.06
Teacher Characteristics														
No college (T)	110	0.19		100	0.08		170	0.11	0.31	0.71	180	0.28	0.45	0.10
Associates (T)	110	0.19		100	0.30		170	0.34	0.47	0.47	180	0.15	0.36	
College + (T)	110	0.62		100	0.62		170	0.55	0.50	0.50	180	0.56	0.50	

III. Math Curriculum Compared With HighScope and Creative Curricula												IV. Creative Curriculum Compared with Locally Developed Curricula											
Math Curriculum (Treat)						HighScope-Creative Curricula Comparison						Creative Curriculum (Treat)						Locally Developed Curricula					
N	Mean	SD	N	Mean	SD	N	Mean	SD	P-value	N	Mean	SD	N	Mean	SD	N	Mean	SD	P-value				
90	35831.42	10605.56	100	40965.06	18510.18	170	32743.28	13856.61	0.39	170	11.68	7.81	180	10.13	6.63	35052.85	11695.57	0.98					
110	17.46	10.92	100	17.92	7.56	170	11.68	7.81	0.88	170	11.68	7.81	180	10.13	6.63	35052.85	11695.57	0.57					
Classroom Observations - Spring 2004																							
100	1.21	0.94	100	0.72	0.46	160	1.43	0.80	0.22	160	1.43	0.80	170	0.91	0.46	35052.85	11695.57	0.02					
100	1.26	0.69	100	0.95	0.33	160	1.42	0.67	0.28	160	1.42	0.67	170	1.00	0.35	35052.85	11695.57	0.06					
100	1.12	0.33	100	1.13	0.39	160	1.53	0.36	0.96	160	1.53	0.36	170	1.18	0.26	35052.85	11695.57	0.04					
100	1.01	0.35	100	0.93	0.42	160	1.47	0.40	0.56	160	1.47	0.40	170	0.92	0.27	35052.85	11695.57	0.00					
100	3.06	0.63	100	2.93	0.60	160	3.35	0.31	0.79	160	3.35	0.31	180	2.93	0.54	35052.85	11695.57	0.11					
110	3.81	0.95	100	3.61	0.86	170	4.21	0.58	0.98	170	4.21	0.58	180	3.50	0.80	35052.85	11695.57	0.06					
110	3.67	1.03	100	3.53	0.75	170	4.19	0.72	0.90	170	4.19	0.72	180	3.35	0.77	35052.85	11695.57	0.03					
110	4.66	1.44	100	4.33	1.40	170	5.33	0.83	0.81	170	5.33	0.83	180	4.06	1.26	35052.85	11695.57	0.02					
Child Outcomes - Fall 2003																							
110	89.28	12.60	100	92.23	14.68	170	86.57	16.21	0.31	170	86.57	16.21	180	87.00	16.37	35052.85	11695.57	0.88					
110	102.86	17.40	100	101.98	13.86	170	94.69	15.39	0.81	170	94.69	15.39	180	92.90	16.49	35052.85	11695.57	0.52					
110	95.30	14.09	100	92.28	11.85	170	89.48	12.43	0.32	170	89.48	12.43	180	89.22	13.12	35052.85	11695.57	0.92					
110	99.78	12.84	100	96.68	13.70	170	92.67	13.55	0.21	170	92.67	13.55	180	90.93	16.41	35052.85	11695.57	0.27					
110	0.44	0.24	100	0.45	0.24	170	0.33	0.23	0.83	170	0.33	0.23	180	0.32	0.22	35052.85	11695.57	0.70					
110	106.06	13.46	100	107.70	14.52	170	99.61	16.97	0.62	170	99.61	16.97	180	99.75	20.64	35052.85	11695.57	0.85					
110	96.00	12.10	100	95.91	13.45	170	101.35	13.78	0.97	170	101.35	13.78	180	102.50	15.62	35052.85	11695.57	0.76					
108	0.00		100	0.00		170	0.00		-	170	0.00		180	0.00		35052.85	11695.57	-					
108	0.02		100	0.04		170	0.01		0.49	170	0.01		180	0.01		35052.85	11695.57	1.00					
Child Outcomes - Spring 2004																							
110	94.84	13.02	100	93.41	15.28	170	92.19	15.08	0.66	170	92.19	15.08	180	90.02	15.12	35052.85	11695.57	0.05					
110	101.46	14.15	100	101.31	14.33	170	99.75	11.32	0.96	170	99.75	11.32	180	99.03	13.23	35052.85	11695.57	0.66					
110	95.90	13.27	100	93.11	11.84	170	91.45	13.08	0.29	170	91.45	13.08	180	91.26	12.99	35052.85	11695.57	0.84					

III. Math Curriculum Compared With HighScope and Creative Curricula										IV. Creative Curriculum Compared with Locally Developed Curricula				
	Math Curriculum (Treat)			HighScope-Creative Curricula Comparison			Diff	Creative Curriculum (Treat)			Diff			
	N	Mean	SD	N	Mean	SD		N	Mean	SD		N	Mean	SD
WJ Applied Problems	110	98.81	13.43	100	94.48	12.96	0.15	170	94.98	12.73	180	93.30	15.85	0.20
CMAA Composite	110	0.66	0.21	100	0.55	0.21	0.01	170	0.49	0.26	180	0.49	0.28	0.85
Social Skills (teacher report)	110	113.37	12.08	100	108.93	14.84	0.23	170	108.36	15.42	180	108.70	13.45	0.99
Behavior Problems (teacher report)	110	96.15	12.77	100	98.98	13.92	0.40	170	99.08	13.24	180	99.38	12.90	0.90
Literacy composite score	110	0.05	0.96	100	-0.08	0.95	0.55	170	-0.22	0.94	180	-0.31	0.98	0.38
Math composite score	110	0.34	0.89	100	-0.10	0.86	0.02	170	-0.22	1.01	180	-0.29	1.18	0.48
Academic composite score	110	0.24	0.91	100	-0.10	0.88	0.09	170	-0.24	0.96	180	-0.32	1.11	0.44
Social skills composite score	110	0.41	0.91	100	0.13	1.03	0.29	170	0.11	1.01	180	0.11	0.89	0.94
Missing any academic outcome	108	0.00		100	0.02		0.29	170	0.02		176	0.03		0.43
Missing any socio-emotional outcome	108	0.00		100	0.00		0.25	170	0.01		176	0.00		0.28

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 Appendix Tables 1a & b. *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	0.24 (0.15)	0.21 (0.16)	0.13 (0.15)	-0.16 (0.19)	-0.12 (0.21)	0.08 (0.19)	0.04 (0.20)	0.16 (0.17)
<i>Classroom N=100</i>	860	860	860	840	840	840	840	850
II. Literacy vs. Locally-Developed Curricula	0.56 (0.27)	0.59 (0.30)	0.48 (0.27)	0.56 (0.39)	0.37 (0.35)	0.74 (0.45)	0.82 (0.43)	0.28 (0.25)
<i>Classroom N=60</i>	430	430	430	410	410	410	410	410
III. Math vs. HighScope and Creative Curriculum	0.21 (0.32)	0.23 (0.26)	0.33 (0.40)	1.28 (0.52)	1.10 (0.54)	0.43 (0.27)	0.34 (0.35)	0.67 (0.52)
<i>Classroom N=30</i>	200	200	200	200	200	200	200	200
IV. Creative vs. Locally-Developed Curricula	0.62 (0.32)	0.45 (0.28)	0.84 (0.38)	0.50 (0.21)	0.50 (0.31)	0.74 (0.22)	0.67 (0.27)	1.00 (0.64)
<i>Classroom N=340</i>	340	340	320	320	320	320	330	

Note: Each entry represents results from a separate regression. Standard errors clustered at the school-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 Appendix Table 1. 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.

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
I. Literacy vs. HighScope and Creative Curriculum							
	0.05 (0.04)	0.07 (0.06)	0.17 (0.06)	0.07 (0.06)	-0.10 (0.07)	-0.25 (0.09)	-0.00 (0.08)
N	860	850	800	840	850	850	860
II. Literacy vs. Locally-Developed Curricula							
	0.09 (0.08)	0.10 (0.13)	0.14 (0.09)	0.05 (0.10)	0.20 (0.09)	-0.26 (0.20)	0.17 (0.19)
N	440	440	440	440	440	440	440
III. Math vs. HighScope and Creative Curriculum							
	0.16 (0.10)	-0.11 (0.12)	0.08 (0.11)	0.27 (0.14)	0.35 (0.11)	0.29 (0.20)	-0.15 (0.17)
N	220	220	220	220	220	210	210
IV. Creative Curriculum vs. Locally-Developed Curricula							
	0.12 (0.07)	-0.01 (0.07)	-0.04 (0.09)	0.09 (0.07)	-0.08 (0.15)	-0.03 (0.21)	0.01 (0.25)
N	360	360	360	360	360	350	350

Note. Each entry represents results from a separate regression. Standard errors clustered at the school-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.

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 (0.10)	0.35 (0.11)	0.25 (0.11)	0.21 (0.27)
N	220	220	220	210
NY Math treatment group included, all NY control classrooms excluded	0.11 (0.13)	0.35 (0.18)	0.27 (0.16)	-0.04 (0.37)
N	210	210	210	200
Only CA math site	0.06 (0.13)	0.30 (0.17)	0.23 (0.16)	-0.01 (0.31)
N	150	150	150	150

Note. Each entry represents results from a separate regression. Standard errors clustered at the school 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.

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 (0.04)	0.10 (0.05)	0.18 (0.06)	0.09 (0.06)	-0.09 (0.06)	-0.25 (0.10)	-0.001 (0.10)
N	890	880	830	870	890	850	860
II. Literacy vs. Locally-Developed Curricula							
	0.06 (0.08)	0.14 (0.11)	0.16 (0.07)	0.06 (0.08)	0.18 (0.07)	-0.27 (0.19)	0.18 (0.19)
N	480	480	480	480	480	450	450
III. Math vs. HighScope and Creative Curriculum							
	0.16 (0.09)	-0.09 (0.10)	0.07 (0.10)	0.27 (0.12)	0.35 (0.10)	0.29 (0.18)	-0.15 (0.16)
N	220	220	220	220	220	210	210
IV. Creative Curriculum vs. Locally-Developed Curricula							
	0.12 (0.07)	-0.04 (0.06)	-0.05 (0.09)	0.10 (0.06)	-0.07 (0.09)	-0.05 (0.21)	0.05 (0.19)
N	360	360	360	360	360	350	350

Note. Each entry represents results from a separate regression. Standard errors clustered at the school 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.

References

- Advisory Committee on Head Start Research and Evaluation 2012 Final Report. Washington, DC: U.S. Department of Health and Human Services.
- Anderson Michael 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.
- Angrist Joshua D. , Dynarski Susan M. , Kane Thomas J. , Pathak Parag A. , and Walters Christopher R. . 2012 “Who Benefits from KIPP?” *Journal of Policy Analysis and Management* 31 (4):837–860. doi: 10.1002/pam.21647.
- Arnett J 1989 Caregiver Interaction Scale. Princeton, NJ: Educational Testing Service.
- Barnett William S. 1995 “Long-term effects of early childhood programs on cognitive and school outcomes.” *Future of Children* 5:25–50.
- Barnett William S. , Carolan ME , Squires JH , Clarke Brown Kirsty , and Horowitz Michelle . 2015 *The State of Preschool 2014* edited by Rutgers School of Education National Institute for Early Education Research. New Brunswick, NJ.
- Barnett William S. , Friedman-Krauss Allison H. , Weisenfeld GG , Horowitz Michelle , Kasmin Richard , and Squires JH . 2017 *The State of Preschool 2016* edited by Rutgers School of Education National Institute for Early Education Research. New Brunswick, NJ.
- Belfield Clive R , Nores Milagros , Barnett William S. , and Schweinhart Lawrence J. . 2006 “The High/Scope Perry Preschool Program.” *Journal of Human Resources* XLI (1):162–190. doi: 10.3368/jhr.XLI.1.162.
- Bertrand Marianne , Duflo Esther , and Mullainathan Sendhil . 2004 “How Much Should We Trust Differences-In-Differences Estimates?” *The Quarterly Journal of Economics* 119 (1):249–275. doi: 10.1162/003355304772839588.
- Bierman Karen L. , Domitrovich Celene E. , Nix Robert L. , Gest Scott D. , Welsh Janet A. , Greenberg Mark T. , Blair Clancy , Nelson Keith E. , and Gill Sukhdeep . 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. [PubMed: 19037951]
- Bierman Karen L. , Nix Robert L. , Greenberg Mark T. , Blair Clancy , and Domitrovich Celene 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: 10.1017/S0954579408000394. [PubMed: 18606033]
- Bitler Marianne P. , Hoynes Hilary W. , and Domina Thurston . 2014 “Experimental evidence on distributional effects of Head Start.” National Bureau of Economic Research Working Paper *Series* No. 20434. doi: 10.3386/w20434.
- Burchinal MR , Howes Carollee , Pianta Robert , Bryant Donna , Early Diane M. , Clifford Richard M. , and Barbarin Oscar A. . 2008 “Predicting Child Outcomes at the End of Kindergarten from the Quality of Pre-Kindergarten Teacher-Child Interactions and Instruction.” *Applied Developmental Science* 12 (3):140–153. doi: 10.1080/10888690802199418.
- Cameron Colin , Gelbach Jonah B. , and Miller Douglas L. . 2008 “Bootstrap-based improvements for inference with clustered errors.” *Review of Economics and Statistics* 90 (3):414–427.
- Campbell Frances A. , Conti Gabriella , Heckman James J. , Seong Hyeok Moon Rodrigo Pinto , Pungello Elizabeth , and Pan Yi . 2014 “Early childhood investments substantially boost adult health.” *Science* 343 (6178):1478–1485. doi: 10.1126/science.1248429. [PubMed: 24675955]
- Campbell Frances A. , Pungello Elizabeth , Burchinal Margaret R. , Kainz Kirsten , Pan Yi , Wasik Barbara H. , Barbarin Oscar A. , Sparling Joseph J. , and Ramey CT . 2012 “Adult outcomes as a function of an early childhood educational program: an Abecedarian Project follow-up.” *Developmental Psychology* 48 (4):1033. [PubMed: 22250997]
- Campbell Frances A. , Wasik Barbara H. , Pungello Elizabeth , Burchinal MR , Barbarin Oscar , Kainz Kirsten , Sparling Joseph J. , and Ramey CT . 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.

- Chetty Raj , Friedman John N. , Hilger Nathaniel , Saez Emmanuel , Schanzenbach Diane Whitmore , and Yagan Danny . 2011 “How does your kindergarten classroom affect your earnings? Evidence from Project Star.” *The Quarterly Journal of Economics* 126 (4):1593–1660. doi: 10.1093/qje/qjr041. [PubMed: 22256342]
- Chetty Raj , Friedman John N. , Hilger Nathaniel , Saez Emmanuel , Schanzenbach Diane Whitmore , and Yagan Danny forthcoming. “How Does Your Kindergarten Classroom Affect Your Earnings? Evidence From Project STAR.” *Quarterly Journal of Economics*. doi: 10.3386/w16381.
- Clements Douglas H. , and Sarama Julie . 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 Douglas H. , and Sarama Julie . 2011 “Early Childhood Mathematics Intervention.” *Science* 333 (6045):968–970. doi: 10.1126/science.1204537. [PubMed: 21852488]
- Clements Douglas H. , Sarama Julie , Wolfe Christopher B. , and Mary Elaine Spitler. 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.
- Conti Gabriella , Heckman James J. , and Pinto Rodrigo . 2015 “The Effects of Two Influential Early Childhood Interventions on Health and Healthy Behaviors.” *National Bureau of Economic Research Working Paper Series* No. 21454. doi: 10.3386/w21454.
- Copple C , and 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 David . 2009 “Early childhood entervention and life-cycle skill development: Evidence from Head Start.” *American Economic Journal: Applied Economics* 1 (3):111–134. doi: 10.2307/25760174.
- DeVries Rheta , and Kohlberg Lawrence . 1987 *Programs of early education: The constructivist view*. White Plains, NY: Longman.
- Diamond Adele , Barnett William S. , Thomas Jessica , and Munro Sarah . 2007 “Preschool program improves cognitive control.” *Science* 318 (5855):1387–1388. doi: 10.1126/science.1151148. [PubMed: 18048670]
- Duncan Greg J. , and Magnuson Katherine . 2013 “Investing in preschool programs.” *The Journal of Economic Perspectives* 27 (2):109–132. doi: 10.1257/jep.27.2.109. [PubMed: 25663745]
- Dunn LM , and Dunn LM . 1997 *Peabody Picture Vocabulary Test--Third Edition (PPVT-III)*. Upper Saddle River, NJ: Pearson Publishing.
- Early Diane M. , Iruka Iheoma U. , Ritchie Sharon , Barbarin Oscar A. , Winn Donna-Marie C. , Crawford Gisele M. , Frome Pamela M. , Clifford Richard M. , Burchinal Margaret R. , Howes Carollee , Bryant Donna M. , and Pianta Robert . 2010 “How do pre-kindergarteners spend their time? Gender, ethnicity, and income as predictors of experiences in pre-kindergarten classrooms.” *Early Childhood Research Quarterly* 25 (2):177–193. doi: 10.1016/j.ecresq.2009.10.003.
- Early Diane M. , Maxwell Kelly L. , Burchinal MR , Alva Soumya , Bender Randall H. , Bryant Donna , Cai Karen , Clifford Richard M. , Ebanks Caroline , Griffin James A. , Henry Gary T. , Howes Carollee , Iriondo-Perez Jeniffer , Jeon Hyun-Joo , Mashburn Andrew J. , Peisner-Feinberg Ellen , Pianta Robert , Vandergrift Nathan , and Zill Nicholas . 2007 “Teachers’ Education, Classroom Quality, and Young Children’s Academic Skills: Results From Seven Studies of Preschool Programs.” *Child Development* 78 (2):558–580. [PubMed: 17381790]
- Fantuzzo John W. , Gadsden Vivian L. , and Paul A. McDermott. 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.
- Fulgini Allison Sidle , Howes Carollee , Huang Yiching , Hong Sandra Soliday , and Lara-Cinisomo Sandraluz . 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. [PubMed: 22665945]

- Gelber Alexander , and Isen Adam . 2013 “Children’s schooling and parents’ behavior: Evidence from the Head Start Impact Study.” *Journal of Public Economics* 101 (0):25–38. doi: 10.1016/j.jpubeco.2013.02.005.
- Gormley William T. 2007 “Early childhood care and education: Lessons and puzzles.” *Journal of Policy Analysis and Management* 26 (3):633–671.
- Gresham Frank M , and Elliott Stephen N . 1990 *Social skills rating system (SSRS)*. Circle Pines, MN: American Guidance Service.
- Harms Thelma , Clifford Richard M. , and Cryer Debby . 1998 *Early Childhood Environment Rating Scale*. New York: Teachers College Press.
- Hart Craig H , Burts Diane C , and Charlesworth Rosalind . 1997 *Integrated curriculum and developmentally appropriate practice: Birth to age eight*: Suny Press.
- Heckman James J , Hyeok Moon Seong , Pinto Rodrigo , Savelyev Peter A , and Yavitz Adam . 2010 “The rate of return to the HighScope Perry Preschool Program.” *Journal of Public Economics* 94 (1):114–128. [PubMed: 21804653]
- Hulsey LK , Aikens Nikki , Kopack Ashley , West J , Moiduddin Emily , and Tarullo Louisa Banks . 2011 *Head Start children, families, and programs: Present and past data from FACES*. Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.
- Isaacs J , Edelstein S , Hahn H , Steele Ellen , and Steuerle CE 2015.
- Jackson C. Kirabo , and Makarin Alexey . 2016 “Simplifying Teaching: A Field Experiment with Online “Off-the-Shelf” Lessons.” *National Bureau of Economic Research Working Paper Series* No. 22398. doi: 10.3386/w22398.
- Jackson C. Kirabo , Rockoff Jonah E. , and Staiger Douglas O. . 2014 “Teacher effects and teacher-related policies.” *Annual Review of Economics* 6 (1):801–825.
- Justice LM , McGinty AS , Cabell SQ , Kilday CR , Knighton K , and 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 Thomas J , Taylor Eric S , Tyler John H , and Wooten Amy L . 2011 “Identifying effective classroom practices using student achievement data.” *Journal of Human Resources* 46 (3):587–613.
- Klein Alice , and Starkey Prentice . 2002 “Child Math Assessment-Abbreviated.”
- Klein Alice , Starkey Prentice , Clements Douglas H. , Sarama Julie , and Iyer Roopa . 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 Patrick , and Walters Christopher . 2016 “Evaluating public programs with close substitutes: The case of Head Start.” *Quarterly Journal of Economics* 131 (4):1795–1848.
- Koedel Cory , Li Diyi , Polikoff Morgan S. , Hardaway Tenice , and Wrabel Stephani L. . 2017 “Mathematics Curriculum Effects on Student Achievement in California.” *AERA Open* 3 (1): 2332858417690511. doi: 10.1177/2332858417690511.
- Landry Susan H , Crawford April , Gunnewig Susan B , and Swank Paul R . 2002 “Teacher Behavior Rating Scale (TBRS).” Center for Improving the Readiness of Children for Learning and Education, unpublished research instrument.
- Lin Ying-Chun , and Magnuson Katherine A. . 2018. “Classroom quality and children’s academic skills in child care centers: Understanding the role of teacher qualifications.” *Early Childhood Research Quarterly* 42 (Supplement C):215–227. doi: 10.1016/j.ecresq.2017.10.003.
- Lonigan Christopher J. , Farver JoAnn M. , Phillips Beth M. , and Clancy-Menchetti Jeanine . 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 PA , Mattera Shira K. , Castells Nina , Bangser Michael , Bierman Karen L. , and Raver CC . 2014 *Impact findings from the Head Start CARES Demonstration: National evaluation of three approaches to improving preschoolers’ social and emotional competence* edited by Office of Planning Research and Evaluation; Administration for Children and Families. Washington, DC.

- Nye Barbara , Hedges Larry V. , and Konstantopoulos Spyros . 2000 “The effects of small classes on academic achievement: The results of the Tennessee class size experiment.” *American Educational Research Journal* 37 (1):123–151. doi: 10.3102/00028312037001123.
- Office of Head Start. 2010 “Head Start Program Fact Sheet Fiscal Year 2010.” Administration for Children and Families, Dept. Health and Human Services, accessed January 2016 <http://eclkc.ohs.acf.hhs.gov/hslc/data/factsheets/fHeadStartProgr.htm>.
- Phillips Deborah , Dodge KA , and Pre-Kindergarten Task Force, eds. 2017 *The Current State of Scientific Knowledge on Pre-Kindergarten Effects*. Washington, D.C.: Brookings Institution and Duke University.
- Piaget Jean . 1976 *Piaget’s theory*. New York, NY: Springer.
- Pianta Robert , Howes Carollee , Burchinal MR , Bryant Donna , Clifford Richard , Early Diane M. , and Barbarin Oscar 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* edited by Institute for Education Sciences; U.S. Department of Education. Washington, DC.
- Puma Michael , Bell Stephen , Cook Ronna , Heid Camilla , Broene Pam , Jenkins Frank , Mashburn Andrew J. , and Downer Jason T. . 2012 Washington, DC.
- Reardon Sean F. , and Portilla Ximena A. . 2016 “Recent Trends in Socioeconomic and Racial School Readiness Gaps.” *AERA Open* 2 (3):1–18. [PubMed: 26942210]
- Ritchie Sharon , and Willer Barbara . 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).
- Sabol Terri J. , Soliday Hong Sandra L. , Pianta Robert , and Burchinal MR . 2013 “Can Rating Pre-K Programs Predict Children’s Learning?” *Science* 341 (6148):845–846. [PubMed: 23970684]
- Schweinhart Lawrence J. 2005 *Lifetime effects: The High/Scope Perry Preschool Study through age 40* Edited by High/Scope Educational Research Foundation. Ypsilanti, M.I.: High/Scope Press.
- Schweinhart Lawrence J. , and Weikart David 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.
- Department of Education US . 2013 “Early childhood education intervention report: The Creative Curriculum for Preschool, Fourth Edition.” Institute of Education Sciences, What Works Clearinghouse.
- Warner Mildred . 2007 “Child care and economic development: Markets, households and public policy.” *International Journal of Economic Development* 9 (3):111–121.
- Wasik Barbara A. , and Hindman Annemarie 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 David P. , and Schweinhart Lawrence . 1987 “The High/Scope cognitively oriented curriculum in early education” In *Approaches to early childhood education*, edited by Roopnarine JL and Johnson JE , 253–268. New York: Merrill/Macmillan.
- Weiland Christina , and Yoshikawa Hirokazu . 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. [PubMed: 23534487]
- Wong Vivian C. , Cook Thomas D. , Barnett William S. , and Jung Kwanghee . 2008 “An effectiveness-based evaluation of five state pre-kindergarten programs.” *Journal of Policy Analysis and Management* 27 (1):122–154.
- Woodcock RW , McGrew KS , and Mather N . 2001 *Woodcock-Johnson-III Tests of Achievement*. Itasca, IL: Riverside.

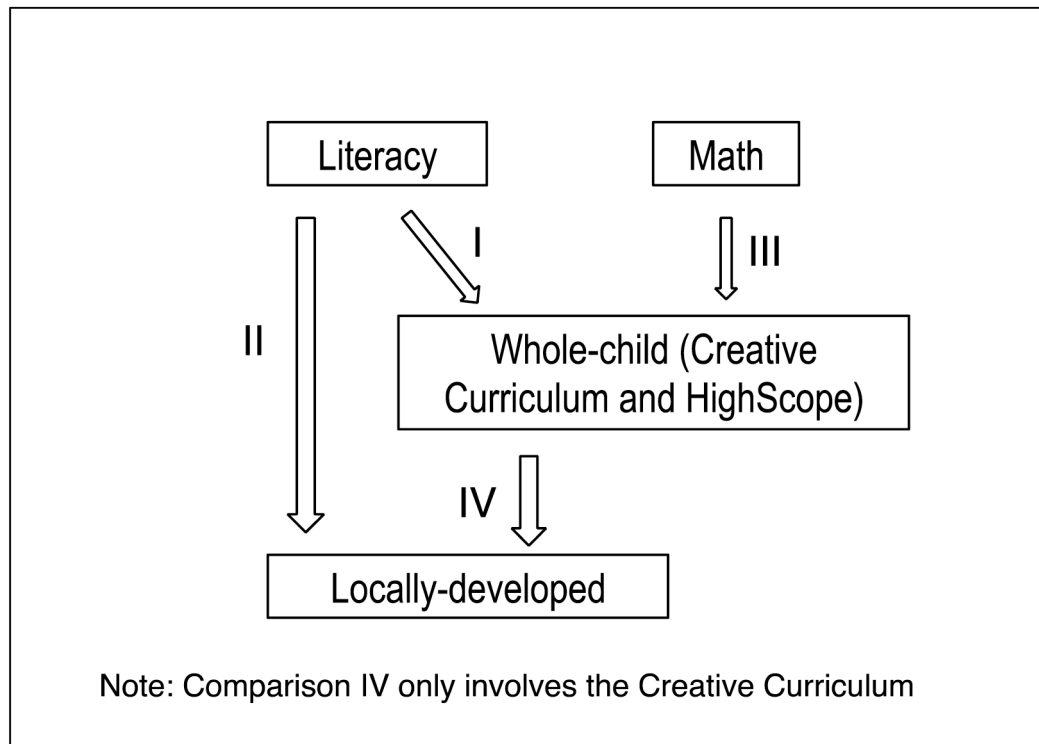


Figure 1.
:Curricula comparisons in study sample *Notes:* All curricula comparisons are within-site comparisons of randomly assigned treatment-control conditions. Curricula and site-specific information are available in Table 1.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

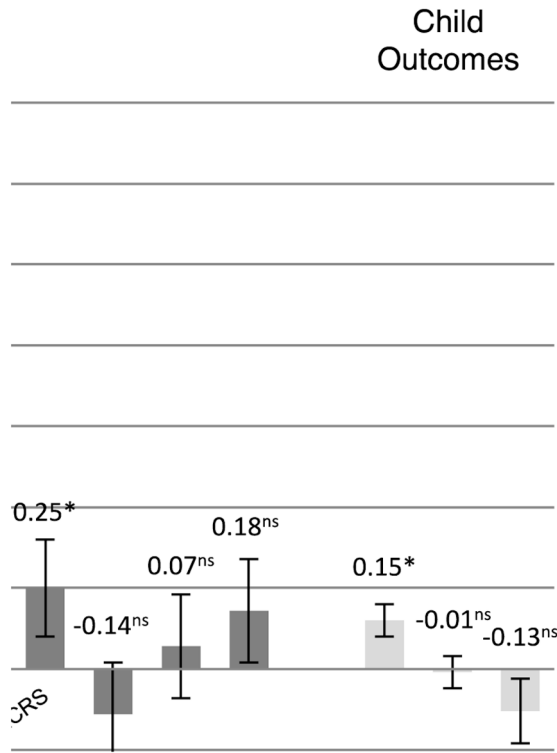


Figure 2. Classroom and child Impacts of comparison I: *Literacy vs Creative and HighScope Curricula*

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

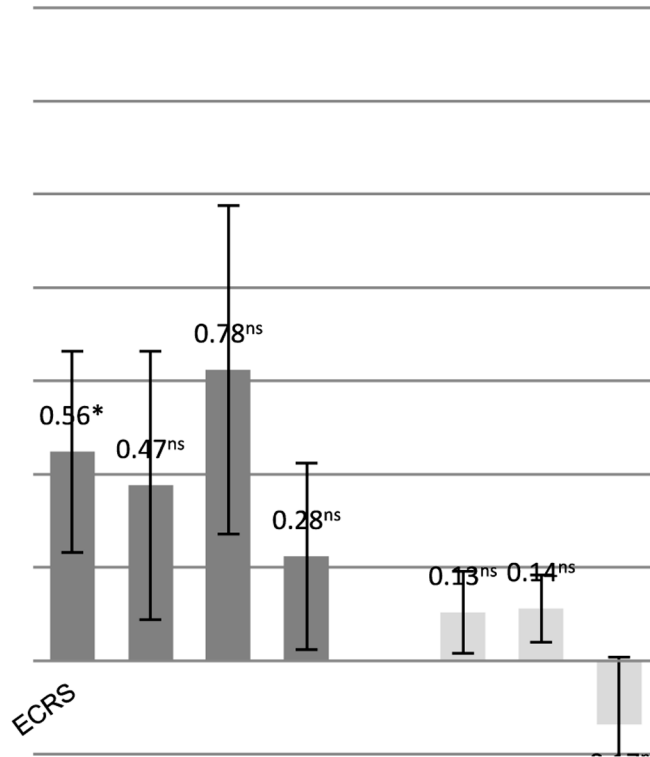


Figure 3. Classroom and child Impacts of comparison II: Literacy vs Locally-Developed Curricula

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

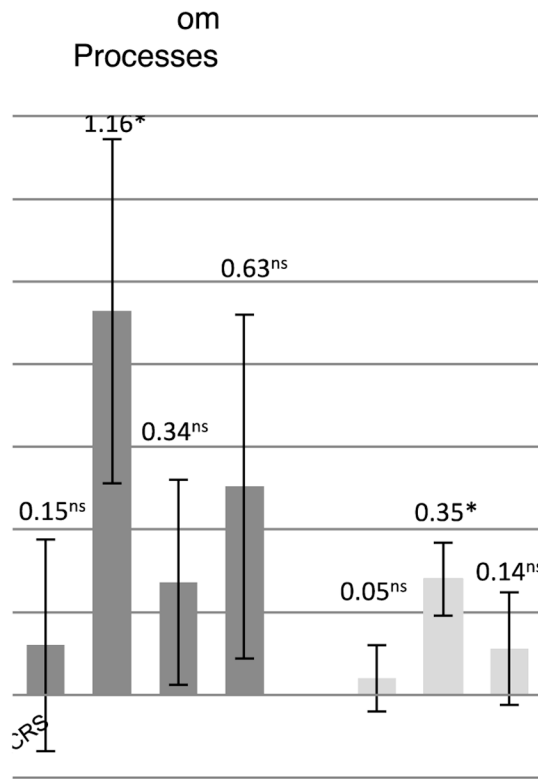


Figure 4. classroom and child Impacts of comparison III: *Math vs. creative and HighScope Curricula*
Notes: Bars show estimated impacts of various curricula comparisons on classroom process quality and child outcomes as measured by composite standardized scores of literacy skills, math skills and socioemotional skills. Each figure is from one of the curricula comparisons described in Figure 1, and each bar is from a separate regression. Standard error bars are shown for each estimate. * $p < .05$.

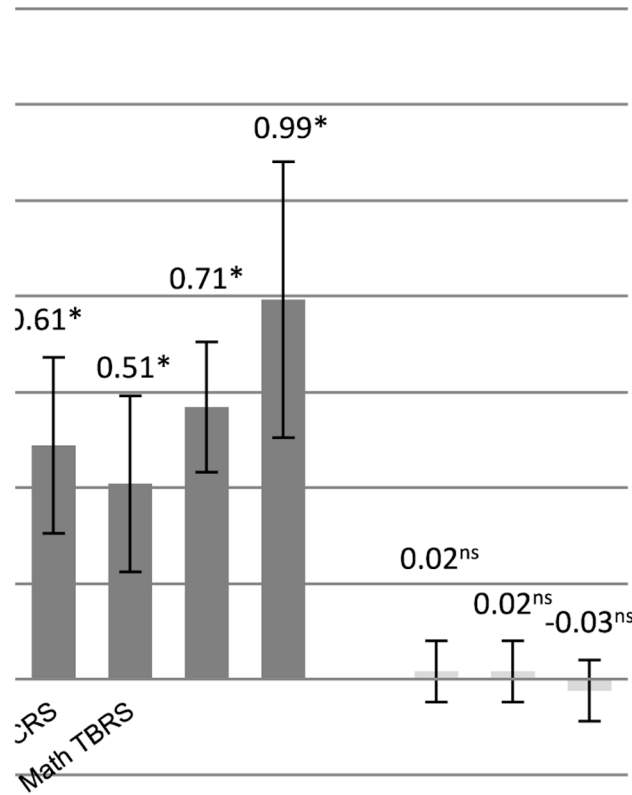


Figure 5. classroom and child Impacts of comparison IV: *Creative Curriculum vs. Locally-Developed*
Notes: Bars show estimated impacts of various curricula comparisons on classroom process quality and child outcomes as measured by composite standardized scores of literacy skills, math skills and socioemotional skills. Each figure is from one of the curricula comparisons described in Figure 1, and each bar is from a separate regression. Standard error bars are shown for each estimate. * $p < .05$.

Table 1:

Description of experimental design

1	2	3	4	5	6	7	8	9	10
Grantee and sample size	Site	Treatment Curriculum	Control Curriculum (a)	Whole school randomized to same Treatment?	Some within-school randomization to treatment?	Children randomized across treatment conditions within school?	Total number of schools (clusters)	Number of schools, classrooms, and students, if school randomization (T/C)	Number of classrooms, and students, if within-school randomization (T/C)
I. Literacy vs. HighScope and Creative Curriculum									
University of North Florida School n= 27; Classroom n= 27; Child n=210	FL	Early Literacy and Learning Model	Creative Curriculum	Y (same elementary school shared same treatment status)	N	N	72	14, 14, 120 13, 13, 90	
Florida State University School n=11; Classroom n=19; Child n=200	FL	Literacy Express	HighScope	Y	N	N	5, 10, 100 6, 9, 100		
Florida State University School n=12; Classroom n=20; Child n=200	FL	DLM Early Childhood Express supplemented with Open Court Reading Pre-K	HighScope	Y	N	N	6, 11, 100 6, 9, 100		
University of	NJ	Ready Set Leap	HighScope	N	Y	N		21, 120 18, 140	

1	2	3	4	5	6	7	8	9	10
Grantee and sample size	Site	Treatment Curriculum	Control Curriculum (a)	Whole school randomized to same Treatment?	Some within-school randomization to treatment?	Children randomized across treatment conditions within school?	Total number of schools (clusters)	Number of schools, classrooms, students, if school randomization (T/C)	Number of classrooms and students, if within-school randomization (T/C)
California-Berkeley School n=17; Classroom n=39; Child n=260									
University of Virginia School n=5; Classroom n=14; Child n=200	VA	Language Focused	HighScope	N	Y	N		7, 100 7, 100	
II. Literacy vs. Locally-Developed Curriculum									
University of Texas Health Science Center at Houston School n=12;	TX	Doors to Discovery	Locally Developed	Y	N	N	41	8, 14, 100 4, 14, 90	
Classroom n=28; Child n=190									
University of Texas Health Science Center at Houston	TX	Let's Begin with the Letter People	Locally Developed	Y	N	N	7, 15, 100 4, 15, 90		

1	2	3	4	5	6	7	8	9	10
Grantee and sample size	Site	Treatment Curriculum	Control Curriculum (a)	Whole school randomized to same Treatment?	Some within-school randomization to treatment?	Children randomized across treatment conditions within school?	Total number of schools (clusters)	Number of schools, classrooms, students, if school randomization (T/C)	Number of classrooms and students, if within-school randomization (T/C)
School n=10;									
Classroom n=30; Child n=200									
Vanderbilt University School n=13;	TN	Bright Beginnings	Locally Developed	Y	N	N	7, 7, 100 6, 7, 100		
Classroom n=14; Child n=200									
III. Math vs. HighScope and Creative Curriculum									
University of California-Berkeley and SUNY University of Buffalo School n=36;	CA and NY	Pre-K Mathematics supplemented with DLM Early Childhood Express (Math Software only)	Creative Curriculum or HighScope	N	Y (but in separate buildings)	N	36	19, 20, 150 17, 20, 150	
Classroom n=40; Child n=300									
IV. Creative Curriculum vs. Locally-Developed Curriculum									
University of North Carolina at Charlotte	NC and GA	Creative Curriculum	Locally Developed	N	Y	Y	17	9, 90 9, 80	

1	2	3	4	5	6	7	8	9	10
Grantee and sample size	Site	Treatment Curriculum	Control Curriculum (a)	Whole school randomized to same Treatment?	Some within-school randomization to treatment?	Children randomized across treatment conditions within school?	Total number of schools (clusters)	Number of schools, classrooms, students, if school randomization (T/C)	Number of classrooms and students, if within-school randomization (T/C)
School n=5;									
Classroom n=18; Child n=170									
Vanderbilt University School n=12;	TN	Creative Curriculum	Locally Developed	Y	N	N	6, 7, 90 6, 7, 100		
Classroom n=14; Child n=210									

Note: Ns are rounded to the nearest 10 in accordance with NCES data policies.

Table 2:

Description of findings across curricula comparisons.

Grantee	Treatment Curriculum	Control Curriculum(a)	Project-reported impacts			Pilot Year	Fidelity of Implementation	
			Literacy	Math	Socio-emotional		Treat	Control
I. Literacy vs. HighScope and Creative Curriculum								
University of North Florida	Early Literacy and Learning Model	Creative Curriculum	ns, ns, ns	ns, ns	ns, ns	Y	2.5	Not Provided
Florida State University	Literacy Express	HighScope	ns, ns, ns	ns, ns	ns, ns		2.5	2
Florida State University	DLM Early Childhood Express supplemented with Open Court Reading Pre-K	HighScope	+, +, +	+, ns	ns, ns		2.3	2
University of California-Berkeley	Ready Set Leap	HighScope	ns, ns, ns	ns, -	ns, ns		1.9	2
University of Virginia	Language Focused	HighScope	ns, ns, ns	ns, ns	ns, ns		2	2
II. Literacy vs. Locally-Developed Curriculum								
University of Texas Health Science Center at Houston	Doors to Discovery	Locally Developed	ns, ns, ns	ns, ns	ns, ns	Y	2.1	1
University of Texas Health Science Center at Houston	Let's Begin with the Letter People	Locally Developed	ns, ns, ns	ns, ns	ns, ns	Y	1.9	1
Vanderbilt University	Bright Beginnings	Locally Developed	ns, ns, ns	ns, ns	ns, ns	Y	1.9	2
III. Math vs. HighScope and Creative Curriculum								
University of California-Berkeley and SUNY University of Buffalo	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	Creative Curriculum	Locally Developed	ns, ns, ns	ns, ns	ns, ns	Y	2.1	1.5
Vanderbilt University	Creative Curriculum	Locally Developed	ns, ns, ns	ns, ns	ns, ns	Y	2.1	2

Note: "NS" = no significant impact of contrast on child skill, $p \geq .05$; "+" = beneficial impact of experimental contrast on child skills with $p < .05$; "-" indicates detrimental impact with $p < .05$. Child outcomes are ordered as follows in the "Project-reported impacts" columns; "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; Fidelity of implementation was rated on a 4-point scale (0 = Not at all; 3 = High).

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 3.

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	0.16
	(0.15)	(0.19)	(0.20)	(0.17)
N	860	840	840	850
<i>Classroom N=100</i>				
II. Literacy vs. Locally-Developed Curricula	0.56	0.47	0.78	0.28
	(0.27)	(0.36)	(0.44)	(0.25)
N	430	410	410	410
<i>Classroom N=60</i>				
III. Math vs. HighScope and Creative Curriculum	0.21	1.19	0.38	0.67
	(0.32)	(0.53)	(0.30)	(0.51)
N	200	200	200	200
<i>Classroom N=30</i>				
IV. Creative Curriculum vs. Locally-Developed Curricula	0.62	0.51	0.71	1.00
	(0.32)	(0.25)	(0.23)	(0.65)
N	340	320	320	330
<i>Classroom N=30</i>				

Note. Each entry represents results from a separate regression. Standard errors clustered at the school 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.

Table 4.

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	0.13 (0.04)	-0.02 (0.06)	0.04 (0.05)	-0.13 (0.09)
N	860	860	860	860
II. Literacy vs. Locally-Developed Curricula	0.13 (0.11)	0.14 (0.09)	0.15 (0.09)	-0.17 (0.18)
N	440	440	440	440
III. Math vs. HighScope and Creative Curriculum	0.05 (0.10)	0.35 (0.11)	0.25 (0.11)	0.15 (0.18)
N	210	210	210	210
IV. Creative Curriculum vs. Locally-Developed Curricula	0.04 (0.07)	0.00 (0.11)	0.02 (0.09)	-0.03 (0.23)
N	350	350	350	350

Note. Each entry represents results from a separate regression. Standard errors clustered at the school 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.