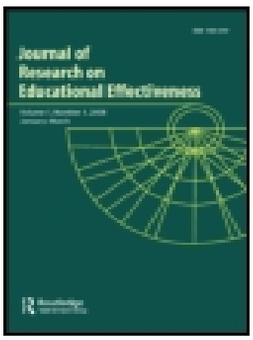

Routledge
Taylor & Francis Group

Journal of Research on Educational Effectiveness

ISSN: 1934-5747 (Print) 1934-5739 (Online) Journal homepage: https://www.tandfonline.com/loi/uree20

Distinctions Without a Difference? Preschool Curricula and Children's Development

Jade Marcus Jenkins, Anamarie Auger Whitaker, Tutrang Nguyen & Winnie Yu

To cite this article: Jade Marcus Jenkins, Anamarie Auger Whitaker, Tutrang Nguyen & Winnie Yu (2019): Distinctions Without a Difference? Preschool Curricula and Children's Development, Journal of Research on Educational Effectiveness, DOI: 10.1080/19345747.2019.1631420

To link to this article: https://doi.org/10.1080/19345747.2019.1631420

JOURNAL OF RESEARCH ON EDUCATIONAL EFFECTIVENESS
https://doi.org/10.1080/19345747.2019.1631420

THEORY, CONTEXTS, AND MECHANISMS

Check for updates

Distinctions Without a Difference? Preschool Curricula and Children's Development

Jade Marcus Jenkins[a], Anamarie Auger Whitaker[b], Tutrang Nguyen[a]* and Winnie Yu[a]

ABSTRACT

Public preschool programs require the use of a research-based, whole-child curriculum, yet limited research examines whether curricula influence classroom experiences and children's development. We use five samples of preschool children to examine differences in classroom processes and children's school readiness by classroom curricular status (curriculum/no curriculum), and across classrooms using different curricular packages. When a teacher reports using a curriculum, their classroom processes are indistinguishable from classrooms where teachers report using no curriculum. Some differences in classroom activities emerged across classrooms using different curricula; however, substantial variability exists across classrooms using the same curriculum. Head Start program fixed effects models and meta-analytic regressions reveal few associations between curricula and children's skills. Findings question whether preschool curricular policy benefit child development.

KEYWORDS

preschool curricula
early childhood education
classroom quality
Head Start
prekindergarten
school readiness

Introduction

Do preschool curricula promote child development? The vast majority of publicly funded preschool programs—center-based early education for three- and four-year-olds—require the use of "research-based curricula." Head Start programs are mandated to use research-based "whole-child" curricula. Federally and state-sponsored quality rating and improvement systems (QRIS) incorporate curriculum into their rankings and consider the use of a developmentally appropriate, research-based curriculum to be an indication of program quality (e.g., Auger, Karoly, & Schwartz, 2015). Tax dollars invested in funding public preschool programs—totaling $18.3 billion in 2015—are thereby also invested in curricula (Barnett, Carolan, Squires, Brown, & Horowitz, 2015; Isaacs, Edelstein, Hahn, Steele, & Steuerle, 2015). With an average price tag of $2,000 per classroom, curricula policies benefit publishers, but it is unclear whether they benefit preschool children.

CONTACT Jade Marcus Jenkins jvjenkin@uci.edu University of California Irvine, 3200 Education, Irvine, CA 92697, USA.
[a]University of California Irvine, Irvine, California, USA;
[b]RAND Corporation, Santa Monica, California, USA
*Tutrang Nguyen is now affiliated with the University of Virginia, Charlottesville, Virginia, USA.

Supplemental data for this article can be accessed on the publisher's website.

In fact, we know very little about whether and how commonly used preschool curricula influence children's school readiness. Although most publishers claim that their curricula are research-based, few describe either the research on which the claim is based or how the curricula materials are explicitly linked to children's development (Clements, 2007). Data from Head Start programs and from a national sample of child-care centers indicate that the most commonly used curriculum is the Creative Curriculum (Hulsey et al., 2011; Jenkins & Duncan, 2017), despite its rating by the What Works Clearinghouse as having "no discernable effectiveness" in promoting school readiness (U.S. Department of Education, 2013). The second most commonly used curriculum is HighScope (Hulsey et al., 2011; Jenkins & Duncan, 2017), whose only rigorous evidence comes from the Perry Preschool study; a small, intensive demonstration program conducted in the 1960s with counterfactual conditions that no longer apply to the current preschool population (i.e., children who did not attend center-based preschool; Belfield, Nores, Barnett, & Schweinhart, 2006; Duncan & Magnuson, 2013; Schweinhart, 2005).

Also unknown is whether different curricular packages vary in terms of their implemented activities and instructional practices (e.g., language and literacy activities, small- versus large-group instruction), which structure the very basis of children's preschool experiences. Furthermore, prior research consists primarily of researcher-designed curricula studies implemented in highly controlled settings or using limited samples of preschool classrooms; few studies have examined the relationship among curricula, classroom activities, and children's school readiness in business-as-usual preschool settings.

Our study is a comprehensive examination of widely used preschool curricula and their associations with preschool classroom environments and children's academic and social-emotional development using five large samples of low-income three- and four-year-old children attending public preschool programs operating at scale. We examine patterns in classroom activities and the emotional, instructional, and overall quality in classrooms with and without a whole-child curriculum in use, and compare associations between curricula and quality by curricular package (e.g., Creative Curriculum vs. HighScope). Our study provides the first detailed description of the curricular landscape in preschool programs using the best available data (samples that include classroom observations, teacher surveys, curricular package information, and child outcome assessments). In addition to these descriptive calculations, we estimate quasi-experimental impact models—Head Start grantee fixed effects or state fixed effects—to analyze the relationship between classroom curricular package and child school readiness outcomes. Examining how different curricula influence the quality and type of activities in preschool classrooms, and subsequently children's development, is essential to understanding the policy levers that make preschool effective for low-income children.

Curricula and Children's Development

Curricula set goals for the knowledge and skills that children should acquire in an educational setting. They guide and support educators' plans for providing the day-to-day learning experiences to cultivate those skills with daily lesson plans, materials, and other pedagogical tools (Goffin & Wilson, 1994; Ritchie & Willer, 2008). Curricula differ across a number of dimensions, such as philosophies, materials, the role of the teacher,

pedagogy or modality (e.g., small or large group setting), classroom design, and child assessment. Preschool programs can choose their own curricula, but their choices are often constrained by a preapproved list developed by state agencies, accrediting bodies, or funding sources (Clifford & Crawford, 2009). Most programs, such as Head Start, require a curriculum that provides enriching experiences across the multiple domains of children's development (e.g., health, social-emotional, academic), known as "whole-child" curricula. The whole-child approach is anchored in Piagetian theory, which emphasizes child-centered active learning cultivated through the strategic arrangement of the classroom environment (DeVries & Kohlberg, 1987; Piaget, 1976; Weikart & Schweinhart, 1987) and sociocultural theory, where the teacher provides supportive and responsive interactions with children (Vygotsky, 1978). Whole-child curricula purport to emphasize critical thinking and problem-solving skills by providing open-ended learning opportunities and simultaneously cultivating the interrelated domains of children's development (Diamond, 2010; Elkind, 2007; Zigler & Bishop-Josef, 2006).

In addition to Creative Curriculum, HighScope, Scholastic, and High Reach are other whole-child curricula widely used in preschool programs, including Head Start and state pre-K (Clifford et al., 2005; Hulsey et al., 2011; Jenkins & Duncan, 2017; Phillips, Gormley, & Lowenstein, 2009). Despite their widespread adoption, little empirical support exists for HighScope, none exists for Creative Curriculum, and neither curriculum has demonstrated effectiveness based on rigorous standards when compared with business-as-usual preschool settings (i.e., teacher-developed curricula or no curricula; Belfield et al., 2006; Preschool Curriculum Evaluation Research Consortium, 2008; Schweinhart, 2005; U.S. Department of Education, 2013).

The dearth of evidence supporting Creative Curriculum and HighScope is not unique to whole-child curricula. Most recently, the National Center on Quality Teaching and Learning of the Office of Head Start (2014) released the "Preschool Curriculum Consumer Report," the first of its kind, which reviewed the most commonly used curricula in Head Start programs nationwide and provided ratings for each based on a set of 13 criteria. One criterion is "Curriculum is Evidence-Based." Of the 14 curricula reviewed in the report, seven had "no evidence," five had "minimal evidence," one had "some evidence," and only one was rated to have "solid, high-quality evidence" (Opening the World of Learning) with demonstrated effects on child outcomes. One of the first Institute for Education Sciences–funded research projects was the Preschool Curriculum Evaluation Research Study Initiative (PCER; 2008), a large multi-site, random assignment experimental study of 14 different preschool curricula. In this study, only two curricula, both of which were content specific (i.e., math or literacy focused), were found to be effective at promoting children's school readiness when compared with business-as-usual counterfactual settings (which included whole-child curricula classrooms).

However, evidence does suggest that other types of less commonly used curricula—when implemented with high-quality professional development, including coaching supports—can have strong impacts on children's early academic and social-emotional development. Findings from small, randomized control trials of well-implemented, content-specific curricula that target single developmental domains show positive, small to moderate impacts on skills targeted in the curricular materials (Bierman et al., 2008; Clements & Sarama, 2008; Diamond, Barnett, Thomas, & Munro, 2007; Fantuzzo,

Gadsden, & McDermott, 2011; Morris et al., 2014). For example, children who received a literacy-targeted curriculum showed improvements in their literacy and language skills compared with business-as-usual conditions (i.e., HighScope, Creative Curriculum, or teacher-developed curricular models; Justice et al., 2010; Lonigan, Farver, Phillips, & Clancy-Menchetti, 2011). Clements and Sarama (2007, 2008) found large gains in math achievement from a targeted preschool mathematics curriculum relative to classrooms using business-as-usual curricula. Results are comparable for curricula aimed at promoting children's social-emotional development (Bierman et al., 2008; Morris et al., 2014). Boston's successful public pre-K program uses a unique curricular approach that combines two content-specific curricula bundled with strong, ongoing professional development, including coaching for its teachers (who are also well paid and highly educated) to achieve its program impacts on children's learning (Weiland & Yoshikawa, 2013).

One might argue that if implemented with similarly strong professional development supports, whole-child curricula may do just as well as the successful content-specific curricula described previously, and it is not the presence of a curriculum per se driving impacts. However, evidence from the PCER study (2008) does not suggest that this is the case. One of the study sites randomly assigned classrooms to the Creative Curriculum as the treatment condition, and therefore received the training and implementation supports afforded to experimental sites to ensure program fidelity. Still, Creative Curriculum classrooms in the treatment condition were no more effective in promoting children's outcomes compared with the locally developed curricular approach that the schools otherwise would have used. Professional development is an important component of any preschool program, but there exists little data to suggest that the lack of evidence on whole-child curricular effectiveness is the result of professional developmental models alone. If early learning policies require the use of whole-child curricula, greater empirical support is needed to understand their value added to the preschool experience.

Curricula and Early Childhood Education Policy

A surfeit of research shows that high-quality preschool can promote children's cognitive and physical development, particularly for low-income children (Barnett, 2011; Duncan & Magnuson, 2013; Gormley, Phillips, & Gayer, 2008; Yoshikawa et al., 2013). Yet the tremendous variability in preschool quality, implementation, and effectiveness both within and between different types of programs (e.g., Head Start and state prekindergarten) and between states reveals how little is known about precisely *what* makes preschool effective (Bloom & Weiland, 2015; Dotterer, Burchinal, Bryant, Early, & Pianta, 2009; Jenkins, 2014; Jenkins, Farkas, Duncan, Burchinal, & Vandell, 2016; Karoly, Ghosh-Dastidar, Zellman, Perlman, & Fernyhough, 2008; Walters, 2015). Furthermore, widespread recent attention to the persistence or fadeout of the impacts of preschool programs raises concerns among policy stakeholders as to how programs can ensure continued learning gains and produce "returns" on these human capital investments as pre-K programs continue to expand (Phillips et al., 2017). Policy efforts at the federal, state, and local levels traditionally use three main levers to improve the effectiveness of public preschool programs: (1) increasing teachers' skills through raising educational requirements and funding

professional development; (2) creating quality improvement, licensing, and monitoring systems; and (3) requiring preschool curricula to guide instruction.

Although often overlooked, curricular requirements and curricula use are embedded in these and other policies that govern early care and education systems. Preschool programs mandate that teachers use a curriculum, curricula prescribe specific classroom activities and practices using various pedagogical approaches, and these activities represent the learning experiences that cultivate children's readiness for school. Therefore, instructional materials and the strategies promoted by curricula constitute some of the most direct policy-relevant connections to learning activities in the classroom, especially in light of the strong impact evidence from studies of content-specific curricula.

Still, such requirements can be vague. For example, a recent survey of state education agencies revealed that states have loose requirements for pre-K curricular decisions (e.g., "research-based" curricula, with "research-based" ill defined) or basic guidelines for selection, such as alignment to state early learning standards (Dahlin & Squires, 2016). In most cases, educators choose among preselected curricular options based on local or state policies with little scientific guidance, a few popular selections, and substantial costs.

Most importantly, however, published curricula packages may differ, on average, in the experiences they shape for children in preschool classrooms. In other words, when enacted in preschool programs at scale, it is unclear whether certain curricular packages are more likely than others to promote developmentally appropriate learning activities. Additionally, there exists no population-level information about the extent to which classroom experiences and instruction using different, or even the same, curricular package vary across classrooms. In theory, the curriculum drives classroom activities, and so classrooms whose teachers report using the same curriculum should be comparable with respect to quantity and type of activities (e.g., math and literacy instruction), and perhaps overall instructional quality. This assumption is dependent on a curriculum being properly enacted with fidelity across preschool classrooms. However, if program features such as length of day or funding for materials vary between classrooms and centers, the classroom experiences generated by curricula packages may differ. Similarly, teacher training and attitudes towards curricula likely affect implementation (e.g., using only part of a curriculum, modifying instruction). Although it is likely that policy-mandated curricula are not, on their own, the primary determinant of children's development in preschool, it is certainly important to know whether curricula steer classroom experiences and raise the overall quality of instruction and support from teachers to promote children's learning. Empirically derived curricula guidance or restrictions may be an efficient mechanism through which policy can improve the consistency and effectiveness of preschool programs.

Another critical policy consideration is that curricula are a significant investment for preschool programs. In the first column of Table 1, we present the approximate costs per classroom for commonly used curricula, which range between $1,125 and $4,190. Not included in these estimates are the additional professional development activities often strongly recommended by publishers to implement the curricula with fidelity, and the costs of supplemental materials. The Head Start program alone has more than 50,000 classrooms, making the costs of such policies nontrivial (Office of Head Start, 2010).

Table 1. Descriptions of analysis sample and key measures by study.

	PCER	NCEDL	HSIS	FACES 2003	FACES 2009
Classroom sample	170	245	997	308	456
Child sample	–	394	1,654	1,565	2,401
Year of preschool data collection	2003	2001	2002	2003	2009
Percent Head Start classrooms	35	29	100	100	100
Percent state prekindergarten classrooms	53	71	–	–	–
Curricula included (cost per classroom)					
HighScope ($1,150)	Classrooms = 40 Children = 380	Classrooms = 84 Children = 359	Classrooms = 350 Children = 656	Classrooms = 53 Children = 325	Classrooms = 66 Children = 424
Creative Curriculum ($2,149)	Classrooms = 50 Children = 360	Classrooms = 38 Children = 161	Classrooms = 416 Children = 854	Classrooms = 161 Children = 1,123	Classrooms = 251 Children = 1,699
DLM Express ($4,108)	Classrooms = 10 Children = 100				
High Reach ($1,125)			Classrooms = 40 Children = 74	Classrooms = 17 Children = 130	Classrooms = 24 Children = 179
Scholastic ($2,900)			Classrooms = 58 Children = 114		Classrooms = 14 Children = 106
Other published curriculum		Classrooms = 32 Children = 135	Classrooms = 193 Children = 361	Classrooms = 77 Children = 514	Classrooms = 101 Children = 709
No published curriculum	Classrooms = 70 Children = 600	Classrooms = 80 Children = 360			
Classroom activities	Teacher Behavior Rating Scale (Literacy and Math)	Early Academic Snapshot	Teacher survey responses	Teacher survey responses (Literacy)	Teacher survey responses (Literacy and Math)
Classroom quality	ECERS-R, Arnett Caregiver Interaction Sale	ECERS-R, CLASS	ECERS-R, Arnett Caregiver Interaction Scale	ECERS-R, Arnett Caregiver Interaction Scale	ECERS-R, CLASS
Child outcomes					
Early academic	–	PPVT, WJ-III Letter Word, Applied Problems	PPVT, WJ-III Letter Word, Spelling, Applied Problems	PPVT, WJ-III Letter Word, Spelling, Applied Problems	PPVT, WJ-III Letter Word, Spelling, Applied Problems
Social and behavioral	–	Teacher-Child Rating Scale	Problem Behavior Rating Scale	Teacher Rating Scale	Teacher Rating Scale

Notes: PCER: Preschool Curriculum Evaluation Research; ECERS: Early Childhood Environment Rating Scale; TBRS: Teacher Behavior Rating Scale; NCEDL: National Center for Early Development and Learning; FACES: Family and Child Experiences Survey; ECERS-R: Early Childhood Environmental Rating Scale-Revised; CLASS: Classroom Assessment Scoring System; WJ: Woodcock-Johnson. PCER classroom count rounded to the nearest 10 per NCES license requirements. PCER child sample and outcome measures are not shown because we could not conduct our child-level analyses using PCER because of the study sample design; see "Analyses" in main text for more detail. Classroom and child samples reflect our study's analytic sample, and not the original study samples. Per-classroom cost estimates are approximated from the cost of purchasing the curriculum teacher's manual or equivalent (in 2015), and the baseline set of materials required to implement the curriculum. Publishers offer different sets of materials, and thus costs will vary by publisher and curriculum.

Given the wide array of curricular choices available, the government expenditures for required curricula, and our insufficient understanding of whether commonly used whole-child curricula promote children's school readiness, a comprehensive study of preschool curricula is badly needed.

Present Study

Our study is an examination of widely used published preschool curricula including Creative Curriculum, HighScope, Scholastic, High Reach, and DLM Express. Four of the five curricula are marketed as "research-based"; however, there exists no or only minimal empirical evidence linking these curricula to children's outcomes (National Center on Quality Teaching and Learning, 2014). Using five large samples of low-income, racially and ethnically diverse preschool children, we aim to understand how preschool curricula relate to classroom activities and quality as they are used in business-as-usual, center-based settings, and subsequently to children's academic and social-emotional development. Specifically, our three research questions (RQ) are:

1. To what extent do classroom activities and quality ratings vary by whether a published curriculum is in use, and in classrooms that do use a published curriculum, do activities vary by the specific curricular package (e.g., HighScope compared with Creative Curriculum)?
2. To what extent is having a published curriculum in use in a preschool classroom associated with children's academic and social-emotional school readiness, and do children's readiness vary by the specific curricular package?
3. To what extent are the classroom activities, overall classroom quality ratings, and teacher's attitudes and perceptions of curriculum consistent among classrooms using the same, or different, curricular packages?

Little prior research exists on whole-child curricula, making predictions about which packages may improve classroom quality and child outcomes difficult. However, because curricula inherently guide classroom processes, we expect differences in classroom process quality between classrooms that do and do not have a published curriculum in use. Because all whole-child curricular packages aim to promote development across multiple domains and are similar in their theoretical approach and pedagogy, we expect that these packages are robust to different classrooms and are similarly related to classroom quality and child outcomes. Specifically, we hypothesize that: (1) there exist differences in classroom process quality between classrooms with and without curricula in use; (2) there are similar levels of process quality in classrooms using different whole-child curricular packages, albeit with different ways of structuring classroom activities; and (3) classroom activities, overall classroom quality ratings, and teacher's attitudes and perceptions of curricula are consistent across classrooms using the same curricular package.

Hereafter, we use the term "curricular status" to describe whether a classroom has any curricula in use (i.e., yes/no), whereas "curricular package" refers to the specific published curriculum in use (e.g., Creative Curriculum).

Method

Data

Our study uses secondary data from five studies of children in preschool settings between the 2001 and 2009 school years: The Preschool Curriculum Evaluation Research Study (PCER), the National Center for Early Learning and Development Multi-State Study of Pre-Kindergarten (NCEDL), the Head Start Impact Study (HSIS), the Head Start Family and Children Experiences Survey, 2003 Cohort (FACES 2003), and the Head Start Family and Children Experiences Survey, 2009 Cohort (FACES 2009). Each data set contains information about curricula, classroom activities, and child academic and social-emotional outcomes. In all five studies, data collection took place in center-based preschool settings, and the child participants were majority low-income and were ethnically and racially diverse. We describe each study's sample and measures in the following sections and summarize this information in Table 1 (additional information about measures is presented in Appendices A.1–A.5).

Before proceeding, we acknowledge that our study data sets are somewhat dated and therefore may not reflect the most current classroom practices and activities. We assessed the extent to which the 2009 FACES cohort—the most recent snapshot of curricula and classroom practices in Head Start centers—compares with both the 2003 FACES cohort and the 2002 HSIS sample to examine differences in practice across years. This comparison indicates that the curricular choices of Head Start centers remained fairly stable over time (Creative Curriculum, HighScope, High Reach, Scholastic, in order of frequency) and closely matches the most recent available national data on curricula use (from the 2012 National Study of Early Care and Education; Jenkins & Duncan, 2017). Descriptive analyses are discussed in greater detail in the "Results" section and are displayed in Table 4. In addition, our data are heavily weighted toward Head Start centers; three of the data sets include only Head Start programs (HSIS and FACES), and the other two include a combination of center-based preschool settings, including state pre-K and Head Start. Although this somewhat limits the interpretation of our results, we also consider this a strength because such programs are universally subjected to the whole-child curricular mandates imposed by federal policy.

Samples

PCER. Beginning in 2003, 12 grantees across the country were funded to study the effect of preschool curricula on children's academic and social-emotional outcomes in the PCER study. Each grantee selected their study curricula for a total of 14 different curricula tested in 18 different locations. Mathematica Policy Research and the Research Triangle Institute assisted with the evaluation to ensure consistent data collection at each site, but each grantee was in charge of its own evaluation. Individual grantees were responsible for recruiting preschool centers to participate in the study. At each grantee site, either classrooms within preschool centers or entire centers themselves were randomly assigned to a treatment (experimental curriculum) or control condition. For feasibility and to preclude cross contamination across classrooms, most research sites

assigned only one curriculum to each preschool center. Baseline data on children, parents, and preschools were collected in the fall of 2003, with post-treatment data collected in the spring of 2004. Approximately 2,900 children in 320 preschool classrooms participated in the study. The subsample of PCER most relevant to our study included the grantee sites and classrooms that used one of our focal whole-child curricula—HighScope, Creative Curriculum, DLM Express—and those classrooms with no published curriculum in use (N = 1,450 children). The data include children who were either in Head Start, private child care, or public preschool. For more information about the study, see the PCER Final Report (2008).

NCEDL. This study comprises two stratified random samples of children within preschool programs across 11 states. States were purposely selected if they had large numbers of children enrolled in preexisting public pre-K programs. The sample for the Multi-State Study of Pre-Kindergarten includes six states (California, Illinois, Georgia, Kentucky, New York, and Ohio). No systematic intervention was tested in NCEDL; data were collected to examine the characteristics of and variations in programs that lead to children's development. The follow-up study, the State-Wide Early Education Programs Study, was not included in our analyses because the data set did not include curriculum indicators. Preschool programs were randomly sampled within states, and 29% were Head Start programs. One classroom was then randomly sampled within each program, and 94% of classroom teachers agreed to participate. Of the selected classrooms, approximately 60% of parents gave consent for their child to participate, and from this subsample four children were randomly selected to participate (N = 1,015). Forty preschool programs were selected in each state for a total of 245 classrooms. Child assessment data were collected during the fall and spring of the 2001–2002 preschool year. For more information, see Early et al. (2005).

***HSIS*.** The HSIS is a nationally representative study of Head Start participants and a group of comparable non-participants from 23 states that were sampled using a complex multistage stratified design. Head Start grantees were divided into geographic clusters and were then stratified based on grantee characteristics, with three grantees or delegate agencies randomly selected from each cluster. Within each delegate agency, Head Start centers were stratified in the same way as grantees and were randomly selected. This resulted in 84 grantees and delegate agencies with a total of 383 individual preschool centers. The full sample included newly entering three- and four-year-old Head Start applicants at randomly selected oversubscribed centers, where children were randomly assigned to receive an offer for Head Start. A total of 4,442 children were selected—2,646 for Head Start and 1,796 for the control condition. Control-group participants either found other child care or the child was cared for at home. Study investigators (Westat) collected baseline surveys and child assessments during the fall of 2002, and posttreatment child assessments were collected at the end of Head Start in the spring of 2003.

We restrict the sample for our study to those children who were randomly assigned to, and actually attended, a Head Start program because only under these conditions were classrooms required to have a curricular package in use. Control children in the HSIS were omitted from our study because of the extensive variation in counterfactual care conditions. For more information, see the HSIS Final Evaluation Report (Puma et al., 2012).

FACES. The Head Start Family and Child Experiences Survey (FACES) study is a multi-wave, large-scale investigation of children, families, and educators in Head Start programs that aims to understand how the program operates and how it contributes to the well-being of the families and children it serves. Similar to NCEDL, the FACES study is not an intervention study. The FACES data contain nationally representative longitudinal data about five cohorts of Head Start children and their families (i.e., FACES 1997, 2000, 2003, 2006, and 2009) as well as staff qualifications, classroom practices, and quality measures including curricula indicators. Our analyses use data from the 2003 and 2009 cohorts. We selected the 2003 cohort because the data time frame closely aligned with our other study data sets. We included the 2009 cohort because they were the most recent FACES data available at the time of our study. The FACES sampling design included a four-stage sampling process to select a representative group of Head Start (1) grantees, (2) centers, (3) classrooms, and (4) newly enrolled children. Sampling at the first three stages was done with probability proportional to size. Data were collected in the fall and spring of the children's first year in Head Start, and the spring of the children's second year in Head Start if they were three years old at first entry. Although teachers were allowed to select multiple published curricula used in their classrooms, the FACES study also asked teachers to name the *primary* curriculum they used in class, which we used as our key independent variable. In total, the FACES 2003 sample included 63 grantees, 182 centers, 409 classrooms, and 2,816 children. The FACES 2009 sample included 60 grantees, 129 centers, 486 classrooms, and 3,349 children. For more information, see the FACES User's Guide (Malone et al., 2013; Zill, Kim, Sorongon, Shapiro, & Herbison, 2008).

Measures

Preschool curricula. Each data set includes classrooms using published curricula. Additionally, both the NCEDL and PCER samples include preschool classrooms with no published curriculum in use. "No published curriculum" means that the classroom did not use a published or packaged curriculum but may have used a locally developed or a teacher-designed curriculum. Although we cannot know the exact content of these curricula or the curricula models on which they are based, we consider the "no published curriculum" and the locally or teacher-developed curriculum designations to represent another common practice in early childhood education, and thus important to include in our study. In the NCEDL, HSIS, FACES 2003, and FACES 2009 studies, a category indicating "Other published curricula" represents those classrooms for which we do not have specific curricular package information, or with fewer than 10 classrooms using a specific curriculum package. These classrooms were collapsed into a single group for analysis. Note that fewer than five classrooms reported using Scholastic in the FACES 2003 and were not included in the analysis.

We acknowledge that teachers may report using a curriculum when it may merely be present on their classroom bookshelves. However, the aim of our study is to understand the implications of policy-mandated curricula. As such, our data represent the de facto classroom environments for children who experienced different curricular choices with at-scale business-as-usual implementation.

Table 2. Research questions, hypotheses, and descriptions of analyses by study.

RQ 1: To what extent do classroom activities and quality ratings vary by whether a published curriculum is in use, and in classrooms that do use a published curriculum, do activities vary by the specific curricula package (e.g., HighScope compared with Creative Curriculum)?					
Hypothesis: Classroom process quality level will differ between classrooms that do and do not have a published curriculum, with classrooms using published curricula having higher levels of quality. We anticipate that quality levels will not significantly differ between specific curricula packages.					
Analysis:	PCER	NCEDL	HSIS	FACES 2003	FACES 2009
t tests of means or *z* tests of proportion to compare classroom activities by curricular status	X	X			
ANOVAs to compare classroom activities and quality ratings by curriculum package	X	X	X	X	X
Regressions of classroom activities on indicators for curricular package	X	X	X	X	X
RQ 2: To what extent is having a published curriculum in use in a preschool classroom associated with children's academic and social-emotional school readiness, and do children's readiness vary by the specific curricular package?					
Hypothesis: Classroom process quality level will differ between classrooms that do and do not have a published curriculum, with classrooms using published curricula having higher levels of quality. We anticipate that quality levels will not significantly differ between specific curricula packages.					
Analysis:	PCER	NCEDL	HSIS	FACES 2003	FACES 2009
Curricular status state fixed effects models		X			
Curricular package grantee fixed effects models				X	X
Curricular package state fixed effects models		X			
Meta-analysis of curricular package estimates		X	X	X	X
RQ 3: To what extent are the classroom activities, overall classroom quality ratings, and teacher's attitudes and perceptions of curriculum consistent among classrooms using the same, or different, curricular packages?					
Hypothesis: Classroom activities, overall classroom quality ratings, and teacher's attitudes and perceptions of curriculum are consistent across classrooms using the same curricular package.					
Analysis:	PCER	NCEDL	HSIS	FACES 2003	FACES 2009
Histograms of classroom quality and classroom activities by curricular package overlaid with no packaged curriculum	X	X			
Histograms of classroom quality and classroom activities by curricular package			X	X	X
Descriptive comparison of teacher survey items on curricula		X	X	X	X

Notes. The term "curricular status" is used to describe whether a classroom has any curricula in use (i.e., yes/no), whereas "curricular package" refers to the specific published curriculum in use (e.g., Creative Curriculum). PCER: Preschool Curriculum Evaluation Research; NCEDL: National Center for Early Development and Learning; FACES: Family and Child Experiences Survey.

To provide some context of curricular implementation and teachers' perspectives on curricula, we use the available teacher survey items related to curriculum in the NCEDL, HSIS, and FACES data sets (teacher curriculum items not collected in PCER) in our descriptive analyses related to curricular variation (RQ 3). Items and their responses are aggregated by curricular packages, are shown in Appendix Tables A.2–A.5, and capture things such as teacher's attitudes toward the curriculum, whether they have training in the curriculum, whether they have the necessary materials to implement the curriculum, and whether the curriculum leaves room for teacher creativity. All items are indicator variables and equal 1 if the teacher responded "yes" to the question prompt.

Classroom quality. Quality of care was measured with several instruments across the three studies. The Early Childhood Environment Rating Scale-Revised (ECERS-R; Harms, Clifford, & Cryer, 1998) is a widely used observer-rated measure of global classroom quality, specifically designed for use in classrooms serving children between 2.5 and 5 years of age, and was used in each study. Scores on the ECERS-R range from 1 to 7, with 1 indicating "inadequate" quality, 3 indicating "minimal" quality, 5 indicating "good" quality, and 7 indicating "excellent" quality. The scale's authors report a total scale internal consistency of .92. We report the total ECERS scale score, and the "Provisions for Learning" and "Interactions" factor scores for each study. We focus our classroom-level quality analyses on the ECERS because it was collected in all four studies. However, we incorporate two additional quality measures, each shared by two or three studies, in our descriptive analyses shown in Tables 3 and 4.

To capture caregiver interactions, the HSIS, PCER, and FACES 2003 studies used the Arnett Caregiver Involvement Scale (Arnett, 1989). This is an observational measure consisting of 26 items reflecting teacher sensitivity, harshness, and detachment that are rated on a scale of 1–4, indicating how characteristic they are of the teacher from not at

Table 3. Classroom activity comparison by presence of published curricula in PCER and NCEDL.

	PCER			NCEDL		
	Published Curriculum	No Published Curriculum	Diff.	Published Curriculum	No Published Curriculum	Diff.
Classroom Activities						
TBRS Math Quantity (0–3 scale)	1.22	.94	*	–	–	
TBRS Literacy Quantity (0–3 scale)	1.51	1.19	*	–	–	
Snapshot: Math Activity (proportion of day)	–	–		.06	.07	
Snapshot: Literacy Activity (proportion of day)	–	–		.15	.15	
Classroom Quality						
Arnett Caregiver Interaction	3.21	2.95	*	–	–	
Total ECERS Score	4.31	3.34	*	3.89	3.59	*
ECERS Factor 1 Language/Interactions	4.94	3.91	*	4.52	4.31	
ECERS Factor 2 Provisions for Learning	4.32	3.26	*	3.98	3.46	*
CLASS Emotional Support Scale	–	–		5.31	5.40	
CLASS Instructional Support Scale	–	–		1.91	1.98	
Observations (Classrooms)	100	70		154	91	

Notes. PCER: Preschool Curriculum Evaluation Research; NCEDL: National Center for Early Development and Learning; ECERS: Early Childhood Environment Rating Scale; TBRS: Teacher Behavior Rating Scale; Snapshot: Emerging Academics Snapshot. $^{*}p < .05$ from t test for differences in means. All PCER classrooms observations rounded to the nearest 10 per NCES data security policy. Comparisons of all classroom characteristics and activities by curriculum are shown in Appendix B.

Table 4. Classroom activity and quality rating comparisons by curriculum.

PCER	HighScope	Creative Curriculum	DLM Express	No Published Curriculum		Diff.	*F* stat.
Classroom Activities (0–3 scale)							
TBRS Math Quantity	1.15	*1.29*	1.21	.94		*	4.06
TBRS Literacy Quantity	1.53	1.47	*1.60*	1.19		*	8.04
Classroom Quality							
Total ECERS Score	4.19	4.31	*4.77*	3.34		*	14.84
ECERS Factor 1 Language/Interactions	4.84	4.95	*5.25*	3.91		*	7.99
ECERS Factor 2 Provisions for Learning	4.33	4.22	*4.65*	3.26		*	16.34
Arnett Caregiver Interaction Score	3.12	*3.30*	*3.25*	2.95		*	4.19
Observations (Classrooms)	40 (23%)	50 (29%)	10 (6%)	70 (41%)			
NCEDL	**HighScope**	**Creative Curriculum**		**No Published Curriculum**	**Other Published Curriculum**	**Diff.**	***F* stat**
Classroom Activities (in proportion of day)							
Snapshot: Math Activity	.05	.07		.07	*.08*	*	3.21
Snapshot: Literacy Activity	.15	.14		.15	.16		.37
Classroom Quality							
Total ECERS Score	*3.95*	3.76		3.60	*3.91*	*	2.92
ECERS Factor 1 Language/Interactions	4.60	4.18		4.33	4.70		1.71
ECERS Factor 2 Provisions for Learning	*4.05*	3.92		3.46	3.87	*	6.87
CLASS Emotional Support Scale	5.34	5.14		5.42	5.45		1.42
CLASS Instructional Support Scale	1.89	1.80		1.99	2.10		1.03
Observations (Classrooms)	84 (36%)	38 (16%)		80 (34%)	32 (14%)		
HSIS	**HighScope**	**Creative Curriculum**	**High Reach**	**Scholastic**	**Other Published Curriculum**	**Diff.**	***F* stat**
Classroom Activities (in times per month)							
Total Math Activities	149.81	153.68	*169.43*	159.71	152.49	*	6.53
Total Literacy Activities	107.40	114.09	119.33	125.80	107.51		1.92
Classroom Quality							
Total ECERS Score	*5.32*	5.05	4.68	4.90	5.06	*	7.31
ECERS Factor 1 Language/Interactions	*5.79*	5.47	5.50	5.41	5.54	*	4.21
ECERS Factor 2 Provisions for Learning	*5.21*	4.96	4.01	4.65	4.88	*	12.04
Arnett Caregiver Interaction Score	2.56	2.50	*2.66*	2.51	2.53	*	2.42
Observations (Classrooms)	350 (33%)	416 (40%)	40 (4%)	58 (5%)	193 (18%)		

(*continued*)

Table 4. Continued.

FACES 2003	HighScope	Creative Curriculum	High Reach		Other Published Curriculum	Diff.	*F* stat
Classroom Activities (in times per month)							
Total Literacy Activities	163.08	164.68	187.76		160.96		1.08
Classroom Quality							
Total ECERS Score	4.21	4.24	3.58		*4.28*	*	3.29
ECERS Factor 1 Language/Interactions	*4.37*	4.31	3.63		*4.38*	*	2.90
ECERS Factor 2 Provisions for Learning	4.30	*4.41*	3.72		*4.43*	*	2.99
Arnett Caregiver Interaction Score	14.47	14.46	14.65		15.09		1.14
Observations (Classrooms)	53 (17%)	161 (52%)	17 (6%)		77 (25%)		

FACES 2009	HighScope	Creative Curriculum	High Reach	Scholastic	Other Published Curriculum	Diff.	*F* stat
Classroom Activities (in times per month)							
Total Math Activities	107.88	114.33	119.50	*125.52*	106.65	*	6.36
Total Literacy Activities	151.40	153.74	*167.33*	160.11	152.62		1.28
Classroom Quality							
Total ECERS Score	4.14	*4.45*	4.16	3.60	4.17	*	43.42
ECERS Factor 1 Language/Interactions	*5.34*	5.04	4.69	4.92	5.01	*	7.79
ECERS Factor 2 Provisions for Learning	2.56	2.49	*2.67*	2.51	2.54	*	2.83
CLASS Emotional Support Scale	5.17	*5.38*	*5.35*	5.13	5.28	*	21.51
CLASS Instructional Support Scale	2.10	*2.32*	2.24	2.11	*2.30*	*	7.54
Observations (Classrooms)	70 (15%)	228 (48%)	31 (6%)	36 (7%)	112 (23%)		

Notes. Largest significant value(s) for each measure is italicized. PCER: Preschool Curriculum Evaluation Research; ECERS: Early Childhood Environment Rating Scale; TBRS: Teacher Behavior Rating Scale; NCEDL: National Center for Early Development and Learning; FACES: Family and Child Experiences Survey; Snapshot: Emerging Academics Snapshot; HSIS: Head Start Impact Study. $^{*}p < .05$ from ANOVA. All PCER classroom observations rounded to the nearest 10 per NCES data security policy. Classroom math activities was not included in the FACES 2003 teacher survey. Comparisons of classroom and center characteristics as well as classroom activities and quality by curriculum are shown in Appendix A.1–A.5.

all (1) to very much (4). Psychometric analyses suggest that the items load onto a single factor (Cronbach's $\alpha = .93$).

The NCEDL and FACES 2009 studies also included the Classroom Assessment Scoring System (CLASS; Pianta, La Paro, & Hamre, 2008), an observer-rated assessment of teacher-child interactions in terms of emotional support (climate, teacher sensitivity, regard), classroom organization (behavior management, productivity, instructional learning formats), and instructional support (concept development, feedback quality, language modeling; Cronbach's α .88 for classroom organization, .90 for emotional support, and .93 for instructional support).

Classroom learning activities. We used different instruments and data sources in each study to create aggregate measures of total classroom literacy and mathematics activities. Detailed lists of the individual items used, along with mean values by curricular package, are available in Appendices A.1–A.5.

The Teacher Behavior Rating Scale from the PCER study used trained observers to rate the quality and quantity of academic activities present in a classroom (Landry et al., 2001). There are two content areas measured by the TBRS: math and literacy. Literacy is composed of five subdomains (written expression, print and letter knowledge, book reading, oral language, and phonological awareness). Quality of activities were rated from 0 to 3 (0 = activity not present; 3 = activity high quality). Quantity of activities was similarly rated from 0 to 3 (0 = activity not present; 3 = activity happened often or many times). We focus only on the quantity measures in our analyses, and this number was derived from taking the average of each of the activities that were rated. Cronbach's α for the math scale is .94, and for the literacy scale is .87.

The Emerging Academic Snapshot (EAS) used in the NCEDL study is also an observer-rated measure of children's classroom engagement that captures children's moment-to-moment activities (Ritchie, Howes, Kraft-Sayre, & Weiser, 2001). Observations were conducted during one or two days in the spring of the preschool year. The data collector observed each study child in 20-second interval "snapshots," followed by a 40-second coding period. The other three study children in the sampled classroom were then coded before coming back to observe the first child again, and this was repeated for the entire observation period. Children were coded with one of six mutually exclusive activity settings in each snapshot (basics, free choice, individual time, meals, small group, and whole group). The activity was also coded for early academic content area (aesthetics, fine motor skills, gross motor skills, letter and sound, mathematics, oral language development, read to, science, social studies, and writing). For example, to obtain the proportion of the day spent in math activities at the classroom level, coders took the average amount of time that each sample child was observed engaged in math activities divided by the total observation time. The last coded component of each snapshot is the type of teacher-child interaction (routine, minimal, simple, elaborated, scaffolding, and didactic). Kappas range from .70 to .87.

End-of-year teacher surveys were used in the HSIS, FACES 2003, and FACES 2009 studies to capture the different types of classroom activities. Teachers were asked how many times in the past week their class engaged in a specific literacy or math activity (shown in Appendices A.3–A.5). We used the teacher-reported items on the type and frequency of classroom literacy and math activities, converted into times per month by

taking the mean value of the answer category (e.g., never = 0; 1–2 times per week = 1.5), and multiplied by 4, following Claessens, Engel, and Curran (2013). We then standardized this measure to have a mean of 0 and standard deviation of 1. Prior research indicates that teacher survey instruments are valid for assessing quantity of instruction but not quality (Herman, Klein, & Abedi, 2000). FACES 2003 did not ask teachers about the quantity of math activities in the classroom, so this outcome was excluded from the analyses for this data set.

Child school readiness skills. Our analyses use multiple literacy, language, math, and social-emotional assessments that are considered valid and reliable, and are widely used within the field of child development. We examine children's skills in several outcome domains because a central tenet of the whole-child curricula model is that the experiences generated by the curricula cultivate all aspects of children's development. In each study, children were assessed at the beginning and end of their preschool year so that the baseline score can be used as a control variable. Note that we do not describe PCER's school readiness measures because we are unable to estimate our child-level analytic models using those data (see the "Analyses" further on).

Receptive language was measured by the Peabody Picture Vocabulary Test (PPVT; Dunn & Dunn, 1997) in each study, which focuses on children's ability to successfully point to the picture that most closely represents the word spoken to them by the test administrator. Reliability for the PPVT ranges from .92 to .98. Children's emergent literacy skills were also measured with the Letter Word (LW) subtest from the Woodcock-Johnson Psycho-Educational Battery-Revised III in the HSIS and FACES (WJ; Woodcock, McGrew, & Mather, 2001). In the LW test, the child is initially asked to identify letters and as the test progresses in difficulty, children are asked to read and pronounce written words correctly. This assessment measures children's ability to correctly recognize and sound out letters and sight words. Reliability is between .97 and .99 for preschool children. HSIS and FACES also included the WJ Spelling subtest. The Spelling subtest requires children to trace letters, write letters in upper- and lowercase, and to spell words, measuring early writing and spelling skills (Cronbach's $\alpha = .90$).

Children's general mathematical knowledge was assessed by the WJ Applied Problems (AP) subtest in all studies (Woodcock et al., 2001). The AP subtest examines early numeracy, and the child's ability to analyze and solve math problems. The reliability coefficient for the three- to five-year-old age group ranges from .92 to .94.

The Teacher-Child Rating Scale (TCRS; Hightower, 1986) was used to measure children's social and emotional skills in the NCEDL study. This is a behavioral rating scale that assesses children's social competence and problem behaviors. The Social Competence scale was computed as the mean of 20 items and had a Cronbach's α of .95. The Problem Behavior scale was computed as the mean of 18 items and had a Cronbach's α of .91. The HSIS study included the 28-item Behavior Problems Index (Zill, 1990). This is a parent report of problem behaviors related to emotional status, school behavior, and interpersonal relationships, with items drawn from several other child-behavior scales (e.g., Child Behavior Checklist). Items are rated on a three-point scale, and have a two-week test-retest reliability of .92. Problem behaviors and social skills were measured in the FACES studies using items from an abbreviated adaptation of the Personal Maturity Scale (Alexander, Entwisle, Blyth, & McAdoo, 1988), Child

Behavior Checklist for Preschool-Aged Children, Teacher Report (Achenbach, Edelbrock, & Howell, 1987), Behavior Problems Index (Zill, 1990), and the Social Skills Rating System (Gresham & Elliott, 1990).

Covariates. Each data set contains several child and parent characteristics that are included as control variables in our analyses. These include gender of child, race of child, mother or primary caregiver educational level and age, and family income. Data about these characteristics were collected via parent report during the preschool year. We also include children's baseline outcome assessments from the fall of the preschool year as covariates. In the NCEDL analyses we include an indicator for family poverty as a control, and in the HSIS analyses we include an indicator for teen mother because of the nature of these two samples (teen mother not reported in FACES). The classroom, teacher, and center covariates are teachers' education, race, and years of experience; classroom-level aggregates of children's race, gender, and parental education; whether the classroom is located in a public school or is a Head Start provider (PCER and NCEDL only); and an indicator for full day (available only in NCEDL and FACES 2009; collected at the center-level in the HSIS). Because PCER was an experimental study, we control for classroom treatment status to adjust for researcher involvement in curricular implementation.

Missing data. Rates of missingness on key study variables across all data sets range from 0% to 14%. The most substantial source of missingness was from curricula information because of teacher or director non-response. We used complete case analysis and compared the characteristics of children and teachers in classrooms with and without curricula information to assess whether the dropped cases differed systematically from the analysis sample. No consistent patterns of missingness emerged across the five data sets, but in three of the data sets teachers with a high school degree or below were less likely to report curricula information. This could bias our estimates of curricula use upward. We assume that data are missing at random (a function of other observable variables), which is plausible given our rich covariates, and also assume that the distribution of missing variables are jointly normal (Allison, 2002).

Analyses

We present an overview of the study hypotheses and analyses by research question in Table 2, indicating the data set in which each analysis was conducted. The Creative Curriculum serves as the reference category for both the classroom- and child-level outcome analyses because it was the most frequently used curricula in each data set, providing a common comparison group for all analyses.[1]

Research Question 1: Descriptive analyses of classroom activities. A first-order question in the investigation of preschool curricula and children's school readiness is whether differences exist in children's preschool classroom experiences by curricular status (i.e., published curriculum in use: yes or no). To answer this question, we first compare the available measures of classroom activities, quality, and other key classroom features (e.g., teachers' education, classroom-level aggregates of child characteristics) by

[1]Because no published curricula is not an option for Head Start centers under curricula mandates, it is not available in the three Head Start samples (HSIS, FACES 2003 and 2009). Analyses with "No curriculum" classrooms as the reference category are presented in Appendix C for NCEDL only.

curricular status in PCER and NCEDL using t tests of means or z tests of proportions. Because all Head Start programs require the use of curricula, HSIS and FACES cannot be used to examine differences in curricular status. We then compare the measures of math and literacy activities, quality, and other classroom features by curricular package (e.g., HighScope, Creative Curriculum, Scholastic, etc.) using ANOVA. For this set of analyses we use all five data sets.

We also test for differences in associations using ordinary least squares (OLS) regression, regressing each measure of math and literacy activities and quality on curricular status or curricular package and controlling for other classroom characteristics that influence the measurement classroom-level of processes and activities (e.g., classroom-level child characteristics) or that may affect implementation (e.g., teacher education), and conduct F tests to determine whether the set of curricular package coefficients jointly equal zero.[2] However, we also recognize that curricula, classroom characteristics and classroom processes may be jointly determined, and therefore controlling for these factors may complicate inference. Because this relative bias calculation is unclear (bias from measurement or implementation context versus confounding from simultaneity), we prefer the straightforward mean comparisons, and focus our results and discussion text on these analyses. Results from regressions of classroom processes on curricular status and curricular package indicators with covariates included are available from the authors.

Research Question 2: Grantee and state fixed effects analyses of child outcomes. Curricula are not randomly assigned to grantees, centers, teachers, or children. Unobserved or unmeasured characteristics may be associated with both curricula and children's outcomes, and thus we cannot causally determine whether a curriculum affects children's school readiness with observational data. To mitigate such bias, we test for associations between curricular status, curricular package, and child school readiness outcomes using two types of fixed effects. Fixed effects is an econometric technique that removes from the estimate of interest any context-specific and time-invariant observable or unobservable characteristics that may influence both the choice of curriculum and children's outcomes. These models compare the outcomes of children who share the same proximal (Head Start grantee) or distal (state policy context) environments. We also conduct F tests of the joint hypothesis of no differences among all curricular packages and children's outcomes to test for systematic variation. There were not enough states or grantees in the PCER sample with variation in curricular status to test for differences in outcomes. Because no common curricular reference group exists across states or grantees in the PCER study, we are also unable to test for differences in child outcomes by curricular package. In total, we examine relationships between curricula and children's outcomes in the HSIS, NCEDL, and FACES samples.

Curricular status state fixed effects models. We estimate the association between curricular status and children's outcomes in the NCEDL data set using state fixed effects models. This model compares children in preschool classrooms within the same state across classrooms who use a curricula package with those in classrooms who do not. We acknowledge that state fixed effects do not address classroom-level

[2]Our data do not include actual curricular implementation measures, such as coaching, and we recognize this as a study limitation.

selection bias, but within the constraints of our data, this approach mitigates bias from cross-state variation in policies, regulations, and funding streams affecting preschool quality and curricular requirements (Barnett et al., 2017; Gilliam & Ripple, 2004; Jenkins, 2014; Kirp, 2007; Pianta, Barnett, Burchinal, & Thornburg, 2009). The general form of this model is as follows:

$$\text{ChildOutcome}_{i} = b_0 + ß_1 \textit{CurricularStatus}_c + \text{X}(\textit{Child Controls}_i) + \text{Q}(\textit{Classroom Controls}_c) + \text{S}(\textit{States}_k) + e_{ic} \quad (1)$$

where ChildOutcome represents a child's (i) school readiness outcome (e.g., PPVT) at the end of the preschool year, *Child Controls* is a vector of child and family control variables, which also includes children's baseline skills assessment scores; *States* is a vector of indicators for each (k) of the states included in the study; and e represents the remaining sources of variation in children's school readiness from unaccounted factors. $ß_1$ is our coefficient of interest, representing the association between classroom curricular status and children's school readiness, indexed by classroom (c). We adjust for the clustered sample designs at the classroom level using Huber-White standard errors.

Because curricula are not randomly assigned, the interpretation of $ß_1$ (and A further on) must allow for the possibility that curricula will be picking up other classroom or center characteristics that are correlated with curricula. We attempt to minimize this problem by including a vector of appropriate teacher-, classroom-, and center-level controls, indicated by *Classroom Controls* (i.e., teacher's education, teacher's years of experience, and ECERS score).

Curricular package grantee fixed effects models. The analysis most robust to bias from unobserved center and classroom characteristics comes from the HSIS and FACES data, where we are able to estimate grantee fixed effects models. For example, in the HSIS data this method takes advantage of differences in classroom curricula within the grantee where families applied for, and were randomly assigned to receive, Head Start services at one of the centers operated by that grantee. In other words, this analysis allows us to compare the outcomes of children living in the same area who received Head Start services from the same grantee, reducing the possibility of omitted variables bias but not eliminating it. The general form of this model is as follows:

$$\text{ChildOutcome}_{ic} = b_0 + \text{A}(\textit{Curricula}_c) + \text{X}(\textit{Child Controls}_i) + \text{Q}(\textit{Classroom Controls}_c) + \text{G}(\textit{Grantee}_z) + e_{icz} \quad (2)$$

where *Curricula* is a vector of curriculum indicator variables, which vary by classroom; *Grantee* is a vector of indicators for each (z) of the Head Start grantees included in the study, and all other terms are identical to those shown in Equation 1. The coefficients in A are our estimates of interest because they represent the differential associations between each preschool curriculum and children's school readiness relative to the reference category. Of the 84 grantees in the HSIS, 62 (75%) had variation across classrooms in curricular package, with Creative Curriculum as the most common curriculum in use. For FACES 2003, 26 (41%) of the 63 grantees had variation across classrooms in curricular package, and 28 (47%) out of the 60 grantees had such variation in FACES 2009. In each of the samples we have 80% power to detect effect sizes of .20. We adjust for the clustered sample designs at the grantee level using Huber-White standard errors.

Curricular package state fixed effects models. Although we are unable to estimate a similar grantee fixed effects model for analyses by curricular package in NCEDL because of the difference in sampling and study designs, we estimate a state fixed effect model with the NCEDL data set. This model compares children in preschool classrooms within the same state across classrooms using different curricula, with Creative Curriculum as the reference group. This model replaces *Grantee* in Equation 2 with indicators for the states ($States_k$, as in Equation 1) included in the NCEDL study.

Meta-analysis of curricular package estimates. We use meta-analytic techniques to summarize the four sets of coefficients produced from the child outcome models of curricular packages. The meta-analysis treats the standardized regression coefficients for each curriculum package of Equation 2 as observations in a regression predicting children's school readiness outcomes at the end of preschool. We follow standard meta-analytic practices and weight each regression coefficient by the inverse of their variance (Hedges & Olkin, 1985).

Research Question 3: Consistency in classroom activities, quality, and teacher perceptions. We conduct several descriptive analyses to examine variation in classroom processes and activities across classrooms using the same curricular package. First, we create histograms of ECERS scores and the frequency of math and literacy activities for the two most commonly used curricula: Creative Curriculum and HighScope. We then overlay these data for the "no published curricula" classrooms on the same histograms to determine how classrooms *without* a published curriculum in use are distributed on classroom variables compared with classrooms *using* a published curriculum. We could not do the comparison overlay in the HSIS and FACES graphs because all Head Start classrooms are required to use a published curriculum, and therefore only conduct these graphical analyses with the PCER and NCEDL data sets.[3] In addition to the graphical analyses, we conduct Kolmogorov-Smirnov tests of the equality of distributions to determine if the distributions of classroom quality were significantly different.

We then descriptively examine responses to the available teacher survey items on classroom curricula aggregated by curricular package to better understand teachers' perspectives on their classroom curricula and the supports they receive to implement the curricula, and look for differences across curricula. We conduct these analyses in the data sets where such items were available (NCEDL, HSIS, FACES 2003 and 2009). Although these data do not capture implementation as assessed by an objective observer, they do provide a better sense of teachers' curriculum use, supports for implementation, and overall perspectives on their curriculum.

Results

Curricular Status and Curricular Package Differences in Classroom Activities and Quality

Curricular status. We computed descriptive statistics and *t* tests to assess whether having a curriculum in use makes a difference in the quality of children's preschool classroom

[3]The measurement scales in NCEDL and PCER are different from the HSIS, and so overlaying those distributions on the HSIS classrooms would be difficult to interpret.

experiences and their classroom's math and literacy activities in the PCER and NCEDL samples presented in Table 3. All Head Start classrooms use curricula, and therefore the HSIS and FACES data are omitted from the curricular status analysis. Here we discuss mean differences between classrooms with and without published curricula on math and literacy activities and quality scores. Mean comparisons of additional classroom characteristics by curricular status are shown in Appendix B. Regression-adjusted comparisons that control classroom characteristics are available from the authors.

PCER. The PCER results indicate that classrooms reporting use of a published curriculum have significantly more literacy and math activities and higher quality ratings from the ECERS (on both subscales) and Arnett Caregiver Interaction scales relative to classrooms where teachers report using no published curriculum. In regression analyses controlling for a comprehensive set of potential confounds (teacher characteristics and classroom-level aggregates of children's race, gender, and parental education), these differences remain but the coefficients do not reach significance.

NCEDL. Descriptive analyses in the NCEDL sample reveal that classrooms using a published curriculum score higher on the total ECERS score and in the Provisions for Learning ECERS factor compared with classrooms not using a curriculum. No significant differences emerge by curricular status in the amount of classroom math and literacy learning activities or in the two CLASS subscales. Regression models including the set of control variables confirm these results.

Curricular package. Table 4 presents descriptive statistics and ANOVAs for each data set to examine differences by curricular package in the means and proportions of classroom activities and quality. Counter to our hypothesis of no differences between whole-child curricular packages, there were significant differences across curricular packages in both the quantity of math activities and overall classroom quality based on the ECERS, Arnett, and CLASS scales in all five samples. Other significant differences emerged between curricular packages in each data set, but without a clear rank ordering of packages in terms of their allocation of literacy and math activities or superior quality.

In PCER, Creative Curriculum had the most math activities, DLM Express had the most literacy activities and highest ECERS scores, and both packages also had the highest Arnett Caregiver Interaction scores. NCEDL revealed the fewest differences between packages, with HighScope and the "other published curriculum" category demonstrating the highest quality on ECERS. HSIS results indicate that HighScope classrooms have the highest ECERS ratings and that High Reach have the most math activities and highest Arnett scores. FACES 2003 results favored the "other published curriculum" category on all ECERS ratings. FACES 2009 reveal Scholastic classrooms implementing the most math activities, while High Reach produced the most literacy activities. Overall ECERS quality was highest in Creative Classrooms, but HighScope had the highest language/interactions subscale score, High Reach had the highest Provisions for Learning subscale score, and CLASS subscale scores also favored Creative Curriculum and High Reach.

Regressions of classroom activities on indicators for curricular package controlling for other classroom characteristics are available from the authors. As a complement to the ANOVAs, this analysis allowed us to directly compare each curriculum with the reference category (Creative Curriculum) while controlling for other classroom

characteristics. Results are very similar to the patterns in Table 4. We tested for differences overall among the curricular packages with joint *F* tests and reject the null hypothesis of no differences in 4 of the 14 estimated models, providing mixed evidence of the unique contribution of curricular packages to classroom processes. Overall, these descriptive analyses did not reveal a top performer across the five data sets.

Curricular Status and Curricular Package Differences in Child School Readiness

Curricular Status.

State fixed effects models testing for differences in children's school readiness in the spring of their preschool year by curricular status in NCEDL are presented in Appendix C. We find no significant differences in children's math, literacy, or social skills depending on whether the classroom used a published curriculum. However, teachers reported significantly fewer problem behaviors in classrooms where a curricular package was used.

Curricular package. Table 5 presents the results for models examining differences in children's outcomes in the spring of their preschool year by curricular package. The reference group is Creative Curriculum in each data set. All outcomes are in standard deviation (SD) units.

HSIS. After controlling for Head Start grantee with grantee fixed effects—and thus as many unobserved grantee-level factors as possible—results suggest that children in Head Start classrooms using the Scholastic curriculum outperform children in other classrooms operated by that grantee using the Creative Curriculum. We detect 0.25 SD difference in children's outcomes between Scholastic and Creative Curriculum classrooms on the WJ-Applied Problems and Letter Word subtests. Children's WJ-Spelling subtest scores were significantly lower in classrooms using Creative Curriculum compared with HighScope and the "other curricular packages" set of classrooms. Children in classrooms using the HighScope curriculum also scored 0.18 SD higher on the WJ-Applied Problems subtest compared with children in Creative Curriculum classrooms. Children in classrooms using High Reach scored significantly worse on PPVT scores relative to Creative Curriculum. *F*-test results indicate that there are overall differences in curricular package associations with children's WJ-Applied Problems and Spelling subtests, marginal differences with PPVT, and no differences with WJ-Letter Word subscale scores and behavior problems.

FACES 2003. Grantee fixed effects models for the FACES 2003 data set indicate very few differences in children's outcomes at the end of preschool by curricular package. Children in classrooms using "other" published curricula scored 0.34 SD lower on social skills compared with children in Creative Curriculum classrooms. *F*-test results indicate that there are marginal differences in curricular package associations with social skills, and no differences with PPVT, WJ subscale scores, or behavior problems.

FACES 2009. Grantee fixed effects models using the FACES 2009 data set show that children in classrooms using High Reach had substantially lower scores on the PPVT and the WJ-Applied Problems subtest compared with children in classrooms using Creative Curriculum (–0.33, –0.18 SD), and marginally significantly lower social skills (–0.29 SD). *F*-test results indicate that there are marginal differences with PPVT, and no differences with WJ subscale scores, behavior problems, or social skills.

Table 5. Fixed effects results for associations between classroom curricula and children's school readiness.

HSIS[t]	PPVT	WJAP	WJLW	WJSP	Behavior Problems	
HighScope	0.02	0.18*	0.09	0.27*	0.12	
	(0.08)	(0.09)	(0.12)	(0.11)	(0.15)	
High Reach	−0.19*	0.19	0.26	0.24	−0.18	
	(0.09)	(0.19)	(0.20)	(0.15)	(0.23)	
Scholastic	0.11	0.25*	0.25*	0.47	0.10	
	(0.07)	(0.11)	(0.12)	(0.29)	(0.15)	
Other Published Curriculum	0.03	0.10	0.04	0.27*	0.20	
	(0.08)	(0.08)	(0.09)	(0.12)	(0.13)	
F test (*p* value)	0.10	0.00	0.22	0.03	0.63	
Observations	1709	1700	1711	1709	1654	
FACES 2003[t]	PPVT	WJAP	WJLW	WJSP	Behavior Problems	Social Skills
HighScope	−0.24	−0.09	−0.04	−0.04	−0.08	−0.26
	(0.16)	(0.18)	(0.17)	(0.13)	(0.21)	(0.22)
High Reach	−0.32	−0.08	−0.10	0.29	−0.43	−0.25
	(0.23)	(0.20)	(0.19)	(0.37)	(0.27)	(0.45)
Other Published Curriculum	−0.01	−0.11	−0.19	−0.17	0.01	−0.34*
	(0.08)	(0.09)	(0.14)	(0.12)	(0.13)	(0.15)
F test (*p* value)	0.52	0.57	0.68	0.43	0.60	0.10
Observations	1637	1631	1628	1565	1787	1777
FACES 2009[t]	PPVT	WJAP	WJLW	WJSP	Behavior Problems	Social Skills
HighScope	−0.05	−0.04	0.18	−0.01	−0.27	−0.06
	(0.05)	(0.09)	(0.12)	(0.06)	(0.25)	(0.26)
High Reach	−0.33**	−0.18*	−0.11	−0.24	0.09	−0.29+
	(0.10)	(0.08)	(0.13)	(0.16)	(0.13)	(0.17)
Scholastic	−0.13	0.04	−0.04	−0.02	−0.29	−0.09
	(0.16)	(0.08)	(0.08)	(0.15)	(0.18)	(0.23)
Other Published Curriculum	0.04	0.12+	0.04	0.06	−0.15	0.16
	(0.06)	(0.07)	(0.09)	(0.07)	(0.10)	(0.11)

(*continued*)

Table 5. Continued.

HSIS[t]	PPVT	WJAP	WJLW	WJSP	Behavior Problems	
F test (*p* value)	0.09	0.11	0.74	0.72	0.45	0.24
Observations	2611	2397	2401	2477	2691	2736
NCEDL	PPVT	WJAP			HT Problem Behaviors	HT Competency
HighScope	0.02	–0.04			0.08	0.15
	(0.09)	(0.13)			(0.12)	(0.13)
Other Published Curriculum	0.09	–0.02			–0.07	0.16
	(0.11)	(0.14)			(0.14)	(0.16)
No Published Curriculum	0.11	0.03			0.33*	0.07
	(0.09)	(0.14)			(0.15)	(0.16)
F test (*p* value)	0.54	0.68			0.10	0.59
Observations	398	394			450	452

Note. [t]Program/grantee fixed effects. [^]State fixed effects. Clustered standard errors in parentheses. Creative Curriculum is reference group. All outcomes are in standard deviation units. Sample includes treated children from the HSIS experiment dataset only. All models include: child race, gender, age, and baseline assessment score for each outcome, mother's education, classroom quality (ECERS), teacher controls (teacher's education, race, and years of experience). An indicator for full-day preschool status was included in the NCEDL and FACES 2009 analysis (information not collected in FACES 2003 and only collected at center-level in HSIS). Teen mom status was included in HSIS analyses (does not exist in NCEDL or FACES). An indicator for income under 150% of the poverty line was included in the NCEDL analyses (all HSIS and FACES participants were considered poor). PPVT: Peabody Picture Vocabulary Test; WJAP: Woodcock Johnson Applied Problems; WJLW: Woodcock Johnson Letter Word; WJSP: Woodcock Johnson Spelling; HT: Hightower. For all Problem Behaviors scores, a higher score indicates a more serious problem. The *F*-test statistic indicates whether the curricular package indicators are jointly equal to 0. $+p<.10$ $^{*}p<.05$ $^{**}p<.01$.

Table 6. Meta-analytic regression results from Table 4 coefficients.

	PPVT	WJAP	WJLW	WJSP	Behavior Problems	Social Skills/Competency
High Scope	–0.03	0.03	0.10	0.04	0.03	0.03
	(0.07)	(0.12)	(0.10)	(0.14)	(0.14)	(0.21)
High Reach	–0.26**	–0.12	–0.03	0.04	–0.04	–0.29**
	(0.08)	(0.15)	(0.19)	(0.30)	(0.24)	(0.12)
Scholastic	0.07	0.11	0.05	0.08	–0.06	–
	(0.12)	(0.14)	(0.19)	(0.28)	(0.27)	
Other Published Curriculum	0.03	0.05	0.00	0.06	–0.02	0.03
	(0.03)	(0.11)	(0.11)	(0.17)	(0.15)	(0.27)
No Published Curriculum	–	–	–	–	–	–
Observations	13	13	11	11	13	8

Notes. All coefficients used in the analyses come from individual study regressions that include full controls from Table 4. Standard errors are corrected for within-study clustering using Huber-White methods. Regression coefficient observations are weighted by the inverse of their variances. Creative Curriculum is the omitted reference group. Dashes indicate meta-analytic regression results not available because of the small number of initial regressions. PPVT = Peabody Picture Vocabulary Test, WJAP = Woodcock Johnson Applied Problems, WJLW = Woodcock Johnson Letter Word, WJSP = Woodcock Johnson Spelling. For all Behavior Problems scores, a higher score indicates a more serious problem. $+p < .10$. $^{*}p < .05$. $^{**}p < .01$.

NCEDL. State fixed effects models in the NCEDL data set indicate that children in classrooms with no published curriculum in use had higher problem behavior (0.33 SD) scores relative to Creative Curriculum classrooms at the end of the preschool year, which corresponds with the RQ1 finding that classrooms with no curriculum had higher levels of problem behaviors. No other significant differences in children's outcomes emerged. *F*-test results indicate that there are marginal differences on the behavior problems measure, and no differences with PPVT, WJ-Applied Problems subtest, or the social competency subscale.

Meta-analyses. We summarize our findings with a meta-analysis of the 74 coefficients drawn from regressions estimating the relationship between curricula packages and children's outcomes (from Table 5), with results shown in Table 6. Because we have as few as eight observations in the meta-analytic regression for each outcome, we have limited statistical power to confidently detect statistically significant, meaningful results. As such, we view these analyses as exploratory; results should be interpreted with caution. Overall, the meta-analytic regressions show that the majority of the curricular packages in our sample are not differentially associated with children's school readiness at the end of preschool. Results for High Reach indicate that children in those classrooms had scores substantially lower on the PPVT (–.26 SD) and on social skills (–.29 SD) compared with children in classrooms using Creative Curriculum.

Variation in the Implementation of Curricula

To examine the variability in classroom experiences across classrooms implementing the *same* curriculum, we present histograms of ECERS scores and the frequency of literacy and math activities for the two most popular curricula, Creative Curriculum and HighScope. Figure 1a shows the distributions of ECERS scores, and of math and literacy activities in NCEDL (left) and PCER (right) for Creative Curriculum classrooms; Figure 1b shows the same distributions for HighScope classrooms. Each

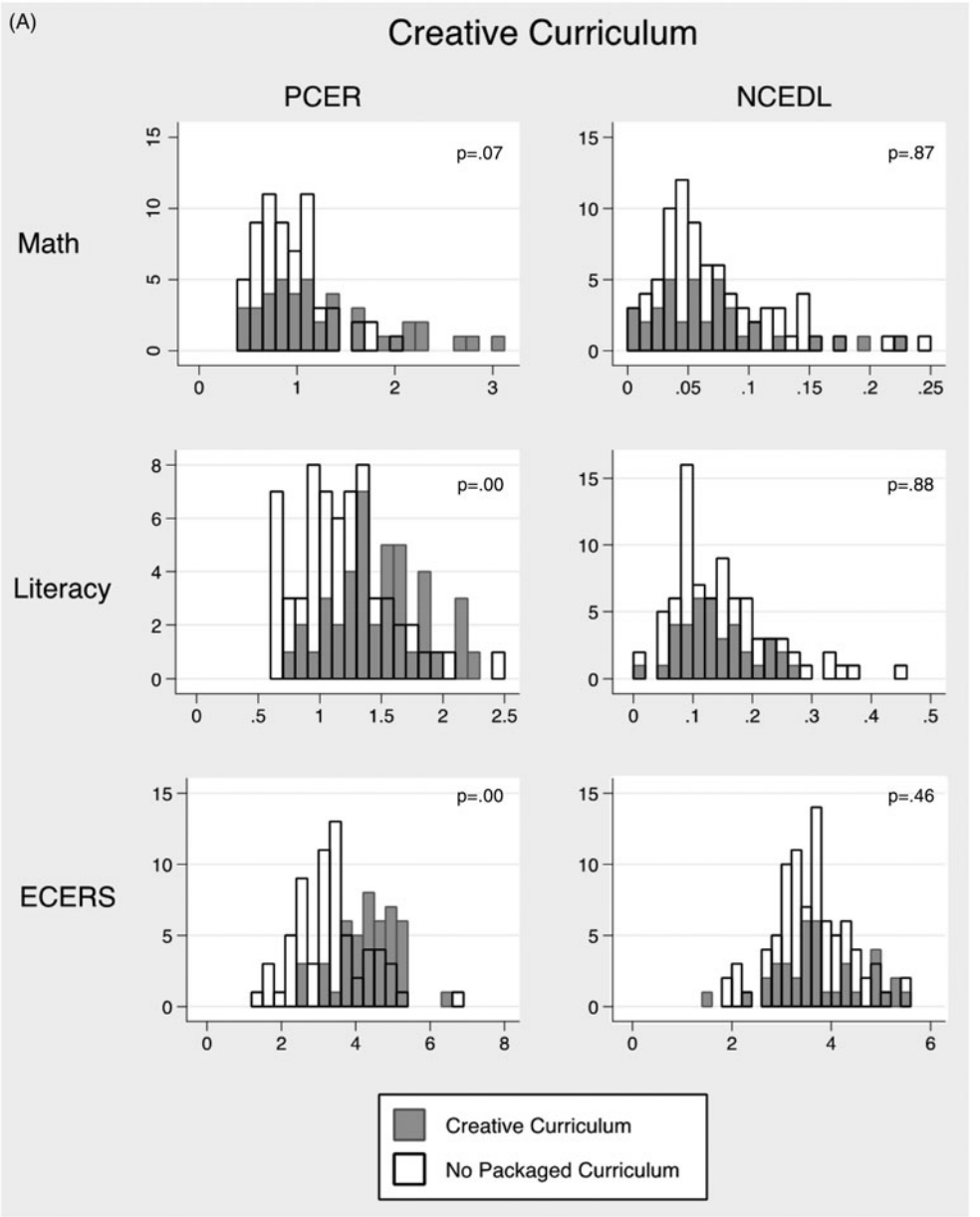

Figure 1. Histograms of classroom quality and classroom activities in Creative Curriculum and High Scope classrooms in the PCER and NCEDL studies.
(A) Creative Curriculum
(B) High Scope Curriculum
Notes: Bins are comprised of classroom-level observations. ECERS, Math, and Literacy activities measures are in raw scale form (X-axis labels are omitted). ECERS scale ranges from 0 to 7, Math and literacy activities are shown as proportion of day in NCEDL (EAS Snapshot), and from a 0 to 3 scale in PCER (TBRS). *p* values from Kolmogorov-Smirnov tests for the equality of the classroom distributions between the focal curriculum (Creative Curriculum or HighScope) and No Packaged (published) Curriculum classrooms are displayed in each graph, where $p < .05$ indicates significantly different distributions of the classroom measure between the two groups. See text for more detail.

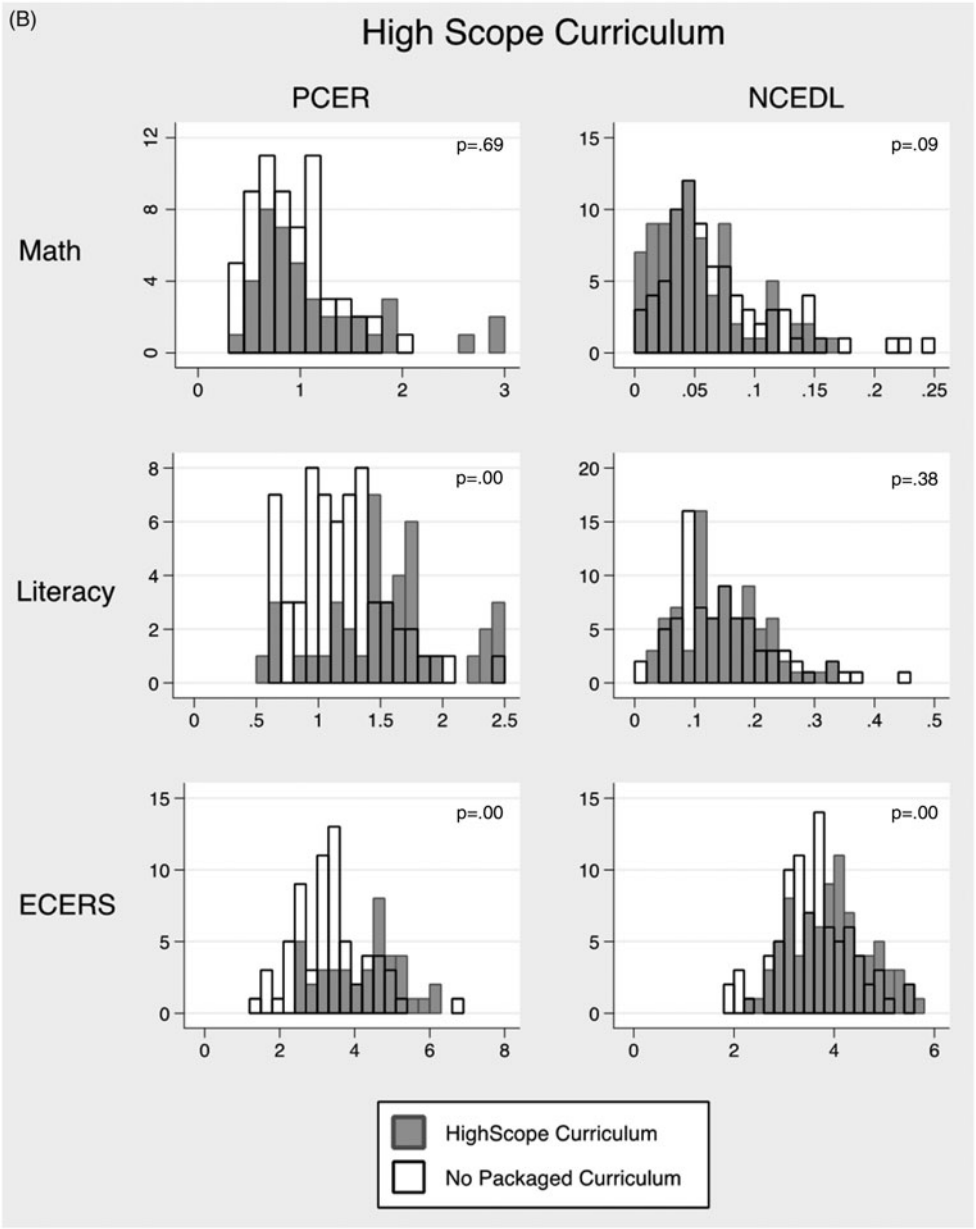

Figure 1. Continued

measure is in its original scale (i.e., not standardized). Overlaid on these graphs are the distributions for classrooms that *do not* report using a published curriculum, for comparison.

The most striking and consistent feature of these graphs is that classrooms using the same curriculum vary widely with respect to their overall quality and learning activities. Results are mixed as to whether the distributions of activities and quality differ in

classrooms with and without published curricula. Appearing on the upper right-hand side of each histogram is the p value from a Kolmogorov-Smirnov test of the equality of the distributions between the focal curriculum classrooms and the no published curriculum classrooms, where the null hypothesis is that the two distributions are equal ($p < .05 =$ significant differences between distributions). Of the 12 distributional tests (three classroom outcomes $\times$ two datasets $\times$ two curricula), half of the comparisons were not significant, indicating similar distributions, and half of the comparisons were significant, favoring the curricula group. These tests reveal mixed and inconsistent statistical support for differences in classrooms with and without a curriculum in use, similar to the findings from our classroom mean comparisons (Appendix B). In PCER, the distributions of both Creative Curriculum and HighScope classrooms were not statistically distinguishable from classrooms using no published curriculum with respect to their math activities. In NCEDL, the distributions of each measure were not statistically distinguishable between Creative Curriculum and no published curriculum classrooms. The distributions of math and literacy activities were also the same between HighScope and "no published curriculum" classrooms. The comparisons indicating significant differences in the distributions by curricular status came primarily from the PCER data, showing the curricula-in-use classroom distributions lying to the right of the "no curricula" classroom distributions.

These illustrations also help to explain our largely null findings thus far; substantial variation in the learning experiences *within* the population of classrooms that report using HighScope or Creative Curriculum would not likely lead to systematic differences in children's outcomes *across* curricular packages. Histograms from the HSIS and FACES samples without the "no published curriculum" overlays are available in Appendix D. Similarly, these figures indicate substantial dispersion or variation in the distribution of activities and quality.

Examining teacher survey items on curricula. To get a better sense of teachers' perspectives on their classroom curricula and the supports they receive to implement curricula, we descriptively compared teacher survey responses to items asking about their classroom curricula by classroom curricula package. Shown in Appendices A.2–A.5, all items are indicator variables and equal 1 if the teacher responded "yes" to the question prompt.

In the NCEDL and HSIS, there were no differences by curricular package in teacher's report of receiving training in the curriculum. The HSIS also included items on teachers' attitudes toward the curriculum. Across all curricula, HSIS teachers reported high agreement (>90%) with such items as liking the curriculum, ease of use, leaving room for teacher creativity, and adequacy of materials to implement the curricula; there were no significant differences across curricular packages. The FACES 2003 included a very similar set of questions to the HSIS, and also indicted strong agreement with survey items (>87%). Although the ANOVA tests indicated differences in agreement across curricular packages, these differences were very small in magnitude (e.g., ranging between 93% and 98% of teachers agreeing). The curricula items included in the FACES 2009 teacher survey focused on the types of support teachers received in using the curriculum. Agreement with these support-related items was lower than agreement with positive attitudes toward the curricula in the

other data sets (ranging between 45% and 88%), with significant differences across curricular packages.

Across the four samples, it appears that preschool teachers receive some initial training on their classroom curricula and that they like the curriculum they use, but they do not consistently receive continual support in implementing the curriculum.

Discussion

Our study comprehensively examined the role of curricula in center-based preschool environments and their relation to children's academic and social-emotional development in five different preschool studies. These five samples captured the authentic preschool experiences of a diverse set of low-income children attending publicly funded state pre-K and Head Start programs and other privately funded preschool centers. Specifically, our research questions were: (1) To what extent do classroom activities and quality ratings vary by whether a published curriculum is in use, and in classrooms that do use a published curriculum, do activities vary by the specific curricular package? (2) To what extent is having a published curriculum in use in a preschool classroom associated with children's academic and social-emotional school readiness, and do children's readiness vary by the specific curricular package? (3) To what extent are the classroom activities, overall classroom quality ratings, and teacher's attitudes and perceptions of curriculum consistent among classrooms using the same, or different, curricular packages? In two data sets, classroom activities were measured with observer-rated protocols (PCER, NCEDL), and in the other three, teachers' reports of classroom activities (HSIS, FACES 2003 and 2009). We do not equate use of a curriculum with fidelity of implementation of a curriculum, and consider what we observe as the de facto use of curriculum in preschool programs operating at scale. Our goal was simply to describe the observed patterns in extant data and glean as much information as possible given the developmental and policy relevance of our research questions.

Taken together, the findings from our study indicate that there are few distinguishing characteristics about the whole-child curricula most commonly used in preschool programs. We found some evidence that implementing a curricular package was associated with higher quality scores and more frequent math and literacy activities compared with classrooms with no published curriculum in use. However, for classrooms that reported using HighScope and Creative Curriculum, the two most commonly used curricula, classroom literacy and math activities and ECERS scores varied as widely within the population of classrooms using each package as it did across the population of preschool classrooms where teachers report not using any published curriculum. The variation within a curriculum that we observed may be because of the fact that these curricula are typically not scripted or manualized for teachers—which would provide consistency in processes—or that teachers do not receive implementation support (Weiland, McCormick, Mattera, Maier, & Morris, 2018). Whole-child curricula do not rely on detailed teacher scripts; rather, they are designed to be flexible to cater to children's rapidly changing interests. This often leaves teachers on their own to interpret how they should implement the curriculum. Indeed, our descriptive analysis of teacher survey

items indicated that teachers receive some initial training on their curricula, but they do not receive continual support in implementation.

One curriculum used in Head Start classrooms (Scholastic) stood out by having more math activities but significantly lower classroom quality scores than Creative Curriculum. Results from the HSIS were suggestive of children in Scholastic classrooms having stronger academic school readiness, but we did not see this in the FACES 2009 data set, which included Scholastic. In both the FACES 2009 and HSIS data sets, children in High Reach classrooms scored lower on both academic and social skills outcomes than children in Creative Curriculum classrooms, and had marginally significantly lower-quality scores. The findings from our meta-analyses confirmed these overall patterns.

Our findings, while primarily descriptive in nature, beg an extremely important and policy-relevant question: What do current curricular investments in early childhood policy yield for children's development and well-being? We do not find evidence to support Creative Curriculum's preeminence in Head Start programs nationwide (between 40% and 52% of classrooms, based on our calculations and that of prior studies; Hulsey et al., 2011; Jenkins & Duncan, 2017), corroborating the What Works Clearinghouse rating of "No Evidence." Nor do we find support for other curricular packages with the exception of Scholastic in one data set. To be clear, we are not suggesting that whole-child approaches are without value, only that these curricula are supported by policies without rigorous evaluation against what teachers are otherwise doing. The average per-classroom cost of a curriculum is approximately $2,000, and thus careful scrutiny of these requirements is imperative. We caution against interpreting these results as causal and instead suggest that they be a starting point for future research and policy discussions.

We also consider the perspective of curricula publishers and developers in interpreting our results. Our data represent the business-as-usual educational environments of low-income preschoolers. They do not necessarily represent classroom experiences when curricula are implemented with high fidelity (which we cannot measure in this study), and with developer-specified professional development; in other words, our analyses do not represent tests of curricular efficacy, but represent the business-as-usual experiences of children in public and private preschool programs. Indeed, teachers may report "using" a curriculum that they only reference on occasion, or not at all. The aim of our study was to understand the implications of policy-mandated curricula, and thus our data represent the de facto educational environments for children attending preschool during 2001–2009. The policy requirements would necessarily need to change and include greater professional development and other supports to implement curricular packages with high fidelity at scale. Still, none of the curricula under study have evidence of efficacy under ideal conditions, so this criticism on its own falls short of how we need to think about curricular choices in public preschool programs for low-income children.

Interestingly, our findings do indicate some classroom quality and math and literacy activity differences between classrooms with and without any curriculum in use, yet we did not find that curricular status was associated with better child outcomes. This raises an important point about why improvements in quality do not translate into

improvements in children's school readiness, a finding shared by other studies of preschool curricula (Jenkins et al., 2018; PCER, 2008) and of quality more generally (Auger, Farkas, Burchinal, Duncan, & Vandell, 2014; Gordon et al., 2017). One possibility is that curricula do not boost classroom quality enough to affect child development, meaning that quality does not reach a sufficient threshold (Burchinal et al., 2016). This is not particularly surprising, given the emerging evidence showing that intensive training and ongoing coaching are essential to improving both quality and child outcomes in curricula interventions (Davidson, Fields, & Yang, 2009; Weiland et al., 2018); curricula use alone is unlikely to lead to high levels of classroom quality without it. Therefore, future research should build on our understanding of what types of support are most beneficial, the cost-effectiveness of those supports, and whether different models, such as expert training sessions or train-the-trainer programs, are equally effective for ensuring consistently high-quality classroom experiences and implementation fidelity.

Still, it is the correct combination of both curriculum *and* professional development that are key for policy makers to improve preschool at scale. Some of the most encouraging results come from studies of content-specific curriculum coupled with both strong teacher supports and continual monitoring of children's progress that, in combination, are important for improving preschool programs (Phillips et al., 2017; Yoshikawa et al., 2013). Encouragingly, the Office of Planning, Research and Evaluation (OPRE) is funding a project to do just this—examine the conditions and supports necessary to implement both whole-child and content-specific curricula in Head Start, child care, and public pre-K centers that lead to improved classroom quality and child outcomes (Office of Planning Research and Evaluation, 2016). Recently, Weiland and colleagues (2018) conducted a detailed examination of the factors related to preschool curricula implementation and professional development. They identified six key features that characterize successful implementation: a focus on instructional content, inclusion of highly detailed teacher scripts, incorporation of teacher voice, time for planning, use of real-time data, and early childhood training for administrators. The current Head Start policy assumes that using a research-based curriculum leads to better classroom environments and child outcomes. However, this is unlikely to occur if curricula are not scripted, and if policies do not provide professional supports for both teachers and administrators, or provide teachers with the ability to adapt, plan, and understand children's progress. More empirical work on each of these elements, both separately and in combination, are clear next steps for the early learning field.

Another direction for future research is the study of specific classroom activities most strongly associated with children's development, along with a content examination of curricula to examine which packages promote the most beneficial activities. Content analyses could also illuminate the extent to which curricula are aligned with early learning standards set by states and national organization such as NAEYC, akin to studies conducted with elementary and secondary school curricula (Polikoff, 2015; Porter, 2002; Schmidt, Wang, & McKnight, 2005).

A strength of our article is replication across five different preschool samples. However, this also means that the unique components of each data set restricted us from universally conducting the same analyses. We recognize that our measures of curricular activities and quality are limited and do not capture the full set of preschool

classroom experiences shaped by curricular packages. Note that many of the PCER classrooms implementing randomly assigned curricula had study administrators in the classroom providing professional development to help teachers implement the curriculum. Although we control for treatment status with the PCER data, these additional supports may not generalize to other preschool classrooms. Our meta-analyses were underpowered to detect differences across the samples included here, and we consider these results as exploratory. We also acknowledge that because our study data sets capture classroom practice from 2001 to 2009, this may limit the relevance of our findings to the current context. Although the patterns of curricula use observed in our data sets match those seen in a recent national sample, additional work of this nature, as more recent sample data become available (that include the key data elements used here), is needed. Developing this evidence base will provide a deeper understanding of factors that may make preschool effective for low-income children.

Acknowledgments

We would like to thank Douglas Clements, Kenneth Dodge, Greg Duncan, George Farkas, Dale Farran, Pamela Morris, Anna Gassman-Pines, Sean Reardon, and Deborah Vandell for helpful comments on prior drafts.

Funding

We are grateful to several institutions that provided generous support for our project. The Institute of Education Sciences (IES) supported this work through grant numbers R305B120013 (UC Irvine) and R305B170002 (UVA). Research reported in this publication was also supported by the Eunice Kennedy Shriver National Institute of Child Health & Human Development of the National Institutes of Health under award number P01HD065704 (UC Irvine). The Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services supported this work under award number 90YE0184 (UVA).

Disclosures

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, the National Institutes of Health, the Office of Planning, Research and Evaluation, the Administration for Children and Families, or the U.S. Department of Health and Human Services.

ORCID

Jade Marcus Jenkins http://orcid.org/0000-0002-2000-3087
Tutrang Nguyen http://orcid.org/0000-0002-6741-6022

ARTICLE HISTORY

Received 15 May 2018
Revised 15 May 2019
Accepted 27 May 2019

References

Achenbach, T. M., Edelbrock, C., & Howell, C. T. (1987). Empirically based assessment of the behavioral/emotional problems of 2- and 3- year-old children. *Journal of Abnormal Child Psychology*, *15*(4), 629–650. doi:10.1007/BF00917246

Alexander, K. L., Entwisle, D. R., Blyth, D. A., & McAdoo, H. P. (1988). Achievement in the first 2 years of school: Patterns and processes. *Monographs of the Society for Research in Child Development*, *53*(2), i–157. doi:10.2307/1166081

Allison, P. D. (2002). *Missing data*. Thousand Oaks, CA: Sage.

Auger, A., Farkas, G., Burchinal, M. R., Duncan, G. J., & Vandell, D. L. (2014). Preschool center care quality effects on academic achievement: An instrumental variables analysis. *Developmental Psychology*, *50*(12), 2559–2571. doi:10.1037/a0037995

Auger, A., Karoly, L., & Schwartz, H. L. (2015). *Evaluation of Delaware stars for early success: Year 2 report*. Santa Monica, CA: RAND Corporation.

Barnett, W. S. (2011). Effectiveness of early educational intervention. *Science*, *333*(6045), 975–978. doi:10.1126/science.1204534

Barnett, W. S., Carolan, M. E., Squires, J. H., Brown, K. C., & Horowitz, M. (2015). *The state of preschool 2014*. New Brunswick, NJ: National Institute for Early Education Research.

Barnett, W. S., Friedman-Krauss, A. H., Weisenfeld, G. G., Horowitz, M., Kasmin, R., & Squires, J. H. (2017). *The state of preschool 2016*. New Brunswick, NJ: National Institute for Early Education Research.

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

Bierman, K. L., Domitrovich, C. E., Nix, R. L., Gest, S. D., Welsh, J. A., Greenberg, M. T., … Gill, S. (2008). Promoting academic and social-emotional school readiness: The Head Start REDI Program. *Child Development*, *79*(6), 1802–1817. doi:10.1111/j.1467-8624.2008.01227.x

Bloom, H. S., & Weiland, C. (2015). *Quantifying variation in Head Start effects on young children's cognitive and socio-emotional skills using data from the National Head Start Impact Study*. New York, NY: MDRC.

Burchinal, M. R., Xue, Y., Auger, A., Tien, H.-C., Mashburn, A., Peisner-Feinberg, E., … Tarullo, L. (2016). Chapter III. Testing for quality thresholds and features in early care and education. *Monographs of the Society for Research in Child Development*, *81*(2), 46–63. doi:10.1111/mono.12238

Claessens, A., Engel, M., & Curran, F. C. (2013). Academic content, student learning, and the persistence of preschool effects. *American Educational Research Journal*, *51*(2), 403–434. doi:10.3102/0002831213513634

Clements, D. H. (2007). Curriculum research: Toward a framework for "research-based curricula". *Journal for Research in Mathematics Education*, *38*(1), 35–70.

Clements, D. H., & Sarama, J. (2007). Effects of a preschool mathematics curriculum: Summative research on the Building Blocks project. *Journal for Research in Mathematics Education*, *38*(2), 136–163.

Clements, D. H., & Sarama, J. (2008). Experimental evaluation of the effects of a research-based preschool mathematics curriculum. *American Educational Research Journal*, *45*(2), 443–494. doi:10.3102/0002831207312908

Clifford, R. M., Barbarin, O. A., Chang, F., Early, D. M., Bryant, D., Howes, C., … Pianta, R. (2005). What is pre-kindergarten? Characteristics of public pre-kindergarten programs. *Applied Developmental Science*, *9*(3), 126–143. doi:10.1207/s1532480xads0903_1

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

Dahlin, M., & Squires, J. H. (2016). *State pre-K approved curricula*. New Brunswick, NJ: National Institute for Early Education Research. Retrieved from http://ceelo.org/wp-content/uploads/2016/05/ceelo_listserv_summary_approved_curriculum_web.pdf

Davidson, M. R., Fields, M. K., & Yang, J. (2009). A randomized trial study of a preschool literacy curriculum: The importance of implementation. *Journal of Research on Educational Effectiveness*, *2*(3), 177–208. doi:10.1080/19345740902770028

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

Diamond, A. (2010). The evidence base for improving school outcomes by addressing the whole child and by addressing skills and attitudes, not just content. *Early Education and Development*, *21*(5), 780–793. doi:10.1080/10409289.2010.514522

Diamond, A., Barnett, W. S., Thomas, J., & Munro, S. (2007). Preschool program improves cognitive control. *Science*, *318*(5855), 1387–1388. doi:10.1126/science.1151148

Dotterer, A. M., Burchinal, M. R., Bryant, D., Early, D. M., & Pianta, R. (2009). Comparing universal and targeted prekindergarten programs. In R. Pianta & C. Howes (Eds.), *The promise of pre-k* (pp. 65–78). Baltimore, MD: Paul H. Brookes Publishing Co.

Duncan, G. J., & Magnuson, K. (2013). Investing in preschool programs. *The Journal of Economic Perspectives*, *27*(2), 109–132. doi:10.1257/jep.27.2.109

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

Early, D. M., Barbarin, O. A., Bryant, D., Burchinal, M. R., Chang, F., Clifford, R. M., … Kraft-Sayre, M. (2005). Pre-kindergarten in eleven states: NCEDL's multi-state study of pre-kindergarten and state-wide early educational programs (SWEEP) study. In National Institute for Early Education Research, Washington, DC.

Elkind, D. (2007). *The hurried child: Growing up too fast too soon*. Cambridge, MA: Da Capo Press.

Fantuzzo, J. W., Gadsden, V. L., & McDermott, P. A. (2011). An integrated curriculum to improve mathematics, language, and literacy for Head Start children. *American Educational Research Journal*, *48*(3), 763–793. doi:10.3102/0002831210385446

Gilliam, W. S., & Ripple, C. H. (2004). What can be learned from state-funded preschool initiatives? A data-based approach to the Head Start devolution debate. In E. F. Zigler & S. J. Styfco (Eds.), *The head start debates* (pp. 477–497). Baltimore, MD: Brookes Publishing.

Goffin, S. G., & Wilson, C. (1994). *Curriculum models and early childhood education: Appraising the relationship*. New York, NY: Merrill.

Gordon, R. A., Hofer, K. G., Aloe, A. M., Wilson, S. J., Peng, F., Gaur, D., & Lambouths III, D. (2017). *Preschool quality and children's vocabulary: A meta-analytic study of heterogeneity and moderation of effects*. Chicago, IL.

Gormley, W. T., Phillips, D., & Gayer, T. (2008). Preschool programs can boost school readiness. *Science*, *320*(5884), 1723–1724. doi:10.1126/science.1156019

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

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

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando, FL: Academic Press.

Herman, J. L., Klein, D. C. D., & Abedi, J. (2000). Assessing students' opportunity to learn: Teacher and student perspectives. *Educational measurement: Issues and practice*, *19*(4), 16–24. doi:10.1111/j.1745-3992.2000.tb00042.x

Hightower, A. D. (1986). The teacher–child rating scale: A brief objective measure of elementary children's school problem behaviors and competencies. *School Psychology Review*, *15*(3), 393–409.

Hulsey, L. K., Aikens, N., Kopack, A., West, J., Moiduddin, E., & Tarullo, L. B. (2011). *Head Start children, families, and programs: Present and past data from FACES* (OPRE Report 2011-33a). 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, E., & Steuerle, C. E. (2015). *Kids share 2015: Report on federal expenditures on children in 2014 and future projections*. Washington, DC: Urban Institute.

Jenkins, J. M. (2014). Early childhood development as economic development: Considerations for state-level policy innovation and experimentation. *Economic Development Quarterly*, *28*(1), 147–165. doi:10.1177/0891242413513791

Jenkins, J. M., & Duncan, G. J. (2017). Do pre-kindergarten curricula matter? In D. Phillips & Pre-Kindergarten Task Force (Eds.), *The current state of scientific knowledge on pre-kindergarten effects* (pp. 37–44). Washington, DC: Brookings Institution and Duke University.

Jenkins, J. M., Duncan, G. J., Auger, A., Bitler, M., Domina, T., & Burchinal, M. R. (2018). Boosting school readiness: Should preschool teachers target skills or the whole child? *Economics of Education Review*, *65*(August), 107–125. doi:10.1016/j.econedurev.2018.05.001

Jenkins, J. M., Farkas, G., Duncan, G. J., Burchinal, M. R., & Vandell, D. (2016). Head Start at ages 3 and 4 versus Head Start followed by state pre-k: Which is more effective? *Educational Evaluation and Policy Analysis*, *38*(1), 88–112. doi:10.3102/0162373715587965

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

Karoly, L., Ghosh-Dastidar, B., Zellman, G. L., Perlman, M., & Fernyhough, L. (2008). *Prepared to learn: The nature and quality of early care and education for preschool-age children in California*. Santa Monica, CA: RAND.

Kirp, D. L. (2007). *The sandbox investment: The preschool movement and kids-first politics*. Cambridge, MA: Harvard University Press.

Landry, S. H., Crawford, A., Gunnewig, S. B., & Swank, P. R. (2001). Teacher Behavior Rating Scale (TBRS). *Center for Improving the Readiness of Children for Learning and Education*, unpublished research instrument.

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

Malone, L., Carlson, B. L., Aikens, N., Moiduddin, E., Klein, A. K., West, J., … Rall, K. (2013). *Head Start family and children experiences survey: 2009, user's manual*. Washington, DC: Mathematica Policy Research, ICPSR.

Morris, P. A., Mattera, S. K., Castells, N., Bangser, M., Bierman, K. L., & Raver, C. C. (2014). *Impact findings from the Head Start CARES Demonstration: National evauation of three approaches to improving preschoolers' social and emotional competence* (OPRE Report 2014-44). Washington, DC.

National Center on Quality Teaching and Learning. (2014). *Preschool curriculum consumer report*. Retrieved from http://eclkc.ohs.acf.hhs.gov/hslc/tta-system/teaching/practice/curricula

Office of Head Start. (2010). *Head Start program fact sheet fiscal year 2010*. Retrieved from http://eclkc.ohs.acf.hhs.gov/hslc/data/factsheets/fHeadStartProgr.htm

Office of Planning Research and Evaluation. (2016). *Variations in Implementation of Quality Interventions: Examining the Quality-Child Outcomes Relationship in Child Care and Early Education (VIQI) 2016–2021*. Washington, DC: Administration for Children and Families, U.S. Department of Health and Human Services.

Phillips, D. A., Gormley, W. T., & Lowenstein, A. E. (2009). Inside the pre-kindergarten door: Classroom climate and instructional time allocation in Tulsa's pre-K programs. *Early Childhood Research Quarterly*, *24*(3), 213–228. doi:10.1016/j.ecresq.2009.05.002

Phillips, D. A., Lipsey, M. W., Dodge, K. A., Haskins, R., Bassok, D., Burchinal, M. R., … Weiland, C. (2017). Puzzling it out: The current state of scientific knowledge on pre-kindergarten effects. In D. A. Phillips & K. A. Dodge (Eds.), *The current state of scientific knowledge on pre-kindergarten effects* (pp. 19–30). Washington, DC: Brookings Institution and Duke University.

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

Pianta, R., Barnett, W. S., Burchinal, M. R., & Thornburg, K. R. (2009). The effects of preschool education: What we know, how public policy is or is not aligned with the evidence base, and

what we need to know. *Psychological Science in the Public Interest*, *10*(2), 49–88. doi:10.1177/1529100610381908

Pianta, R., La Paro, K. M., & Hamre, B. K. (2008). *Classroom assessment scoring system*. Baltimore, MD: Brookes.

Polikoff, M. S. (2015). How well aligned are textbooks to the common core standards in mathematics? *American Educational Research Journal*, *52*(6), 1185–1211. doi:10.3102/0002831215584435

Porter, A. C. (2002). Measuring the content of instruction: Uses in research and practice. *Educational Researcher*, *31*(7), 3–14. doi:10.3102/0013189X031007003

Preschool Curriculum Evaluation Research Consortium (PCER). (2008). *Effects of preschool curriculum programs on school readiness* (NCER 2008–2009). Washington, DC: National Center for Education Research, Institute of Education Sciences, U.S. Department of Education.

Puma, M., Bell, S., Cook, R., Heid, C., Broene, P., Jenkins, F., … Downer, J. T. (2012). *Third grade follow-up to the Head Start Impact Study: Final report*. Washington, DC: Office of Planning, Research and Evauation, Administration for Children and Families.

Ritchie, S., Howes, C., Kraft-Sayre, M., & Weiser, B. (2001). *Emerging academic snapshot* (Unpublished measure), University of California at Los Angeles.

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

Schmidt, W. H., Wang, H. C., & McKnight, C. C. (2005). Curriculum coherence: An examination of US mathematics and science content standards from an international perspective. *Journal of Curriculum Studies*, *37*(5), 525–559. doi:10.1080/0022027042000294682

Schweinhart, L. J. (2005). *Lifetime effects: The High/Scope Perry preschool Study through age 40*. Ypsilanti, MI: High/Scope Press.

U.S. Department of Education. (2013). *Early childhood education intervention report: The creative curriculum for preschool*, (4th ed.). Washington, DC: Institute of Education Sciences, What Works Clearinghouse.

Vygotsky, L. S. (1978). Interaction between learning and development. *Readings on the Development of Children*, *23*(3), 34–41.

Walters, C. (2015). Inputs in the production of early childhood human capital: Evidence from Head Start. *American Economic Journal: Applied Economics*, *7*(4), 76–102. doi:10.1257/app.20140184

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

Weiland, C., McCormick, M., Mattera, S., Maier, M., & Morris, P. (2018). Preschool curricula and professional development features for getting to high-quality implementation at scale: A comparative review across five trials. *AERA Open*, *4*(1), 1–16. doi:10.1177/2332858418757735

Weiland, C., & Yoshikawa, H. (2013). Impacts of a prekindergarten program on children's mathematics, language, literacy, executive function, and emotional skills. *Child Development*, *84*(6), 2112–2130. doi:10.1111/cdev.12099

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

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

Zigler, E. F., & Bishop-Josef, S. J. (2006). The cognitive child versus the whole child: Lessons from 40 years of Head Start. In D. G. Singer, R. Golinkoff, & K. Hirsh-Pasek (Eds.), *Play = learning: How play motivates and enhances children's cognitive and social-emotional growth* (pp. 15–35). New York, NY: Oxford University Press.

Zill, N. (1990). *Behavior problems index based on parent report*. Washington, DC: Child Trends.

Zill, N., Kim, K., Sorongon, A., Shapiro, G., & Herbison, R. (2008). *Head Start FACES 2003: Data user's guide*. Washington, DC: Westat, ICPSR.