Preschool Curricula and Professional Development Features for Getting to High-Quality Implementation at Scale: A Comparative Review Across Five Trials

Christina Weiland
University of Michigan
Meghan McCormick
Shira Mattera
Michelle Maier
MDRC
Pamela Morris
New York University

Experts have heralded domain-specific play-based curricula coupled with regular coaching and training as our "strongest hope" for improving instructional quality in large-scale public preschool programs. Yet, details from different evaluations of the strongest hope model are not systematically compiled, making it difficult to identify specific features across studies that distinguish the most successful implementation efforts. We performed a cross-study review across five diverse large-scale evaluations (n = 6,500 children and n = 750 teachers across 19 localities and multiple auspice types) to identify common features that have characterized successful implementations of this model to date. We identified six features in our exploratory review that may help to flesh out the strongest hope model for localities considering it—a significant focus on specific instructional content, inclusion of highly detailed scripts, incorporation of teacher voice, time for planning, use of real-time data, and early childhood training for administrators. These six features provide more specific guidance for practitioners and help meet calls in preschool and K–12 for more synthesis of implementation lessons from large-scale research trials.

Keywords: *descriptive analysis, early childhood, multisite studies, professional development, program evaluation, teacher education/development*

THERE is strong public support for expanding the availability of high-quality public preschool in the United States. A bipartisan research team found that 86% of Americans thought that the federal government should help states and local communities build better preschool services and improve access to preschool (Public Opinion Strategies and Hart Research, 2014). The Head Start program presently serves approximately 800,000 preschoolers nationally (Office of Head Start, 2015), and 42 states and a handful of cities fund their own public preschool programs (Barnett, Carolan, Squires, Clarke Brown, & Horowitz, 2015b). Some cities—New York, Seattle, San Antonio, and Boston—have dramatically expanded access to public preschool in the last few years, adding thousands of publicly funded slots in a short time. In November 2016, voters in Dayton and Cincinnati, Ohio, passed ballot initiatives to expand public preschool, and in April 2017, New York announced an expansion to serve 3-year-olds.

To date, increasing the number of slots has proven to be the easier part in expanding preschool nationally. The harder part has been ensuring that all slots are of high quality. In particular, instructional quality—the aspect of a preschool classroom most related to children's school readiness (Pianta, Downer, & Hamre, 2016)—is stubbornly low in most public programs (Weiland, 2016). Work to date suggests that a combination of preschool curricula intentionally focused on specific domains, such as literacy, math, or social-emotional skills, and supported by teacher training and coaching—called the "strongest hope" model (Yoshikawa et al., 2013; Yoshikawa, Weiland, & Brooks-Gunn, 2016) or, most recently, the "good bet" model (Phillips et al., 2017)—may be our most promising tool for moving the needle on preschool instructional quality. Analysis of curricula in this category has highlighted two shared traits: first, they were developed by experts in the given domain (e.g., a math curriculum developed by math development experts); second, they are characterized by intentional play-based activities[1] that are fun and interactive for early childhood classrooms and are designed to support children's

natural trajectories of development within a domain via a specific scope and sequence (Chaudry, Morrissey, Weiland, & Yoshikawa, 2017). After teachers are trained on how to implement these curricula, coaches then serve as mentors who observe teachers' in-classroom work with children on a regular basis, troubleshoot problems in instructional practice, provide constructive feedback, and support teachers to implement curricula at high levels of quality.

However, the success rate of the strongest hope model has not been 100%—not even for different implementations of the same curriculum (e.g., Barnett et al., 2008; Farran & Wilson, 2014). Details on features of different implementations are not systematically compiled; as such, it is difficult for localities and researchers to identify the specific features characterizing more versus less successful trials and thereby increase the likelihood of successful replication in their own contexts. Our goal in the present article is to provide a first-hand cross-study comparative synthesis of five implementations of the strongest hope model. In so doing, we aim to identify actionable common elements of successful large-scale implementations of strongest hope models that will aid local efforts to replicate the models and meet calls in preschool and K–12 education for more research and information on implementation features (Durlak, 2010; Harris, 2016).

We have jointly worked on three large-scale studies of five domain-specific curricula coupled with training and coaching. All told, our evaluations included approximately 6,500 children and 750 lead teachers in 19 localities spanning urban, suburban, and rural settings in Boston, New York City, California, Colorado, Illinois, Texas, Ohio, Mississippi, Pennsylvania, Maryland, New Jersey, and Massachusetts. Notably, like potential implementers of these models, we were not the authors of any of these curricula. We largely had to learn about the programs the way that localities would—through discussions with program developers, attending trainings, and reading materials. In each trial, we found that the strongest hope model was able to be implemented with good fidelity at scale and led to an uptick in high-quality instructional strategies—no small feat given that poor to middling curriculum implementation is a common and vexing problem in K–12 (Cohen & Ball, 2001) and preschool education (Davidson, Fields, & Yang, 2009; Kuperschmidt, Bryant, & Willoughby, 2000).

We begin by providing background on the origins and definition of the strongest hope model, followed by information on the present preschool landscape and details about the five focal trials in this article. Then, through an exploratory and hypothesis-generating review, we address two central questions:

Question 1: Did the curricula, training, and coaching in these trials share specific features?

Question 2: Beyond curricula, training, and coaching, were there common supports across the evaluations that were critical in attaining good curriculum fidelity?

We conclude with specific actionable recommendations for researchers and localities as they engage in the on-the-ground work of improving preschool instructional quality at scale.

Background on the Strongest Hope Model

The strongest hope model concept and definition emerged from a joint Society for Research on Child Development–Foundation for Child Development policy brief in 2013 on the evidence base for preschool, authored by 10 leading early childhood experts (Yoshikawa et al., 2013).

The brief's authors defined the model as including the following components:

- *Domain-specific curricula*: play-based curricula that "aim to provide intensive exposure to a given content area based on the assumption that skills can be better fostered with a more focused scope" (p. 7)—as compared with global curricula, which focus on overall teacher pedagogy across domains and which do not have a specified scope and sequence for building children's skills.
- *Intensive professional development*: curriculum-focused training and "coaching at least twice a month, in which an expert teacher provides feedback and support for in-classroom practice, either in person or in some cases through observation of videos of classroom teaching" (p. 8).
- *Monitoring of child progress*: for some curricula—specifically "assessments of child progress that are used to inform and individualize instruction, carried out at multiple points during the preschool year" (p. 8).

In reaching their conclusions, the authors were effectively summarizing a pattern across existing studies—a pattern that has a theoretical basis in other literature. For example, beginning in the mid-1980s and updated since, the National Association for the Education of Young Children (2009) has put forth a framework of developmentally appropriate practice that emphasizes (a) that there are known sequences in which children gain specific concepts, skills, and abilities; (b) that familiarity with these sequences should inform teachers' practices; (c) that good teaching is intentional and goal oriented; and (d) that teachers must know where each child is relative to classroom learning goals, to be intentional about helping individual children to progress. Domain-specific curricula with monitoring of child progress provide a concrete path for increasing teachers' knowledge of practices that support children's developmental trajectories and for actualizing these principles daily in the classroom.

The theoretical basis for coaching comes from a different source: the science of adult learning theory, which holds (a) that adults are most interested in learning when it has immediate relevance to their jobs; (b) that adults learn from reflecting

on problems that arise when they try to apply their new knowledge/skills; and (c) that adults want to be actively involved in directing and evaluating their own learning (Knowles, Holton, & Swanson, 2005; Michigan Association of Intermediate School Administrators General Education Leadership Network Early Literacy Task Force, 2016). The coaching models in the strongest hope studies largely match these adult learning tenets, as do all of our five focal models. Notably, coaching also solves the "problem of transfer," or the challenge of taking information learned in a training and transferring it effectively for use in real-world classroom conditions (Joyce & Showers, 1980, 1982; Kraft, Blazer, & Hogan, 2016). Following sociocultural learning theory, individuals can learn best when they are provided with opportunities to discuss and reflect with others, apply new ideas and skills in practice while receiving feedback from an expert, and have effective practices modeled for them (Collins, Brown, & Holum, 1991; Honig & Ikemoto, 2008). To solve the transfer problem following this theory, these opportunities must take place within the specific context where actual instruction is set to take place—the teacher's classroom (Marsh, 2012).

Beyond this theoretical basis, three new meta-analyses provided empirical evidence for some (but notably not all) of these key features. For example, one meta-analysis summarized 34 experimental and quasi-experimental studies published since 2007 on preschool curricula and found small to moderate effect sizes of targeted language, literacy, and math curricula relative to global curricula on child outcomes (Nguyen, 2016), demonstrating the value of domain-specific curricula. Likewise, a meta-analysis of 32 studies published since 1999 reached the same conclusion in its evaluation of 22 language/literacy and globally focused programs (Chambers, Cheung, & Slavin, 2016). Moreover, a meta-analysis examining language- and literacy-focused professional development across 33 trials within 25 studies found that professional development had medium to large effects on classroom quality and small to medium effects on children's literacy skills (phonological awareness and alphabet knowledge), with stronger benefits when professional development supports included coaching versus not (Markussen-Brown et al., 2017), suggesting the benefit of intensive professional development, the second component. No meta-analysis to date has examined the third component of the strongest hope model nor all three strongest hope model components together (curriculum, coaching, and assessments). However, these recent works do lend some initial empirical support for these approaches as preschool quality improvement strategies that have been heretofore identified in the practice and theoretical literature.

To be sure, a practical limitation of the evidence base on domain-specific curricula to date is the real-world reality that preschool programs are charged with improving multiple domains of school readiness. Adopting only a math curriculum, for instance, would not meet the full needs of enrolled children. Some localities/interventions have responded to this challenge by layering different domain-specific curricula (e.g., a language/literacy and socioemotional-focused curricula with a math curricula, as in our Boston focal study), with extensive support for teachers to integrate these curricula. Another team recently brought together experts in four domains to develop one integrated curriculum that addresses all domains (language and literacy, math, science, and socioemotional; Connect4Learning, n.d.). Analyses so far show no detriments of adopting multiple domain-specific curricula. For example, there were benefits in Boston on the nontargeted domain of executive function (Weiland & Yoshikawa, 2013), and an analysis across 14 early childhood curricula found no differences on child socioemotional skills among targeted language, literacy, and math curricula relative to global curricula (Duncan et al., 2017). Connect4Learning has not yet been evaluated.

Present Preschool Landscape

Despite a growing evidence base for the strongest hope model (see Weiland, 2016; Yoshikawa et al., 2013; Yoshikawa et al., 2016), few programs nationally have adopted this approach. Rather, dominating the preschool landscape are so-called whole-child or global curricula—that is, curricula that purport to address all domains of child development but do not have a specified scope and sequence within each domain and do not allow for much depth of focus on any one domain (Jenkins & Duncan, 2017). The What Works Clearinghouse (WWC; 2016), which reviews rigorous studies of preschool curricula and gives each curriculum an effectiveness rating, rates one of these common choices—Creative Curriculum—as having an effectiveness rating of zero for children's mathematics, oral language, phonological processing, and print knowledge skills.[2] The Preschool Curriculum Consumer Report developed specifically for Head Start programs, which uses a different set of criteria in its rating system, likewise determined that there is no evidence that Creative Curriculum has positive impacts on child outcomes (National Center on Quality Teaching and Learning, 2015).

Similarly, although detailed data on preschool professional development are not tracked nationally, there is work showing that teachers have traditionally received training through brief workshops that take place in summer and short school breaks, often with a specific number of hours required per year (Joyce & Showers, 2002). These training sessions can be described as "one-off" because their themes are not connected and they are not designed to follow a cohesive and aligned strategy to support teachers in achieving well-defined goals that are guided by a curriculum with a clear scope and sequence. There is little to no evidence that this approach to professional development leads to enduring changes in teachers' interactions with children and their ability to provide high-quality instruction in early childhood classrooms (Wasik, Mattera, Lloyd, & Boller, 2013; Zaslow, Tout, Halle, Whittaker, & Lavelle, 2010).

Coaching for early childhood educators, often used to supplement and support a training model, is on the rise. A compendium summarizing approaches used in existing Quality Rating and Improvement Systems (QRIS) found that 26 state QRISs reported that some type of coaching or individualized on-site assistance was available to preschool programs to help them improve (Tout, Halle, Zaslow, & Starr, 2011). Notably, however, this report also found that the strategies for providing coaching varied considerably across QRISs and did not necessarily reflect the type of approach argued for in the strongest hope model. Indeed, although some coaching models provide systematic regular coaching for all preschool teachers, a more typical approach used in many localities is one where coaches can be available when teachers request a meeting with them or when a principal notes that a teacher needs support from a coach (Cohen & Kaufmann, 2000). It is also clear that coaching models in practice are rarely paired with intentional support for implementation of domain-specific curricula, a feature that the strongest hope model argues is essential. Recent work showed that coaching models not tied to a curriculum but focused instead on improving general teacher practices have been largely unsuccessful in improving instructional quality (Piasta et al., 2017; Yoshikawa et al., 2015).[3]

The reasons for these research-to-practice chasms are undoubtedly varied and complex, ranging from program requirements and funding to marketing and historical curricular choices. However, there have been recent inklings of a shift toward the strongest hope model. One option for sites in New York City's universal preschool program known as Pre-K for All, for example, is an evidence-based mathematics curriculum, Building Blocks, alongside training supported by the curriculum developer and tiered coaching. The not-yet-implemented new Head Start standards also call for curricula that are "sufficiently content rich" and "have an organized developmental scope and sequence that include plans and materials for learning experiences based on developmental progressions" (Head Start Performance Standards, 2016). We suspect that other sites and localities will similarly be following suit, as the message about the success of these kinds of models is increasingly understood.

Background: Five Focal Strongest Hope Implementations

The five strongest hope implementations discussed in this article occurred within three large-scale studies:

Boston Study: An age-based regression discontinuity evaluation of the Boston prekindergarten program (N = 2,018 children) that implemented a language and literacy curriculum (Opening the World of Learning [OWL]) with a mathematics curriculum (Building Blocks) and supported teachers via ongoing training and approximately biweekly in-classroom coaching (Weiland & Yoshikawa, 2013).

Head Start CARES Study: A national large-scale randomized controlled trial (N = 2,114 children) of three theoretically distinct social-emotional curricula, supported by weekly in-classroom coaching and ongoing training and technical assistance in Head Start classrooms (Mattera, Lloyd, Fishman, & Bangser, 2012; Morris et al., 2014), versus business-as-usual Head Start. The individual curricula were supported by the CARES training and coaching model and are referred to as Head Start CARES (HS CARES): Incredible Years, Preschool PATHS, and Tools of the Mind–Play (adapted from the original Tools of the Mind curriculum).

Making Pre-K Count Study: A large cluster-randomized controlled trial (N = 2,715 children) testing the efficacy of the Building Blocks math curriculum, coupled with in-classroom coaching and ongoing training in 4-year-old classrooms, versus business-as-usual preschool. The Making Pre-K Count (MPC) study took place across 2 years in New York City preschools serving low-income children (Mattera & Morris, 2017; Morris, Mattera, & Maier, 2016).

In addition to their collective large sample sizes and our aforementioned status as nondeveloper evaluators, these trials represent an excellent sample for an exploratory comparative review of common features because of their diversity in settings and participants. See Table 1 for a full comparison of the studies' background features and information on the demographics of participating schools, teachers, and children. As illustrated, auspice varied from only public schools (Boston) to only Head Start (CARES) and a combination of public school and community-based preschool settings (MPC). Participating children were diverse in terms of racial/ethnic and socioeconomic status. For example, the Boston study included children from a range of racial and family income backgrounds, while the HS CARES study included only low-income children who qualified for Head Start. The majority of participating children in all three studies were non-White. Teacher characteristics also varied within and across studies. Although the majority of teachers had at least 5 years of experience across the studies, as illustrated in Table 1, the percentage of teachers holding a graduate degree ranged from 7% (HS CARES) to 86% (MPC). Accordingly, the settings represent a cross section of the preschool landscape today, which is characterized by diversity in auspice (Barnett et al., 2015a), student population (Phillips, Johnson, Weiland, & Hutchison, 2017), and teacher workforce (Phillips, Austin, & Whitebrook, 2016).

Key findings across focal studies. To contextualize the common features highlighted in the next section, we briefly summarize key results from the three studies. Importantly, each program model evaluated had shown evidence of efficacy in prior trials (e.g., Barnett et al., 2008; Clements, Sarama, Spitler, Lange, & Wolfe, 2011; Domitrovich, Cortes, & Greenberg,

TABLE 1
Background Characteristics of Three Strongest Hope Implementation Models

Background characteristic	Strongest hope model		
	Boston	HS CARES	MPC
Location	Boston, MA	CA, CO, IL, TX, OH, MS, PA, MD, NJ, MA	New York, NY
Program auspice	Public preschools	Head Start centers	Public preschools and community-based centers
Program eligibility	All 4-year-olds	Income targeted	All 4-year-olds
Period			
Year 1	2008–2009	2009–2010	2013–2014
Year 2	N/A	2010–2011	2014–2015
Study sample sizes, *n*			
Preschool center or school	69	104	69
Lead teacher/classroom	250	307	350
Students	2,018	2,114	2,715
Teacher characteristics			
Education	78%, master degree	37%, associate degree or less; 54%, BA; 7%, master degree	86%, master degree
Experience	75%, 5 years of teaching experience	30%, 3–10 years of teaching experience; 63%, >10 years of experience	Mean years teaching, 15.23 (*SD* = 8.87)
Student demographics[a]			
Black, %	26	33	37
Hispanic, %	41	43	56
White, %	18	16	3
Asian or other race, %	14	7	3
Eligible for free/reduced-price lunch, %	69	—	—
Average monthly income, $	—	1,763	—
Parent has at least high school degree/ GED, %	—	—	74

Note. HS CARES = Head Start CARES; MPC = Making Pre-K Count; N/A = not applicable. Dashes (—) indicate that data are not available.
[a]In MPC, parent, not student, demographic data were collected. Different indicators of family socioeconomic status were available across studies; as such, there is no comparative reference.

2007; Webster-Stratton, Reid, & Hammond, 2001, 2004; S. J. Wilson, Morse, & Dickinson, 2009).[4] The studies reported here represent efforts to implement these evidence-based programs at scale in a set of participating localities. Details on each study's findings are illustrated in Table 2. As described there, all studies achieved a moderate to high level of fidelity and demonstrated positive upticks in teachers' practices in the spring of the school year. For example, in the Boston study, descriptive data on treatment classrooms showed that after 2 years of implementation of the coaching and curriculum model, average instructional quality was the highest to date in the literature on large-scale preschool programs (Weiland, Ulvestad, Sachs, & Yoshikawa, 2013). The HS CARES study demonstrated moderate to large positive impacts of each of the three tested curricula on targeted teacher practices. Incredible Years improved teachers' classroom management—and, to a lesser extent, teachers' social-emotional instruction—while

Preschool PATHS enhanced teachers' direct social-emotional instruction. Tools of the Mind–Play improved teachers' scaffolding of pretend play. In MPC, teachers assigned to the intervention condition met most fidelity benchmarks and engaged in more and slightly higher-quality math instruction than control group teachers—spending an additional 12 minutes on math and offering an average of nearly two more math activities in a 3-hour observation period.

However, Tools of the Mind–Play did not improve children's outcomes in preschool, and effects in MPC did not emerge until a year following preschool, when more sensitive measurement was gathered and some children received an intervention boost in kindergarten (the MPC study is still underway; initial findings show no effects in preschool but benefits to children's academic and executive functioning outcomes at the end of kindergarten; Mattera & Morris, 2017). To be clear, our bar for including these five studies

TABLE 2

Comparison of Strongest Hope Model Impact and Implementation Study Designs and Key Findings

Study characteristic	Strongest hope model		
	Boston	HS CARES	MPC
Years of implementation before impact study	2	0	1
Impact study			
Design overview	Impacts on students evaluated via age cutoff regression discontinuity	Cluster randomized trial; centers randomly assigned to one of three curricula (IY, PP, or TM) or business-as-usual control condition	Cluster randomized trial; preschool sites randomly assigned to Building Blocks intervention or business-as-usual control condition
Counterfactual condition	No public preschool; preschool and child care outside of formal public preschool	Business-as-usual Head Start	Business-as-usual preschool program
Findings—positive impacts on			
Teacher instructional practices	Yes[a]	All curricula: yes	Yes
Effect sizes	N/A	0.26–1.36[b]	0.40–1.16[c]
Student academic outcomes	Yes	All curricula: no consistent impacts	No (in preschool)[d]
Effect sizes	0.44–0.62[e]	−0.06 to 0.09	0.05–0.08[c]
Student social-emotional outcomes	Yes	All curricula: yes	Not examined
Effect sizes	0.19[e]	0.12–0.29[b]	N/A
Student executive function outcomes	Yes	IY, yes; PP, yes; TM, no	Yes
Effect sizes	0.21–0.28[e]	−0.06 to 0.20[b]	0.10[c]
Implementation study			
Design overview	Descriptive approach to measuring quality and curricula fidelity (no control group classrooms)	Descriptive approach captured extensive information on implementation of training and coaching and teachers' use of the programs. Implementation monitored across 2 years. Impacts on students examined at the end of preschool.	Descriptive approach captured extensive information on implementation of training and coaching and teachers' use of Building Blocks. Implementation monitored across 2 years. Impacts for teachers and children examined at the end of the second year of implementation.
Implementation study findings	Moderate to high implementation of curricula; strong instructional quality on average (mean CLASS score = 4.27)	Strong implementation of coaching and training model; satisfactory implementation of the three models at a prespecified level; TM implementation was somewhat lower than other programs, possibly due to program complexity	Teachers implemented most curricular components successfully at levels prespecified by the research team.

Note. HS CARES = Head Start CARES; MPC = Making Pre-K Count; IY = Incredible Years; PP = Preschool Paths; TM = Tools of the Mind–Play; N/A = not applicable.
[a]The Boston study utilized a regression discontinuity design and therefore did not directly estimate effects on teacher instructional practices. [b]See Morris et al. (2014) for more information. [c]See Morris, Mattera, and Maier (2016) for more information. [d]No statistically significant effects in preschool; positive effects in kindergarten not described here (for more information, see Mattera & Morris, 2017). [e]See Weiland and Yoshikawa (2013) for more information.

was that the focal curricula were implemented with high fidelity, with some level of evidence that classroom quality increased as a result. The broader literature is clear that fidelity matters; effects on participants are generally stronger when implementation is higher (Wilson & Lipsey, 2001). Furthermore, strong fidelity is hard to achieve; in at-scale

interventions, fidelity is typically around 60% and rarely reaches 80% (Durlak & DuPre, 2008).

Heterogeneity of results across trials. The diverse range of implementation contexts and varying research designs for our focal models allows us to theorize why we observed heterogeneity in program impacts on children across trials, although admittedly only within the range of this particular set of studies. For example, the counterfactual conditions across the studies may explain heterogeneity of impacts. The Boston counterfactual was a combination of community-based preschool and child care (two-thirds of the control group) and staying at home with a parent or family member (one-third of the control group; Weiland & Yoshikawa, 2013). In contrast, all the children enrolled in the HS CARES study were enrolled in Head Start, and the study estimated the impacts of three social-emotional enhancements plus Head Start versus Head Start on its own. Similarly, all children included in the MPC study were enrolled in a formal public or community-based preschool, and the study estimated the impact of MPC versus business-as-usual preschool, which already included a lot of time spent on math activities (i.e., teachers in the control sites were teaching 35 minutes of math, considerably more than the 10–27 minutes documented in previous trials; e.g., Clements & Sarama, 2008).

The studies also took place across varied settings for implementing preschool. For example, the Boston study tested the impact of a scaled program in a moderately large city, where the district had some central control and direction over the program model implemented in schools. Moreover, the program was implemented only in public preschool settings, located in the same building as early elementary school grades. The HS CARES study was implemented in 17 Head Start grantees and >100 centers distributed across the country. All were required to follow Head Start regulations, but they differed greatly in their teacher and student populations. MPC was implemented in community-based and public school preschools in the largest city in the United States, at a time of marked change in its prekindergarten landscape. At the time of implementation, there was growing district-level oversight over preschool programs that arguably provided more structure to the prekindergarten day, although that varied depending on whether sites were community-based organizations or public schools. In comparing these contexts, program implementation in Boston was likely to be most consistent across settings, due to the scale of the program and the district-level leadership in that study.

In addition, study implementation findings suggest differences across studies in implementation quality. For example, Tools of the Mind–Play implementation was somewhat lower than other curricula tested in HS CARES, with teachers rated on average just below the benchmark of a 3 on a 1-to-5 scale

of implementation quality that examined indicators such as whether the classroom looked like a Tools of the Mind–Play classroom, whether teachers were comfortable with and using the curricular activities, and whether they were integrating the curriculum's theory into everyday practice. This may possibly have been due to the complexity of the Tools of the Mind–Play curriculum and its unique requirement that teachers substantially change the structure and setup of the preschool day and classroom. Programs that require less change to the day-to-day functioning of a classroom and that work within the structures already in place—such as whole group, center time, or small group—may be easier to implement and therefore more successful than those that substantially change a teacher's daily schedule or classroom setup. In the MPC study, implementation of individualized curricular components (e.g., computer games) was difficult to implement consistently.

Notably, the MPC study used a measure in preschool that did not deeply assess children's competencies in geometry, one of the unique foci of Building Blocks. When a more comprehensive measure assessing geometry was collected in kindergarten (similar to the measure used in the Boston Study), effects were positive (Mattera & Morris, 2017). This suggests that the program may have had an effect in preschool that went unmeasured and that program effects (or the lack thereof) may be sensitive to child assessment measurement choices.

We are not able to pinpoint which difference (or differences)—counterfactual, setting, intervention complexity, or measurement—is responsible for the heterogeneity of findings at the child level. However, such sources of heterogeneity are ripe areas for future work on preschool curricula and professional development. In our synthesis, as we explain in the next session, we examine commonalities across the trials that can clarify the strongest hope model while fully acknowledging that heterogeneity in effects on teachers and children is generally the norm in early childhood education (Bloom & Weiland, 2015; Morris et al., 2017) and is likely to be so in future implementations of the strongest hope model.

Approach

To address our two central questions, we undertook a cross-study comparative review that took advantage of our deep knowledge of the five trials. We chose to focus on the five models in which we had intimate on-the-ground knowledge of the curricula, training, and coaching components because there is no common repository nor agreed-on categories for implementation data and details across early childhood curricular interventions. While this choice limits the external validity of our analysis, our evaluations did span a combined 6,500 children and 750 lead teachers in 19 localities in urban, suburban, and rural settings in Boston, New York City, California, Colorado, Illinois, Texas, Ohio, Mississippi,

TABLE 3

Model and Implementation Data Used in the Present Study

Data type	Strongest hope models		
	Boston	HS CARES	MPC
Curriculum manuals and materials	×	×	×
Fidelity observations	×	×	×
Field notes from fidelity visits	×	×	
Coach interviews	×	×	×
Teacher interviews		×	×
Center director interviews		×	
Field notes from coaching meetings	×		
Coach meeting attendance			×
Training attendance logs		×	×
Training content		×	×
Coaching logs		×	×
Coaching manuals or protocols		×	×
Teacher surveys	×	×	×
Coach surveys	×	×	×

Note. HS CARES = Head Start CARES; MPC = Making Pre-K Count.

Pennsylvania, Maryland, New Jersey, and Massachusetts—an unusually large sample in the early childhood context. Furthermore, because our five focal models spanned language, literacy, mathematics, socioemotional skills, and self-regulation, our purposive sampling approach covered the major domains commonly targeted by preschool curricula. Importantly, focusing on these five trials ensures that our identified common features and components are not based on misinterpretations of a given model's components nor on incomplete data for a given model. For example, a study might not have reported the presence or absence of a particular curricular feature (e.g., scripting), and/or study authors might disagree with our categorization of what constitutes a scripted versus nonscripted curriculum. Simply put, in our view, a detailed review of the strongest hope models at this early stage of implementation research in the early childhood field requires direct experience with them.

After we made the decision to use this approach, a member of the original study team for each trial identified project data pertinent to our review. For parsimony, data types used in the present analysis are shown, by trial, in Table 3. Four data types—curriculum manuals and materials, fidelity observations, coach interviews, and teacher and coach surveys—were available for all focal studies. Others were available for a subset of the trials (e.g., center director interviews available only in HS CARES). The key study findings that we reviewed are summarized in Tables 2 and 4. As illustrated, Table 2 provides an overview of each research trial's impacts on teacher instructional practices, as well as children's socioemotional, executive functioning, and academic outcomes. This table also reviews implementation findings

from each trial and describes whether each model was implemented with fidelity. Table 4 lists specific intervention components and details on curriculum and training in each trial. Table 5 does the same for coaching.

Next, these data were reviewed by a researcher with deep involvement in the original study to identify salient features and implementation supports, particularly those that had not emerged in prior publications from the original study team. Then, via a collaborative and iterative process, each researcher shared study-specific features and implementation supports with the full team, and we examined whether these features and supports were present in the other trials as well. We paid attention to features and supports that were common across at least four of the five trials and to those that were unique but were also viewed as being especially influential in a given study. As we continue to detail in the next section, our identified commonalities span curricular, professional development, and organization support features.

Common Features and Supports

Common Curricular Features: Instructional Content and Scripts

As detailed earlier and in Table 4, the curricula examined across the studies covered a variety of child development domains, including language and literacy, mathematics, socioemotional development, and executive function. In accordance with the strongest hope model (Yoshikawa et al., 2013), all were designed by experts in the domain targeted by the curriculum, and all were strongly based in theory and empirical research about how such skills develop as part of

TABLE 4

Comparison of Strongest Hope Model Intervention Components: Curricula, Assessments, and Training

Study characteristic	Strongest hope model		
	Boston	HS CARES	MPC
Curriculum			
Curricula used	OWL and BB	IY, PP, or TM	BB
Learning domains targeted	Language/literacy (OWL) and math (BB)	Social-emotional skills	Math
Manualized curriculum	Yes	IY: yes; PP: yes; TM: yes	Yes
Scripted curriculum	Yes	IY: yes; PP: yes; TM: no	Yes
Formative assessments used	Yes	No	Yes
Training			
Total number of training days	13 total	IY: 6; PP: 4; TM: 5	12 total
Days in beginning of year	OWL: 5; BB: 2	IY: 1; PP: 2; TM: 2	Year 1: 2; Year 2: 0[a]
Days throughout the year	OWL: 2; BB: 4	IY: 5; PP: 2; TM: 3	Year 1: 5; Year 2: 5
Training provider and structure	External trainers conducted Year 1 training with BPS coaches; BPS coaches led training in Year 2	Developer-certified trainers trained through train-the-trainer model	Curriculum developers and staff; supported by trained coaches
Training manualized	No	Yes	Yes
Train-the-trainer certification	No	Yes	Yes
Focus of Year 1 training	Building teachers' content knowledge and strengthening knowledge of the curriculum components	Training on pedagogical approach and theory behind curricula	Building content and pedagogical knowledge about mathematics and curricular components
Focus of Year 2 training (if applicable)	Same as Year 1 (focus on new teachers)	N/A	Planning for and differentiating math activities for children at different levels of knowledge and skill
Training structure	Full group presentation of curriculum components. Focus on theory/procedures; content delivered through breakout groups/hands-on activities.	Large group presentations, as well as small group discussions/practice (including video watching and discussion)	Large group overview followed by small group hands-on activities

Note. HS CARES = Head Start CARES; MPC = Making Pre-K Count; OWL = Opening the World of Learning; BB = Building Blocks; IY = Incredible Years; PP = Preschool PATHS; TM = Tools of the Mind–Play; BPS = Boston Public Schools; N/A = not applicable.
[a]The MPC evaluation rolled out across 2 years. The first year focused on supporting teachers to achieve high-quality implementation. During Year 2, impacts on students were examined. As such, teachers implemented the intervention for 2 years prior to impact data collection.

typical child development. All drew teachers' attention to how children's skills develop in a given domain and what to do next to support those skills. Notably, increasing teacher knowledge of skill-specific developmental trajectories was one of the National Academy of Science's recent recommendations for better preparation of the early childhood workforce (Institute of Medicine and National Research Council, 2015). Beyond this shared foundation, our comparative synthesis raised two other similarities of the evaluated curricula that we hypothesize contributed to their successful implementation: a direct and depth of focus on instructional content and inclusion of highly detailed teacher scripts.

Regarding instructional content, most of the curricula included in our studies diverged from common early childhood practice in that they taught specific content aimed at improving children's conceptual knowledge. For example, OWL (Boston) teaches rich science content beyond basic skills, such as the specifics of the life cycles of plants and how to discuss, record, and describe plant growth. Preschool PATHS provides lessons on emotional knowledge and social problem-solving skills—for example, directly teaching children what facial features are associated with which emotion and how to solve a problem with another child—and program impacts clustered in these areas directly addressed by the curriculum. Tools of the Mind–Play, which did not have impacts on its targeted child outcomes (executive function and problem behaviors), was the least content focused of the five focal curricula as it was implemented in the HS CARES trial. Tools of the Mind–Play focused solely on the pretend play aspects of the larger Tools of the Mind curriculum and

TABLE 5
Comparison of Strongest Hope Model Intervention Components: Coaching

Characteristics of coaching model	Strongest hope models		
	Boston	HS CARES	MPC
Coaches, *n*	8	52	15
Education	All had master's degrees	Most had graduate degree	Most had graduate degree
Experience	Average: 8.8 years, teaching experience; 3.3 years, coaching experience	Minimum experience: 11 years in early childhood; 0–4 years in adult education	All had at least 3 years of experience in early childhood; majority had 3–10 years in adult education
Employer	Boston Public Schools	Head Start grantee; funded explicitly for the demonstration	Bank Street College; funded explicitly for the demonstration
Coaching model duration			
Year 1	1.25 hours every other week	1.5 hours per week	2 hours per week
Year 2	1.25 hours every other week	N/A	2 hours every other week
Group coaches trained by	Curriculum developer on content; some outside consultation provided on how coaching works (minimal and not core to implementation)	Curriculum developers	Curriculum developers
Structure of coaching sessions	Watching live instruction, giving feedback/reflecting with teacher; modeling or helping teacher with planning/planning process	A 30-minute reflection meeting to reflect on practice and set priorities/ goals, followed by 60-minute period with a variety of strategies for observation (e.g., commenting, modeling)	Observe live instruction and offer curriculum guidance in the moment; meet with teachers to debrief from the observation, reflect on implementation, set goals, and plan for upcoming weeks
Structure of coach supervision	Coaches worked collaboratively and were considered practice experts; some supervision from executive director of Boston Public Schools Department of Early Childhood. Two coaches out of group did coach evaluations.	Regular supervisory phone calls between coaches and trainers; varied by program from weekly supervision (PP) to every other week (TM) to approximately monthly (IY)	Group and one-on-one supervision between coaches and one coaching supervisor, hired by the Bank Street College of Education
Coach role at training	Year 1: coaches facilitated small breakouts. Year 2: coaches led all components	Coaches attend training, set up logistics, and may help prompt discussion in training	Attended all training sessions. facilitated small group breakout sessions, helped develop Year 2 training content
Coaching manualized	In process	Yes	Yes

Note. HS CARES = Head Start CARES; MPC = Making Pre-K Count; PP = Preschool PATHS; TM = Tools of the Mind–Play; IY = Incredible Years; N/A = not applicable.

not the self-regulation games. It was also theorized to improve children's executive function, not by teaching specific content knowledge, but by setting up the classroom and children's activities in such a way that opportunities for self-regulation were maximized (Morris et al., 2014).

Second, across the five curricula, four were relatively well scripted—that is, in addition to instructions for each curricular activity, they included what the teacher should specifically say when conducting the activity. For example, one component of the OWL language and literacy curriculum used in Boston is full group read-alouds (i.e., the same book is read once per day for four days, with each reading focusing on a specific goal; Schickedanz & Dickinson, 2005). The curriculum includes specific scripts for each book, directly illustrating how to accomplish the goals of the different read-alouds (goals range from basic comprehension and recall to high-level inference making, depending on the read number). For instance, for the book *Corduroy* in Unit 1, the model script for Read 1 details, page by page, which vocabulary words to support and how (e.g., on pages 8–9, "As you read, explain

escalator by saying it is a set of stairs that move; people move on them"), as well as specific language for supporting children's comprehension of the story line (e.g., on pages 24–25, the teacher should add to the text: "four flights of stairs is a lot, and Lisa ran up all the way. Lisa was very excited about bringing Corduroy home, wasn't she?"). To be clear, simply robotically reading a script was not intended by the developers of the focal curricula, nor would doing so be considered a high-quality practice by any early childhood expert. Rather, the scripts serve as physical reminders that teachers can read in advance or have in front of them as they teach—not unlike the checklists that airplane pilots and surgeons use to ensure that they take the necessary specific steps for success in their endeavors (Gwande, 2009).

Similarly, the Preschool PATHS curriculum in the HS CARES study used highly scripted weekly lessons to introduce social and emotional concepts to children, often drawing on the curriculum's puppets to deliver the content. For example, one week introduces children to the emotion "happy." With background information and instructions for the teacher, the curriculum materials give a script for a happy conversation between Twiggle the Turtle and Henrietta the Hedgehog, as well as scripting for the teacher to explain what the word *happy* means, how it looks on people's faces, and what makes people happy.

Notably, the least scripted of the five curricula, Tools of the Mind–Play, showed the lowest level of implementation. While Tools of the Mind–Play included extensive manuals and documentation of the theory behind activities, highly defined scripts that the teacher could immediately use to guide delivery of daily lessons were less common. The HS CARES research team hypothesized that without a defined set of scripts, it was difficult for the teachers to pick up the curriculum and provide content right away to children without extensive preparation and planning each week. In a highly controlled setting with extensive support, teachers may be able to plan, integrate, and execute many activities with minute details. However, when the curriculum is implemented at scale, other demands may make it difficult to prepare extensively every week, such as additional program requirements, assessments, special activities and field trips, breaks, and other curricula. Scripting helps alleviate this problem of scale by "pre-preparing" some of the materials and language for teachers, thereby reducing their logistical and cognitive loads. Scripting might also reduce implementation variability until teachers truly "get" the model.

Common Coaching Features: Teacher Voice and Real-Time Data

The strongest hope model description to date has emphasized that training and coaching occur regularly (approximately one or two times per month for coaching; Yoshikawa et al., 2013). All our implementations approximately matched this guideline, ranging from weekly to biweekly coaching across studies and localities. As illustrated in Table 4, all five curricula were supported by considerable training that continued throughout the year, ranging from 4 to 7 days per year. Also consistent with the strongest hope model, training across all models was designed to provide teachers with information about the domain of interest, how to use the curriculum, and strategies to support high-quality implementation. Training sessions were structured to build teachers' theoretical and content knowledge in a given area and to provide hands-on guided opportunities to try out curriculum materials. Following initial training, teachers would return to their classrooms, practice the new skill or activity, and then return to training to debrief from their on-the-ground experience and learn new content (see Table 5). Likewise, coaching was curriculum focused across the studies, with coaches providing feedback on teachers' in-the-moment implementation of lessons and activities.

Beyond these characteristics, another common feature that we hypothesize to have been important to the success of all five implementation models was that each made room for *teacher voice*. Coaches were not merely imposing curriculum or monitoring teachers to see if they were implementing it. Rather, the mentoring relationships were designed to be a "two-way street," with teachers raising questions and struggles from their practice for coach feedback. In return, coaches shared supportive but constructive feedback on teachers' practice in relation to curriculum implementation. Abby Morales, a longtime coach in Boston, expressed the coaching philosophy regarding teacher learning: "Adults like to learn but they don't like to be taught." The two-way street approach in each model facilitated relationships that were respectful of teachers' perspectives and goals and, we hypothesize, led teachers to be more receptive to making changes in their practice and to effectively implementing the curricula. Notably, coaches in all five models were not direct supervisors of the teachers whom they coached, which may have allowed teachers to feel more comfortable and safe in using their voices.

Coaches also made key use of real-time data. Four of the five focal implementations (MPC and the three HS CARES) used large real-time management information systems for coaches to keep track of coaching dosage and curriculum implementation levels. Via an online management information system, coaches provided information on a weekly and monthly basis about how teachers were implementing in the classrooms. A technical assistance team regularly monitored those data to identify patterns of implementation challenges in individual classrooms or across classrooms. For example, if a specific issue arose consistently across sites, such as how to rotate children through small groups, coaches could work with teachers to discuss how to work with the assistant teacher and set up classroom management practices that allowed the children to rotate—an issue that might not have

been addressed if each teacher was struggling with it independently. This level of data collection was also fed back into the implementation support system through coach supervision. In MPC, coaches and their supervisors examined the summaries of the data to help identify areas for technical assistance, while in HS CARES, the technical assistance team reached out to coaches and supervisors with specific issues that their teachers were having that arose from the data. The MPC model went a step further and also collected real-time child-level data. Each week, teachers collected written information—that is, formative assessments—about children's skills and abilities based on their responses in small group activities. When the teachers met with coaches, they worked together to write down specific goals for individual children and tie those goals to implementation behaviors that the teacher would try the following week, including how to individualize instruction.

Notably, the more weakly implemented Tools of the Mind–Play model also included this feature, suggesting that it may not be sufficient in the absence of other features, such as scripting (which Tools of the Mind–Play lacked). Another possibility, however, is that without the use of real-time data, the implementation of Tools of the Mind–Play may have been weaker.

Common Feature: Organizational Supports

Work by implementation researchers has drawn the field's attention to the organizational supports necessary for an intervention's success, including factors such as staff selection, structural supports, materials, staff and program evaluation, and facilitative administrative support (Fixsen, Naoom, Blasé, Friedman, & Walle, 2005; Hulleman & Cordray, 2009). Across all three studies and all five curricula, we identified one organizational support that facilitated implementation: regular common planning time for teachers. In all models, teachers had an easier time implementing when they had dedicated time in their schedule for drawing up instructional plans, preparing materials with their coteachers or other teachers, discussing individual children's progress, and brainstorming barriers and how to address them. Teachers engaged in common planning time on a daily basis in the Boston study and on a weekly basis in the MPC study. The ability to work flexibility and creatively within the schedule and system and to provide teachers and coaches with this time and space was often a marker for strong organizational support and commitment to the curriculum. Work examining the recent pilot expansion of the Boston public preschool model to community-based preschool settings in the city showed that the implementation of the model was relatively low (Yudron, Weiland, & Sachs, 2016). Most teachers lacked common planning in the community-based settings and cited it in interview as a key reason for this low-level implementation.

Another organizational support that we hypothesize may lead to better strongest hope implementation is trainings for principals on early childhood development and good early childhood practice. These trainings were part of the Boston model, as district staff found it critical to help principals understand why preschool classrooms were organized and run differently than classrooms for older children in their building. To be clear, this feature was present only in the Boston model, not in our other focal models that also achieved good implementation levels. In Boston, coaches reported that trainings helped shift the perceptions of principals such that they viewed children's engagement in centers and play as core learning activities, rather than as time wasted on nonproductive instruction. Working with principals and administrators in this way may be essential for effective strongest hope implementation but is rarely part of preschool interventions aimed at improving instructional quality. The aforementioned report on Boston's pilot expansion to community-based preschools cited a lack of administrator curricular understanding and engagement as a key barrier in replicating the model in those settings (Yudron et al., 2016).

Discussion

To date, our five focal implementations across 6,500 children and 750 lead teachers in 19 localities spanning urban, suburban, and rural settings and public school, private, and Head Start auspices have added to the evidence base on how to promote preschool instructional quality in five large-scale systems, using different versions of the strongest hope model. Our exploratory hypothesis-generating cross-study review of these five implementations identified five common features—specific instructional content, inclusion of highly detailed teacher scripts, incorporation of teacher voice, time for planning, and use of real-time data—that may help further flesh out the strongest hope model at large scale for localities considering this approach. A sixth feature—early childhood training for principals—was characterized in only one of the five implementations but rose to the top as a best bet for an important but nonshared feature. Notably, these features fit with the theoretical basis for the strongest hope model discussed earlier—for example, the principles of developmentally appropriate practice and adult learning. The focus on specific instructional content and use of real-time data, for example, connect to the emphasis on intentional teaching and the necessity of information on children's current skills to help identify progress toward learning goals in the principles of developmentally appropriate practice (National Association for the Education of Young Children, 2009). Inclusion of highly detailed scripts, incorporation of teacher voice, and time for planning all match adult learning principles of self-direction in learning, the role of reflection on problems that arise when knowledge is applied, and preference for job-relevant learning (Knowles et al., 2005; Michigan Association of Intermediate School Administrators General Education Leadership Network Early Literacy Task Force, 2016).

There were also divergent features across implementations that are important to note, as they might offer more flexible

features of the strongest hope model. The Boston coaching model, which was not manualized, was organic and flexible; coaches were free to focus on any weaknesses in a teacher's practice. In contrast, MPC and CARES included well-manualized coaching models with specific objectives, coaching activities, and foci. The staff responsible for leading training also ranged across the studies. In Boston, the district's instructional coaches led most teacher trainings, while developer-certified trainers provided up-front staff training in HS CARES. Curriculum developers supported by trained external partners led the MPC teacher trainings. The study team was also far more involved in creating the coaching systems in MPC and HS CARES versus Boston, where the district made all coaching structure and focus decisions. Importantly, each study had very different sets of teacher and coach characteristics—from highly professionalized educated teachers and coaches in Boston to a range of backgrounds in HS CARES. These differences were due to variation in contexts across studies, with Boston only in public schools, HS CARES only in Head Start programs, and MPC split between community-based and public school settings. At least for degree of coaching and training manualization, different roads may lead to success in changing teacher practices in the intended directions, depending on the levels of coach and teacher expertise in the applied context and on the focal developmental domain. Additional and systematic attention to the role of auspice in facilitating or impeding the implementation of a given model would provide needed practical guidance in the field. Presently, there is very little such work to guide decisions and models (Yudron et al., 2016).

Also, one of our common features—content—may be more crucial for mathematics, language, and literacy than for socioemotional skills, given that the latter develop less linearly and are more contextually dependent (Raver, 2004). Incredible Years, for example, had much less of a content focus than the curricula used in Boston, MPC, and Preschool PATHS: it includes some content, but it primarily focuses on teaching strategies to support teachers' language in extending children's skills. It still produced positive effects on teacher practices and targeted child outcomes (though its impacts were concentrated on outcomes that measured specific knowledge that children learned; Morris et al., 2014).

To be clear, we do not present any of these features, convergent or divergent, as causing any particular effect on teacher practice or child outcomes. Rather, we envision them as being useful to the field in several specific ways at this juncture in time. For research, they identify additional features to describe in preschool model evaluations. For example, if there was a coaching model, to what extent was there teacher voice and use of real-time data? These features also may be ripe for experimental manipulation and, thus, causal examination of these features in preschool impact evaluations. For example, if a new preschool program were launched in a given area, there could be systematic variation in and investigation of principal training in early childhood.

For localities, as they move forward on preschool quality and access, their implementation plans need to address not only theoretical issues (e.g., what the coaching philosophy of the district is or what child domains are most important to promote) but also concrete logistical issues (e.g., how scripted the curriculum will be or how to get teachers the time that they need in their schedule to plan for it). Our six features may help them better articulate these plans as well as help them choose among models. Three of the five features are also highly actionable: incorporation of teacher voice, use of real-time data, and early childhood training for principals. That is, localities could add them to strongest hope models that otherwise lack them.

Importantly, we focused in this article on the prekindergarten year only. Fadeout of intervention benefits is common across educational interventions, and in the preschool years, there is little understanding of the mechanisms behind this phenomenon (Bailey, Duncan, Odgers, & Yu, 2017). It could be that retaining a preschool boost requires a stronger kindergarten experience, more peers who also experienced enhanced preschool, or more sensitive measures in postprekindergarten follow-up (McCormick, Hsueh, Weiland, & Bangser, 2017). Two of our focal studies, HS CARES and MPC, did follow children into kindergarten. For CARES, the key outcomes for which there were effects in preschool were not collected in kindergarten (i.e., direct assessments of children's emotion knowledge and social problem-solving skills); however, there were few lasting benefits on the other teacher- and parent-reported outcomes (Morris et al., 2014). It is not clear if the lack of lasting benefits found was due to the more limited measurement or the kindergarten context. For MPC, the follow-up study is still underway, and analyses so far have revealed impacts on children's academic and executive functioning outcomes at the end of kindergarten (Mattera & Morris, 2017). Once results are available, MPC will provide information about what an enhanced boost for math in kindergarten (provided to a randomized group of children through math clubs) has for children's learning and development. Our focus in the present study was on active ingredients that may help localities get the most out of the preschool year. Active ingredients that best sustain the preschool boost through elementary school is an important topic for future research.

In close, beyond preschool, the present exploratory review exemplifies the wealth of knowledge generated from field-based trials that rarely make it into the evidence base, particularly in a synthesized format accessible to localities. The WWC (2016) currently shares extensive details on research design and program impacts to help practitioners choose effective programs; a current review of the 16 early childhood programs labeled as effective by the WWC demonstrates a wealth of information on research study design and impact study findings. Yet, the WWC provides little information about practical implementation successes and challenges that can inform the next steps for schools and districts—for

example, how to actually implement models so that they mirror the levels of fidelity observed in research trials. We echo the calls elsewhere for more research and information on implementation features and for a common repository for this information (Durlak, 2010; Harris, 2016). In the meantime, it is our hope that our cross-comparative synthesis represents one small step toward using the full range of learning from educational impact studies to improve opportunities for children.

Notes

1. What constitutes "play" is a long-standing debate in the field. Vygotsky, for example, limited his definition to pretend play only, excluding movement and manipulation of objects and explorations (Bodrova, 2008). We define "play-based learning" more broadly and in accord with Walsh, Sproule, McGuinness, and Trew (2011): playfulness as a characteristic of adult-child interactions, compatible with adult and child initiation of a given activity and with activities that have a specific learning goal.

2. Some global curricula, such as Creative Curricula, do include a set of activities that get progressively more difficult throughout the year. However, teachers have considerable choice about what they actually choose to implement, and specific activities do not need to take place in an established order. In these curricula, there is significant onus on the teacher to craft the actual scope and sequence. This is very different from a curriculum in which a specific place/day in the year has a specific plan—for example, Unit 6 Day 10 of the Opening the World of Learning curriculum specifies exactly what activities should occur. This is part of the fundamental distinction that experts (Yoshikawa et al., 2013) made in distinguishing "strongest hope" curricula from global or whole-child curricula. Accordingly, for the purposes of this article, we do not these include global curricula in our definition of curricula with an established scope and sequence.

3. Yoshikawa and colleagues (2015) did find positive effects of coaching and training not tied to a curriculum on improving the emotional support and organizational components of classroom quality.

4. The Opening the World of Learning (OWL) curriculum, as used in Boston, was summarized by Weiland and Yoshikawa (2013) in a study based on a pretest/posttest design with no control group and a sample of children in eight programs that implemented OWL, and it showed consistently positive associations with gains in students' language and literacy skills (Wilson, Morse, & Dickinson, 2009). However, a randomized controlled trial in Head Start centers (Dickinson, Kaiser, et al., 2011; Dickinson, Freiberg, & Barnes, 2011) revealed no impacts of OWL on children's language and literacy outcomes at the end of preschool. It also indicated some negative effects at the end of kindergarten and the end of first grade. Notably, the fidelity of implementation in the treatment groups was relatively low, and control classrooms had partially implemented OWL. Teachers were also on average better educated in the eight programs showing positive effects than in the randomized controlled trial (65% vs. 17%, respectively).

References

Bailey, D., Duncan, G. J., Odgers, C. L., & Yu, W. (2017). Persistence and fadeout in the impacts of child and adolescent interventions. *Journal of Research on Educational Effectiveness*, *10*(1), 7–39.

Barnett, W. S., Carolan, M. E., Squires, J. H., Clarke Brown, K., & Horowitz, M. (2015a). *Appendix A: State survey data, 2014*. New Brunswick, NJ: National Institute for Early Education Research.

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

Barnett, W. S., Jung, K., Yarosz, D. J., Thomas, J., Hornbeck, A., Stechuk, R., & Burns, S. (2008). Educational effects of the Tools of the Mind curriculum: A randomized trial. *Early Childhood Research Quarterly*, *23*(3), 299–313.

Bloom, H., & 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.

Bodrova, E. (2008). Make-believe play versus academic skills: A Vygotskian approach to today's dilemma of early childhood education. *European Early Childhood Education Research Journal*, *16*(3), 357–369.

Chambers, B., Cheung, A. C., & Slavin, R. E. (2016). Literacy and language outcomes of comprehensive and developmental-constructivist approaches to early childhood education: A systematic review. *Educational Research Review*, *18*, 88–111.

Chaudry, A., Morrissey, T., Weiland, C., & Yoshikawa, H. (2017). *Cradle to kindergarten: A new plan to combat inequality*. New York, NY: Russell Sage.

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.

Clements, D. H., Sarama, J. H., Spitler, M. E., Lange, A. A., & Wolfe, C. B. (2011). Mathematics learned by young children in an intervention based on learning trajectories: A large-scale cluster randomized trial. *Journal for Research in Mathematics Education*, *4*, 127–166.

Cohen, D. K., & Ball, D. L. (2001). Making change: Instruction and its improvement. *Phi Delta Kappan*, *83*(1), 73–77.

Cohen, E., & Kaufmann, R. K. (2000). *Early childhood mental health consultation*. Rockville, MD: Center for Mental Health Services.

Collins, A., Brown, J. S., & Holum, A. (1991). Cognitive apprenticeship: Making thinking visible. *American Educator*, *15*(3), 6–11.

Connect4Learning. (n.d.). *The prekindergarten curriculum: Interdisciplinary teaching across domains*. Retrieved from http://www.c4lcurriculum.com/research/whitepapers/

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.

Dickinson, D. K., Freiberg, J. B., & Barnes, E. (2011). Why are so few interventions really effective? A call for fine-grained research methodology. In S. B. Neuman, & D. K. Dickinson (Eds.), *Handbook of early literacy research* (Vol. 3, pp. 337–357). New York, NY: Guilford Press.

Dickinson, D. K., Kaiser, A., Roberts, M., Hofer, K. G., Darrow, C. L., & Griffenhagen, J. B. (2011, March). *The effects of two language focused preschool curricula on children's achievement through first grade*. Paper presented at the Society for Research in Educational Effectiveness, Washington, DC.

Domitrovich, C. E., Cortes, R. C., & Greenberg, M. T. (2007). Improving young children's social and emotional competence: A randomized trial of the preschool "PATHS" curriculum. *Journal of Primary Prevention*, *28*(2), 67–91.

Duncan, G., Jenkins, J., Auger, A., Bitler, M., Domina, T., & Burchinal, P. (2017). *Boosting school readiness skills: Should preschools target skills or the whole child?* Working paper.

Durlak, J. A. (2010). The importance of doing well in whatever you do: A commentary on the special section "Implementation research in early childhood education." *Early Childhood Research Quarterly*, *25*(3), 348–357.

Durlak, J. A., & DuPre, E. P. (2008). Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting the implementation. *American Journal of Community Psychology*, *41*, 327–350.

Farran, D., & Wilson, S. (2014). *Achievement and self-regulation in pre-kindergarten classrooms: Effects of the Tools of the Mind curriculum*. Working paper.

Fixsen, D. L., Naoom, S. F., Blasé, K. A., Friedman, R. M., & Wallace, F. (2005). *Implementation research: A synthesis of the literature* (FMHI Publication No. 231). Tampa, FL: Louis de la Parte Florida Mental Health Institute.

Gwande, A. (2009). *The checklist manifesto: How to get things right*. London, UK: Picador.

Harris, D. (2016). *Improving the federal role in education research*. Retrieved from https://www.brookings.edu/blog/brown-center-chalkboard/2016/12/23/memo-improving-the-federal-role-in-education-research/

Head Start Performance Standards, 45 CFR § 13 (2016).

Honig, M. I., & Ikemoto, G. S. (2008). Adaptive assistance for learning improvement efforts: The case of the Institute for Learning. *Peabody Journal of Education*, *83*(3), 328–363.

Hulleman, C. S., & Cordray, D. S. (2009). Moving from the lab to the field: The role of fidelity and achieved relative intervention strength. *Journal of Research on Educational Effectiveness*, *2*(1), 88–110.

Institute of Medicine and National Research Council. (2015). *Transforming the workforce for children birth through age 8: A unifying foundation*. Washington, DC: The National Academies Press.

Jenkins, J. M., & Duncan, G. J. (2017). *Do prekindergarten curricula matter?* Washington, DC: Brookings Institute. Retrieved from https://www.brookings.edu/wpcontent/uploads/2017/04/duke_prekstudy_final_4-4-17_hires.pdf

Joyce, B., & Showers, B. (1980). Improving inservice training: The messages of research. *Educational Leadership*, *37*(5), 379–85.

Joyce, B., & Showers, B. (1982). The coaching of teaching. *Educational Leadership*, *40*(1), 4–10.

Joyce, B. R., & Showers, B. (2002). *Student achievement through staff development*. New York, NY: Longman.

Knowles, M. S., Holton, E. F., & Swanson, R. A. (2005). *The adult learner*. Burlington, MA: Elsevier.

Kraft, M. A., Blazar, D., & Hogan, D. (2016). *The effect of teacher coaching on instruction and achievement: A meta-analysis of the causal evidence*. Working paper.

Kuperschmidt, J. B., Bryant, D., & Willoughby, M. (2000). Prevalence of aggressive behaviors among preschoolers in Head Start and community child care programs. *Behavioral Disorders*, *26*, 42–52.

Markussen-Brown, J., Juhl, C. B., Piasta, S. B., Bleses, D., Højen, A., & Justice, L. M. (2017). The effects of language-and literacy-focused professional development on early educators and children: A best-evidence meta-analysis. *Early Childhood Research Quarterly*, *38*, 97–115.

Marsh, J. A. (2012). Interventions promoting educators' use of data: Research insights and gaps. *Teachers College Record*, *114*(11), 1–48.

Mattera, S. K., Lloyd, C. M., Fishman, M., & Bangser, M. (2012). *A first look at the Head Start CARES demonstration: Large-scale implementation of programs to improve children's social-emotional competence* (OPRE Report No. 2013-47). Washington, DC: Office of Planning, Research and Evaluation.

Mattera, S. K., & Morris, P. (2017). *Making pre-K count and high 5s: Preliminary kindergarten impacts*. New York, NY: MDRC.

McCormick, M., Hsueh, J., Weiland, C., & Bangser, M. *The challenge of sustaining preschool impacts*. New York, NY: MDRC. Retrieved from https://www.mdrc.org/sites/default/files/ExCEL_SustainingPreschoolImpacts.REV_.pdf

Michigan Association of Intermediate School Administrators General Education Leadership Network Early Literacy Task Force. (2016). *Essential coaching practices for elementary literacy*. Retrieved from http://www.gomaisa.org/sites/default/files/Essential_Coaching_Practices_12_8_16.pdf#overlay-context=general-education-leadership-network

Morris, P., Connors, M., Friedman-Krauss, A., McCoy, D., Weiland, C., Feller, A., . . . Yoshikawa, H. (2017). *New findings on impact variation from the Head Start Impact Study: Informing the scale-up of early childhood programs*. Working paper.

Morris, P., Mattera, S. K., Castells, N., Bangser, M., Bierman, K., & Raver, C. (2014). *Impact findings from the Head Start CARES Demonstration: National evaluation of three approaches to improving preschoolers' social and emotional competence. Executive summary* (OPRE Report No. 2014-44). Washington, DC: Office of Planning, Research and Evaluation.

Morris, P. A., Mattera, S. K., & Maier, M. F. (2016). *Making pre-K count: Improving math instruction in New York City*. New York, NY: MDRC.

National Association for the Education of Young Children. (2009). *Developmentally appropriate practice in early childhood programs serving children from birth through age 8: Position statement*. Retrieved from https://www.naeyc.org/files/naeyc/file/positions/position%20statement%20Web.pdf

National Center on Quality Teaching and Learning. (2015). *Preschool curriculum consumer report*. Retrieved from https://eclkc.ohs.acf.hhs.gov/sites/default/files/pdf/curriculum-consumer-report.pdf

Nguyen, T. (2016). *Impacts of targeted and global preschool curricula on children's school readiness: A meta-analytic review*. Working paper.

Office of Head Start. (2015). *Head Start program facts for 2014*. Retrieved from https://eclkc.ohs.acf.hhs.gov/hslc/data/factsheets/2014-hs-program-factsheet.html

Phillips, D., Austin, L. J., & Whitebrook, M. (2016). The early care and education workforce. *The Future of Children*, *26*(2), 139–158.

Phillips, D., Johnson, A., Weiland, C., & Hutchison, J. E. (2017). *Public preschool in a more diverse America: Implications for next-generation evaluation research*. Ann Arbor, MI: Poverty Solutions.

Phillips, D., Lipsey, M., 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. Washington, DC: Brookings Institution. https://www.brookings.edu/wp-content/uploads/2017/04/consensus-statement_final.pdf

Pianta, R., Downer, J., & Hamre, B. (2016). Quality in early education classrooms: Definitions, gaps, and systems. *The Future of Children*, *26*(2), 119–137.

Piasta, S. B., Justice, L. M., O'Connell, A. A., Mauck, S. A., Weber-Mayrer, M., Schachter, R. E., . . . Spear, C. F. (2017). Effectiveness of large-scale, state-sponsored language and literacy professional development on early childhood educator outcomes. *Journal of Research on Educational Effectiveness*, *10*, 354–378.

Public Opinion Strategies and Hart Research. (2014). *Key findings from a national survey of 800 registered voters conducted May 28–June 1, 2014*. Retrieved from http://growamericastronger.org/wp-content/uploads/2014/06/FINAL_FFYF-National-Results_071414.pdf?a63b61

Raver, C. C. (2004). Placing emotional self-regulation in sociocultural and socioeconomic contexts. *Child Development*, *75*(2), 346–353.

Schickedanz, J., & Dickinson, D. (2005). *Opening the world of learning*. Iowa City, IA: Pearson.

Tout, K., Halle, T., Zaslow, M., & Starr, R. (2011). *Evaluation of the early childhood educator professional development program: Final report*. Washington, DC: Child Trends.

Walsh, G., Sproule, L., McGuinness, C., & Trew, K. (2011). Playful structure: A novel image of early years pedagogy for primary school classrooms. *Early Years*, *31*(2), 107–119.

Wasik, B. A., Mattera, S. K., Lloyd, C. M., & Boller, K. (2013). *Intervention dosage in early childhood care and education: It's complicated*. Princeton, NJ: Mathematica Policy Research.

Webster-Stratton, C., Reid, M. J., & Hammond, M. (2001). Preventing conduct problems, promoting social competence: A parent and teacher training partnership in Head Start. *Journal of Clinical Child Psychology*, *30*(3), 283–302.

Webster-Stratton, C., Reid, M. J., & Hammond, M. (2004). Treating children with early-onset conduct problems: Intervention outcomes for parent, child, and teacher training. *Journal of Clinical Child and Adolescent Psychology*, *33*(1), 105–124.

Weiland, C. (2016). Launching Preschool 2.0: A road map to high-quality public programs at scale. *Behavioral Science & Policy*, *2*(1), 37–46.

Weiland, C., Ulvestad, K., Sachs, J., & Yoshikawa, H. (2013). Associations between classroom quality and children's vocabulary and executive function skills in an urban public prekindergarten program. *Early Childhood Research Quarterly*, *28*(2), 199–209.

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.

What Works Clearinghouse. (2016). *The Creative Curriculum for preschool: Fourth edition*. Retrieved from https://ies.ed.gov/ncee/wwc/Docs/InterventionReports/wwc_creativecurriculum_030513.pdf

Wilson, D. B., & Lipsey, M. W. (2001). The role of method in treatment effectiveness research: Evidence from a meta-analysis. *Psychological Methods*, *6*, 413–429.

Wilson, S. J., Morse, A. B., & Dickinson, D. K. (2009). *Examining the effectiveness of OWL as used in ERF projects: Final report of results from the OWL Consortium Project*. Nashville, TN: Vanderbilt University Center for Evaluation Research and Methodology.

Yoshikawa, H., Leyva, D., Snow, C. E., Treviño, E., Barata, M., Weiland, C., . . . Arbour, M. C. (2015). Experimental impacts of a teacher professional development program in Chile on preschool classroom quality and child outcomes. *Developmental Psychology*, *51*(3), 309.

Yoshikawa, H., Weiland, C., & Brooks-Gunn, J. (2016). When does preschool matter? *The Future of Children*, *26*(2), 21–35.

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

Yudron, M., Weiland, C., & Sachs, J. (2016). *BPS K1DS: Piloting the Boston Public Schools' prekindergarten model in community-based organizations*. Boston, MA: Boston Public Schools.

Zaslow, M., Tout, K., Halle, T., Whittaker, J. V., & Lavelle, B. (2010). *Toward the identification of features of effective professional development for early childhood educators*. Washington, DC: Office of Planning, Evaluation and Policy Development.

Authors

CHRISTINA WEILAND is an assistant professor at the University of Michigan. Her research focuses on the effects of early interventions and policies on young children in poverty.

MEGHAN MCCORMICK is a research associate at MDRC. Her work uses experimental and quasi-experimental approaches to estimate the impacts of school- and center-based programs and policies on low-income children's academic, behavioral, social, and emotional outcomes.

SHIRA MATTERA is a research associate at MDRC. Her research focuses on early childhood education and intervention and children's development.

MICHELLE MAIER is a research associate at MDRC. Her research focuses on enhancing young children's preschool experience so that they are better prepared for school both academically and emotionally.

PAMELA MORRIS is a vice dean for Research and Faculty Affairs and professor of applied psychology at New York University. Her research is characterized by the study of theoretically-informed interventions, strong attention to measurement of developmental outcomes for children, and cutting-edge analytic strategies on causal inference and strong research designs.