Literacy and Language Outcomes of Comprehensive and Developmental-Constructivist Approaches to Early Childhood Education:
A Systematic Review

Bette Chambers
University of York

Alan C. K. Cheung
The Chinese University of Hong Kong

Robert E. Slavin
Johns Hopkins University

March, 2016

Abstract

This systematic review of research on early childhood programs seeks to identify effective approaches capable of improving literacy and language outcomes for preschoolers. It applies consistent standards to determine the strength of evidence supporting a variety of approaches, which fell into two main categories: *comprehensive approaches*, which include phonemic awareness, phonics, and other skills along with child-initiated activities, and *developmental-constructivist approaches* that focus on child-initiated activities with little direct teaching of early literacy skills. Inclusion criteria included use of randomized or matched control groups, evidence of initial equality, a minimum study duration of 12 weeks, and valid measures of literacy and language. Thirty-two studies evaluating 22 programs found that comprehensive early childhood programs that have a balance of skill-focused and child-initiated activities programs had significant evidence of positive literacy and language outcomes at the end of preschool and on kindergarten follow-up measures. Effects were smaller and not statistically significant for developmental-constructivist programs.

Key words/phrases: early childhood programs, preschool teaching methods, emergent literacy, Head Start, systematic reviews

Literacy and Language Outcomes of Comprehensive and Developmental-Constructivist Approaches to Early Childhood Education: A Systematic Review

1. Introduction

Early childhood education, particularly preschool education for three- and four-year-olds, can have a lasting impact on the educational success and life chances of disadvantaged children. Numerous reviews of longitudinal studies have found that in comparison to children who do not attend preschool at all, high-quality preschool experiences have strong impacts on cognitive outcomes at the end of preschool, and these effects can last into elementary and secondary school and beyond (e.g., Authors, 2006; Authors, 2013; Camilli, Vargas, Ryan, & Barnett, 2009; Coghlan et al., 2009; Gorey, 2001; Jacobs, Creps, & Boulay, 2004; Nelson, Westhues, & MacLeod, 2003; Waldfogel & Washbrook, 2010). Even though the types of programs with long-lasting impacts are far more intensive and extensive than ordinary early childhood programs, these long-term impacts for very disadvantaged children make intensive programs highly cost-effective in the long run (Carneiro & Heckman, 2003; Heckman & Masterov, 2007; Karoly, Kilburn, &Cannon, 2005). In particular, programs like the Perry Preschool (Schweinhart & Weikart, 1997) and the Abecedarian Project (Ramey & Ramey, 1998) are frequently held up as models of what early childhood education could be.

The evaluation of these and other intensive models are important in demonstrating that disadvantage can be overcome with intensive and comprehensive intervention, involving children, families, health, and social factors (see Gorey, 2001). However, evaluations of Head Start in the U.S., Sure Start in the U.K., and other run-of-the-mill preschool models have shown strong immediate impacts but not the lasting effects documented for some of the intensive models (Camilli et al., 2009; Melhuish, Belsky, & Leyland, 2010).

For example, a large-scale national randomized evaluation of Head Start (Puma, Bell, Cook, & Heid, 2010) found positive impacts on several measures at the end of the preschool year that were fading by the end of kindergarten and gone by the end of first grade. This pattern closely matches findings from the recent large-scale randomized evaluation of the Tennessee Voluntary Prekindergarten Program (Lipsey, Hofer, Dong, Farran, & Bilbrey, 2013 a, b), a long-term evaluation of early childhood programs in Australia (Claessens, & Garrett, 2014), a follow-up to age 5 of Sure Start in the U.K. (Melhuish et al., 2010) and long-term studies in England of preschool programs for three year olds (Belsky & Melhuish, 2007; Brewer, Cattan, Crawford, & Rabe, 2014), as well as many earlier studies (see Karweit, 1994).

The disappointing findings of these evaluations of ordinary preschool programs, coupled with political support for expanding and improving preschool experiences, have led to a search for pragmatic strategies capable of improving immediate and lasting impacts of preschool for disadvantaged children. In particular, interest has moved from studies of preschool vs. no preschool to evaluations of alternative approaches to preschool education (Diamond, Justice, Siegler, & Snydor, 2013). Previous reviews (e.g., Camilli et al., 2009; Jacob et al., 2004) have reviewed effects of alternative programs, but only cited studies published up to 2000. A great

3

deal of experimental research comparing different preschool approaches has taken place since 2000. This review considers the evidence on alternative strategies for use in preschool in terms of language and literacy outcomes, measured at the end of the preschool and kindergarten years, synthesizing findings of studies of contrasting approaches that were published from 1990 to 2015.

1.1 Theoretical and Historical Background

From the beginning, preschools have had a primary emphasis on socialization and general cognitive development. In the transition from home to school, children have long been encouraged to play, sing, build with blocks, and do art, dress-up, and drama. The theories of Piaget (1952) and Vygotsky (1987) strongly reinforced the idea that cognitive and social development were the appropriate goals of early childhood education and that self-chosen activities, interactions among children, and experience with make-believe, construction, art, and music were the key to cognitive and social development. Approaches of this kind are called "inquiry-based" or as using "constructivist principles" by Camilli et al. (2009) and "developmentally appropriate" by the National Association for the Education of Young Children (Neuman, Copple, & Bredekamp, 1999) and many others. We combine these concepts and use the term "developmental-constructivist" to refer to such programs. This term is intended to reflect the expectations of the NAEYC standards and the Early Childhood Environment Rating Scale, or ECERS (Harms, Clifford, & Cryer, 2014) both of which discourage teacher-directed instruction or direct teaching of pre-reading skills.

In the 1960s, as part of the behaviorist movement that was asserting itself at the time, Bereiter & Engelmann (1966) threw down a major challenge to the longstanding traditions of early childhood education. In their view, the goal of early education is not only to stimulate intellectual development, but to explicitly teach reading, math, and other skills. They pointed to the findings of early evaluations of preschool provision (compared to no preschool) in which IQ scores were increased at the end of preschool but fell back to pretest levels a year or two later. Rather than worrying about IQ, they argued, go directly to the target skills. They created and evaluated in a small experiment an approach later called Direct Instruction (DI), in which preschoolers were taught in groups of 4-7. Children participated in structured lessons on reading, arithmetic, and language concepts, and then engaged in semi-structured writing, drawing, reading-readiness, music, and snack activities. Teachers were asked to use very specific methods and very simple, direct language. The children in the initial Bereiter & Engelmann (1966) study achieved remarkable gains, on average, over a two-year period (ages 4-5), both on an IQ measure and on measures of reading, arithmetic, and spelling.

The Bereiter and Engelmann study, and subsequent research, set off an extraordinary debate, with strong feelings on both sides. David Weikart (1995) and Lawrence Schweinhart (Schweinhart & Weikart, 1997) demonstrated long-term positive impacts of a developmental-constructivist model, the Perry Preschool, in comparison to no preschool experience. However, the Perry Preschool was an extremely intensive approach that could not be replicated in ordinary schools. As noted earlier, long-term evaluations of ordinary Head Start and other run-of-the-mill

4

programs, which generally implement developmental-constructivist approaches, have found that initial effects tend to fade out within a few years (Camilli, Vargas, Ryan, & Barnett, 2009; Lipsey et al., 2013a, b; Puma, et al., 2010).

Until the 1990s, the debate about early childhood was dominated by the differing claims of the two extreme positions. In practice and in academia, the developmental-constructivist argument clearly prevailed, even though comparisons of alternative approaches up to 2000 by Camilli et al. (2009) and Nelson et al. (2003) concluded that programs with an element of direct teaching were associated with superior outcomes. Camilli et al. (2009) noted that "many early childhood educators would be surprised by this finding in light of the field's consensus that a developmentally appropriate approach…is not one in which children are drilled in basic concepts" (p. 599; also see Jacob et al., 2004). Head Start centers and preschool programs in public schools overwhelmingly used, and most still use, programs consistent with the developmental-constructivist view, and DI has been relegated to a fringe position.

New approaches to early childhood education began to appear in the 1990s. These comprehensive approaches balance the teaching of early literacy skills with the language and socialization skills of developmental-constructivist models. In some cases developers have created complete preschool models that incorporate teaching of phonemic awareness, phonics, alphabet, writing, and math, with traditional creative play, art, music, drama, and story time. Examples include *Curiosity Corner* (Authors, 2001), *ELLM* (Cosgrove et al., 2006), and *Breakthrough to Literacy* (www.breakthroughtoliteracy.com). In other cases, developers have created supplemental skills-based approaches that add to any developmental model well-planned activities focused on literacy, language, and sometimes numeracy objectives. Taken together, the supplement and the developmental-constructivist activities form a comprehensive approach. One example is *DLM Early Childhood Express,* a developmental-constructivist approach, plus *Open Court Reading Pre-K,* a literacy-focused model (Lonigan et al., 2011). Comprehensive programs vary a great deal, but most teach literacy skills primarily through rhymes, songs, games, and interactive reading with children.

Only a few reviews in the past decade have made comparisons among different types of preschool interventions in terms of language and literacy outcomes (Authors, 2006; Camilli et al., 2009). The review by Authors (2006) compared traditional, academic, and cognitive-developmental early childhood programs, and found that academic programs generally produced better immediate and mid-term cognitive outcomes. However, a small set of studies of the Perry Preschool, a developmental-constructivist approach, found this model to produce better long-term educational and social adjustment outcomes. Camilli et al. (2009) compared what they called direct instruction and inquiry-based teaching in studies through 2000, and found better effects for the programs emphasizing direct instruction. Note that this does not refer to the Direct Instruction program, but to all programs that directly teaching reading and language skills in preschool.

In a meta-analysis of the effects of early childhood curricula limited to children's receptive and expressive vocabulary, Darrow (2009) evaluated 17 early childhood curricula and

concluded that taken together, programs did not differ from their respective control groups on vocabulary development by the end of preschool, nor at the end of kindergarten. Nor could she determine the impacts of particular programs.

1.2 Research Questions

The purpose of this article is to compare the evidence for the effectiveness of alternative preschool programs provided in group settings for children's language and literacy outcomes for disadvantaged children. The scope of the review includes preschool approaches designed for use in regularly scheduled group programs that educators might consider adopting to prepare their children for success in elementary school and beyond. It focuses on the main approaches teachers and schools might emphasize, not on smaller targeted interventions such as shared reading (e.g., Mason, Kerr, Sinha, & McCormick, 1990; Piasta, Justice, McGinty, & Kaderavek, 2012; Wasik & Bond, 2001; What Works Clearinghouse, 2015; Whitehurst et al., 1994), and vocabulary development (e.g., Neuman, Newman, & Dwyer, 2011; Pollard-Duradola et al., 2011).

Specific research questions are as follows:

1. What are the average effect sizes on language and literacy measures of attending preschools using:

   a) Comprehensive approaches (vs. controls)
      -or-
   b) Developmental-constructivist approaches (vs. controls)?

2. What are the lasting effects, through the end of kindergarten, of attending comprehensive vs. developmental-constructivist preschools?

The present review places the findings of studies of early childhood programs intended to enhance school readiness on a common scale, to provide educators and policy makers with meaningful, unbiased information that they can use to select approaches most likely to benefit their children's school readiness. It also updates the evidence, particularly in light of the findings of an extraordinary set of studies, funded by the U. S. Department of Education, called the Preschool Curriculum Evaluation Research (PCER, 2008), as well as other research that has added substantially to the number and quality of studies of outcomes of alternative preschool approaches.

To make the review most useful to educators and policy makers, it emphasizes large studies done over significant time periods that used standard individually-administered measures. Such studies generally evaluate programs as they are used in practical, larger-scale implementations, rather than in the hothouse conditions characteristic of the Perry Preschool evaluation (Schweinhart, Barnes, & Weikart, 1993) and the Abecedarian Project (Ramey &

Ramey, 1998), among others. It also identifies common characteristics of programs likely to make a difference in children's literacy and language outcomes.

1.3 Methodological Issues Unique to Early Childhood Education

There are several problems characteristic of research on child outcomes of early childhood programs that are important to understand. First, many outcomes of early education are difficult to measure with young children, so it may be that impacts of a given approach may not be detected at the end of a four-year-old program but might show up on related measures a year or two later, not because of a "sleeper effect" but because a true but difficult-to-measure impact became measurable in later years. For example, difficult-to-measure impacts on general vocabulary might show up in reading comprehension or reading vocabulary assessments in the primary grades. As one example of this, a five-year study by Lipsey, Farran, Hurley, Hofer, and Bilbrey (2009) randomly assigned preschools to *Bright Beginnings*, a comprehensive approach, *Creative Curriculum*, a developmental-constructivist approach, or control conditions for one year and found modest positive literacy effects for *Bright Beginnings* (ES = +0.18) and none for *Creative Curriculum* (ES = -0.11). Yet third-grade state reading tests for the children remaining in the same schools showed positive follow-up effects for *Bright Beginnings* (ES = +0.27) and *Creative Curriculum* (ES = +0.16). Note that these effect sizes are computed as experimental-control posttest differences adjusted for pretests.

Secondly, studies of early childhood programs are particularly susceptible to bias due to use of measures inherent to the experimental treatment, or overly aligned with the treatment group's objectives but not the control group's objectives (see Authors, 2011, on this topic). For example, imagine that an experimental treatment for four-year-olds emphasizes a specific list of vocabulary words, and then the assessment consists of a subset of these words, which the treatment group would have heard far more often than the control group. As one example, Neuman, Newman, & Dwyer (2011) evaluated a vocabulary intervention called *World of Words*, or WOW. Averaging across three units over the course of a year, the mean effect size for "word knowledge" for a subset of target words was +0.40 ($p<.001$). However, on the treatment-independent Woodcock-Johnson Picture Vocabulary test, the difference was an effect size of +0.07 (n.s.). In a study by Pollard-Duradola et al. (2011), the effects of a vocabulary intervention on a researcher-developed receptive vocabulary measure focusing on target words was +1.56. On PPVT, a treatment-independent measure, the effect size was +0.09. Such treatment-inherent measures are excluded in this review.

Finally, preschool measures are always administered individually. Individual assessment can create opportunities for bias, especially if testers are the children's own teachers or other school staff who would be aware of the child's treatment assignment and might have motivations to make the program or their class or school look good on tests. For this reason, studies were rejected from this review if testing was done by the children's teachers or other persons who were not independent of the school or program.

2. Method

This review uses a form of meta-analysis called best-evidence synthesis, designed for use in reviewing literatures in which there are relatively few studies evaluating each of many programs (see Author, 2008). Best-evidence syntheses apply consistent standards to identify unbiased, meaningful information from experimental studies, discuss the evidence for each program, and pool effect sizes across studies in substantively justified categories. The method uses standard meta-analytic techniques (Cooper, 1998; Lipsey & Wilson, 2001), with adaptations described later in this section.

2.1 Search Procedures

We conducted an exhaustive initial search to locate all studies that took place from 1990 to 2015 which compared child learning outcomes of alternative approaches to early childhood education. The 1990 start date reflects the changes that have taken place in early childhood education, as preschool programs have become far more widespread and more focused on literacy, and as many innovative preschool approaches have been developed and evaluated. Studies from all countries were included, as long as the studies were available in English. In practice, all qualifying studies were from the U.S. Studies published in refereed journals, technical reports, dissertations, or unpublished papers, were all included. Electronic searches were made of educational databases (e.g., JSTOR, ERIC, EBSCO, Psych INFO, NHS EED, Dissertation Abstracts) and web-based repositories (e.g., Google Scholar). Search terms used different combination of key words (e.g., preschool, nursery, prekindergarten, compensatory education, school readiness, child care, low income, disadvantaged, social deprivation, high poverty) and program names (e.g., High Scope, Creative Curriculum, Abecedarian, Perry Preschool, Montessori, Reggio Emilia, Direct Instruction, Tools of the Mind, Project Approach, Project Construct). In addition to thorough electronic searches, manual searches of the major education journals were conducted: *American Educational Research Journal, British Educational Research Journal, Early Childhood Research Quarterly, Child Development, Contemporary Issues in Early Childhood, Early Education and Development, Reading and Writing, Journal of Educational Psychology, Review of Educational Research,* and *Educational Evaluation and Policy Analysis*). References in reviews and primary research articles were followed up.

All potentially relevant papers were retrieved. Data were extracted and coded by one reviewer using a standard procedure and were checked by another reviewer. Disagreements were resolved by discussion and consensus and, if necessary, a third reviewer was consulted. The search yielded 1,698 articles; of these, 32 experimental-control comparisons evaluating 22 different programs met the inclusion criteria described in the following section (each experimental-control comparison is referred to hereafter as a "study.").

2.2 Inclusion Criteria

The studies evaluated educational programs for groups of children between the ages of 3 and 5, in the year before they begin kindergarten. The studies compared children taught in classes using a given program to those using an alternative program or business as usual. All control groups used developmental-constructivist models, either a specific program such as *Creative Curriculum* or *High/Scope*, or a teacher-designed approach. Studies that only compared preschool attendance to non-attendance were not included. Any early childhood setting that offered a regularly scheduled educational program to a group of preschoolers was included except for those that only provided narrow, time-limited supplements to developmental programs, such as shared reading interventions (Mason, et al., 1990; Piasta, et al., 2012) or vocabulary interventions (e.g., Neuman, Newman, & Dwyer, 2011), as noted earlier. Only four studies of shared book reading and two of vocabulary interventions would have met the inclusion criteria. Other inclusion criteria were as follows.

1. Initial equivalence. Random assignment or matching with appropriate adjustments for any pretest differences (e.g., analyses of covariance) had to be used. Studies with differences of more than 50% of a standard deviation on a pretest were excluded because, even with analyses of covariance, large pretest differences cannot be adequately controlled for (Shadish, Cook, & Campbell, 2002).

Studies without control groups, such as pre-post comparisons, were excluded. Studies in which parents selected their children into treatments (e.g., chose to attend a particular preschool program) were excluded. Regression discontinuity studies that permitted comparisons of alternative preschool approaches would have been included, but none were found. Studies that compared innovative preschool treatments to no preschool were excluded (e.g., Gormley, Phillips, & Gayer, 2008; Weiland & Yoshikawa, 2013; Wilson, Dickinson, & Rowe, 2013).

2. Sample size. Studies needed to have least 2 teachers and 25 children in each condition in the analysis. The purpose of the two-teacher requirement was to reduce confounding of teachers and treatments. The sample size requirement was intended to reduce confounding with class effects, and to reduce the "small study size effect" in which very small studies have been found to greatly inflate reported program impacts (Authors, 2009).

3. Outcome measures. The dependent measures included quantitative measures of literacy (e.g., alphabet knowledge, phonemic awareness, phonics, and concepts of print) and language (e.g., expressive and receptive language). Experimenter-made measures were accepted only if it could be determined that they assessed skills equally addressed in the control groups as well as the experimental groups.

Measures of objectives inherent to the intervention were excluded, for reasons discussed previously. Also excluded were measures in which the children's teachers rated cognitive skills or behaviors or administered the outcome measures. Teachers in the treatment groups might have

had their perceptions of the children's skills or behaviors influenced by their knowledge of being in a study and knowing the goals of the intervention.

We included studies that followed children into kindergarten, or further into elementary school, and measured children's language or literacy outcomes.

4. Duration. A minimum study duration of 12 weeks was required, to focus the review on practical programs and practices intended for extended use. Very brief studies often create artificial conditions that could not be maintained over time. In practice, almost all studies that met the 12-week criterion implemented for the whole preschool year.

Sometimes the impacts of an intervention become more apparent well after the immediate posttest. This is especially true for literacy outcomes, because reading is not generally assessed in preschool, although gains in vocabulary or other cognitive skills can have later effects on reading. For this reason, in the summary table (Table 3), we report outcomes for the end of preschool, the end of kindergarten, and later in elementary school (if available), and put particular emphasis on findings as of the end of kindergarten or later.

5. Replicability. Studies were excluded if they evaluated interventions that could not be broadly replicated as they were implemented, because they depend on resources far beyond those that other schools could provide. Examples are programs with extremely small class sizes, exceptional numbers of staff, locations in university lab schools with many researchers and students in daily participation, and so on. Studies in which the researcher or graduate students delivered the treatment were excluded (e.g., Byrne & Fielding-Barnsley, 1993). Such studies may be of theoretical interest, but are not relevant to practice.

6. High-poverty schools. Studies here were included only if they described their sample as high-poverty, or low socioeconomic status.

2.3 Computation of Effect Sizes

In general, effect sizes were computed as the difference between experimental and control individual child posttests after adjustment for pretests and other covariates, divided by the unadjusted posttest control group standard deviation (Lipsey & Wilson, 2001). If there were multiple studies of a given program, effect sizes were pooled across studies for the program, and effect sizes were also pooled across studies for various categories of programs. This pooling used means weighted by the final sample sizes to maximize the importance of large studies, as small studies tend to overstate effect sizes (see Authors, 2009, 2016; Rothstein, Sutton, & Borenstein, 2005). Effect sizes were categorized as measures of language (including receptive and expressive vocabulary) and literacy (including phonological awareness, print knowledge, letter and word identification, and spelling).

1.2.6 Categories of Research Design

Three categories of research designs were included in this review. *Randomized experiments* were those in which children, classes, or schools were randomly assigned to treatments, and data analyses were at the level of random assignment. In a *cluster randomized* design, when classes, teachers, or schools were randomly assigned to experimental and control conditions, analysis should be at the class or school level, usually using hierarchical linear modeling (HLM; Raudenbush & Bryk, 2002). When schools, teachers, or classes were randomly assigned but there were too few schools or classes to justify analysis at the level of random assignment, the study was categorized as a *randomized quasi-experiment* (Author, 2008). *Quasi-experimental control group* studies were ones in which experimental and control groups were matched on key variables at pretest, before posttests were known. Studies using fully randomized designs are preferable to randomized quasi-experiments, but all randomized experiments are less subject to bias than matched studies.

3. Results

Key study characteristics and child outcomes are summarized in the following sections. Table 1 and Table 2 present the effect sizes for all measures in qualifying studies, as well as averages across language and literacy measures in preschool, kindergarten, and beyond. Tables 1 and 2 also include information on study demographics, sample sizes, nature of control groups, and other details. Where appropriate data were available from two or more studies of similar interventions, program means weighted by sample sizes were computed. Descriptions of the experimental and control treatments and outcomes appear in the text along with any unique or remarkable aspects of the design, measures, or findings, but otherwise information appearing in Tables 1 or 2 is not repeated in the narrative.

=====
Tables 1 and 2 Here
====

3.1 Program Categories

To facilitate understanding and discussion, the 22 programs that had at least one qualifying study were organized in categories according to their main intentions, focus, and characteristics. The programs fell into two main categories: *comprehensive* and *developmental-constructivist*. These are described below.

Decisions about where to place each program were based on information on websites and in the studies that were summarized in this review. This discrimination was relatively easy to make, and there were few disagreements among reviewers.

Comprehensive programs. Comprehensive programs are intended to use the best aspects of both developmental-constructivist and skills-focused approaches. Like developmental-

11

constructivist approaches, they are likely to emphasize child-initiated activities, activity stations, art, and music. Activities are likely to be organized in themes, and are likely to provide many opportunities for make-believe, experimentation, and unstructured group play. However, some of the day is devoted to whole-class or small-group activities specifically focused on building language and early reading skills. Further, comprehensive programs are likely to regularly assess children's progress and to carefully plan both teacher-directed and child-initiated activities that contribute to progress toward specific language and literacy goals. Examples of comprehensive programs include *Curiosity Corner* (Authors, 2001), ELLM (Cosgrove et al., 2006), and DLM plus Open Court (Lonigan et al., 2011).

Developmental-constructivist programs. Programs falling into the developmental-constructivist category base their theories of action on the work of Piaget and Vygotsky. These programs have also been referred to as "child-centered" in other reviews (e.g., Gorey, 2001). Programs in this category have a strong emphasis on child-initiated activity, play, make-believe, art, music, and movement. Children typically play at self-chosen activity stations or tables, which offer them materials to stimulate language and cognitive development, with open-ended activities such as finger-painting, sand and water tables, a dress-up corner, a puppet theatre, blocks, cars and trucks, and so on. Teachers' roles are primarily supportive rather than directive, with teachers setting out and introducing activities, engaging individuals or groups in conversations about what they are doing or plan to do. Teachers introduce themes, often based on the children's interests, and discuss key concepts such as numbers, shapes, colors, and vocabulary. They read books to the class, and seek ways to encourage and expand appropriate uses of language. Direct teaching of phonemic awareness and phonics are rarely emphasized, and if it does take place it is likely to be incidental, in the context of thematic activities, rarely in a whole-class setting. There is usually a strong emphasis on social-emotional development, social skills (such as taking turns), and parent involvement. *Creative Curriculum* and *High/Scope* models are widely known and longstanding examples of this approach.

Preschool Curriculum Evaluation Research (PCER). The evaluation of outcomes of alternative approaches to early childhood education was greatly accelerated by a large federal initiative, the Preschool Curriculum Evaluation Research (PCER), funded by the U.S. Institute for Education Sciences between 2002 and 2005. PCER was of particular importance both because it applied consistent standards and assessments across all programs and because it followed up outcomes to the end of kindergarten. PCER commissioned third-party evaluations of 14 different preschool curricula with two independent external evaluators and 12 PCER grantees to compare one or two different curricula to a control condition in cluster randomized experiments. In each study, preschool classes or whole schools were randomly assigned to experimental or control conditions for two years. The external evaluators (Mathematica and RTI) administered a battery of nine measures designed to assess children's cognitive, language, beginning reading, math, and writing skills. The child assessments included: Social Awareness Tasks; Peabody Picture Vocabulary Test—3; Test of Early Language Development—Phonemic Awareness Subtest and Grammatical Understanding Subtest; Test of Early Reading Ability—3rd Edition; and the Letter-Word Identification and Spelling Subtests of the Woodcock-Johnson III. The measures were administered in the fall of preschool, in the spring of preschool, and again in

the spring of kindergarten. Posttest data, controlling for pretests, were analysed at the cluster level using hierarchical linear modeling (HLM; Raudebush & Bryk, 2002). In some cases, experimenters built on the PCER studies by following up PCER samples past kindergarten or by supplementing the PCER samples with additional students and schools recruited in the same way.

Of the 32 studies that met the inclusion criteria, 22 were conducted as part of PCER, and some of the remaining 10 studies were associated with PCER (e.g., they added additional samples to schools previously studied in the PCER study). Most of the PCER studies failed to find statistically significant differences between experimental and control groups, but the PCER studies were generally underpowered as cluster randomized trials, with only 12 to 39 schools/classes in each of the studies (in general, HLM requires 40-50 clusters to detect effect sizes as small as +0.20 with 80% probability [Raudenbush & Bryk, 2002]). Meta-analysis was largely developed to allow for pooling of underpowered but unbiased studies (Glass, McGaw, & Smith, 1981), and is particularly appropriate as a way to learn from small randomized experiments by clustering treatments in pragmatic categories. Adding in non-PCER studies adds to the diversity and usefulness of the review.

3.2 Program Summaries

The programs are clustered into the comprehensive and developmental-constructivist categories described above and presented in alphabetical order within each category. The programs and outcomes are described in the text, but information on the research design is discussed only if there are features that are not obvious from the tables.

Comprehensive programs. As noted earlier, comprehensive programs explicitly emphasize all aspects of children's development, teaching specific skills, language and school behaviors in a deliberate and planful way, but still make time for play, discovery, music, art, gross motor activity, and parent involvement. Eighteen programs fell into this category. Table 1 summarizes the findings of these studies.

Breakthrough to Literacy. *Breakthrough to Literacy* is a systematic and integrated literacy and language program published by the Wright Group that seeks to promote language development and literacy skills among preschool children. The program uses systematic, direct instruction built around a series of weekly books in the classroom. Interactive computer programs are used to engage pupils in individualized activities to support literacy skills and print knowledge.

Abt Associates (2007) carried out an 18-month study to examine the impacts of three intervention programs on teacher behaviors, classroom environments, and child outcomes—*Ready, Set, Leap!, Building Early Language and Literacy (BELL)*, and *Breakthrough to Literacy*. At the end of kindergarten, children who received *Breakthrough to Literacy* (*N*=354) outperformed the control group on average literacy measures and "definitional vocabulary" (ES = +0.44).

13

Bright Beginnings. *Bright Beginnings* is an integrated curriculum with a focus on language and early literacy. The curriculum goals are to provide a consistent, child-centered, literacy-focused program and to include instruction that addresses the needs of the whole child. The curriculum was especially designed to provide continuity in the preschool to second-grade curricula. BB includes nine curriculum units that focus on all domains of learning. The classroom environment is designed to encourage children's active exploration and interaction with adults, other children, and concrete materials.

As part of the PCER (2008) evaluation, researchers evaluated *Bright Beginnings* and the *Creative Curriculum*. In control classrooms, teachers used teacher-developed curricula with a focus on basic school readiness. A non-significant mean effect size of +0.31 across literacy outcomes at the end of preschool had faded by the spring of kindergarten to +0.02. Limited effects were found for two language measures at preschool (ES = +0.11) and at kindergarten (ES = +0.12).

A linked study by Lipsey, et al., (2009) followed a larger number of teachers and children over a five-year period. At the end of the preschool year, HLM analyses controlling for pretests found positive but non-significant effects on two literacy measures (mean ES = +0.18) but no differences on five language measures (mean ES = -0.03). Children in *Bright Beginnings* scored significantly higher than those in *Creative Curriculum* on Woodcock Johnson Letter-Word ID and Spelling scales, and on PPVT. Non-significant differences on individually-administered measures were also found on kindergarten and first grade assessments. However, on state tests in third grade, controlling for pretests, there were significant differences favoring former *Bright Beginnings* students in reading (ES = +0.27).

Building Early Language and Literacy (BELL). BELL is a supplementary program aimed at promoting preschoolers' general language proficiency, phonological awareness, shared reading skills, and print knowledge. Children receive two 15-20 minutes lessons daily. Children's literature is used in the classroom to build vocabulary and promote awareness of story sequencing and characters.

Abt Associates (2007) carried out an 18-month study to examine the impacts of three intervention programs on teacher behaviors, classroom environments, and child outcomes—*Ready, Set, Leap!*, *BELL*, and *Breakthrough to Literacy*. No statistically significant differences were found between the BELL group and the control group. Effect sizes averaged +0.06 for literacy measures and +0.07 for definitional vocabulary.

Classroom Links to Early Literacy. Powell, Diamond, Burchinal, and Koehler (2010) created and evaluated a professional development approach for preschool teachers designed to help them implement balanced, literacy-focused methods. The professional development was provided either in person or using distance technology, but a sub-study comparing these two PD methods found them to be equally effective. In both cases, teachers made and returned to project coaches videos of themselves using program elements.

Teachers were randomly assigned to receive the PD or to continue current practices (most implemented *Creative Curriculum).* By the end of preschool, children in the experimental group scored significantly better than control children on measures of letter knowledge, concepts about print, writing, and blending, with an average effect size across five literacy measures of +0.19. However, there were no differences on PPVT (ES=-0.03).

Curiosity Corner. *Curiosity Corner* is a comprehensive cognitive-developmental program created and disseminated by the Success for All Foundation.   It aims to develop the attitudes, skills, and knowledge necessary for later school success with an emphasis on children's language and literacy skills. *Curiosity Corner* comprises two sets of 38 weekly thematic units that include sequential daily activities. The program provides training, support, and teaching materials for teaching staff and administrators.

Curiosity Corner was evaluated in the PCER (2008) project. Adjusting for pretests, there were no significant differences at the end of preschool, but there were significant differences favoring the *Curiosity Corner* preschool attendees at the end of kindergarten on literacy measures (ES = +0.33) but not language (ES = +0.15).

Authors (2001), in an earlier matched study, evaluated *Curiosity Corner* in high-poverty communities in New Jersey. Two age groups participated in the study. Children in the three-year-old *Curiosity Corner* classes scored significantly higher on expressive language than their counterparts in the control group, but there were no differences for four-year-olds. Overall effects averaged ES = +0.15.

Dialogic Reading plus Sound Foundations. Whitehurst et al. (1999) evaluated a program that combined *Dialogic Reading*, an interactive story reading approach (Whitehurst et al., 1994), with the *Sound Foundations* program (Byrne & Fielding-Barnsley, 1993), which focuses on phonics and phonemic awareness. The average effect size (controlling for pretests) for two literacy measures was +0.12 at pre-k and +0.08 at kindergarten. On three language measures, average effect sizes were +0.12 at pre-k and +0.13 at kindergarten. Follow-up assessments in first grade found that control students scored non-significantly higher than former experimental students on Stanford Word Reading (ES = -0.16) and WRMT Word Attack (ES = -0.10). At the end of second grade, effect sizes were -0.29 for Word Reading and -0.26 for Word Attack.

Direct Instruction (**DI**). DI is a program first developed by Bereiter and Englemann (1966) as an instructional method for at-risk children. DI is a teacher-directed program in which specific cognitive and literacy skills are broken down into small units and taught explicitly. Teachers follow highly scripted lesson plans and techniques in their lessons. The main focus of the program is on basic academic concepts, such as arithmetic and reading. DI was evaluated as a comprehensive program by Engelmann (1968), but this study had too few students in the treatment group and was conducted before the start date of this review. More recently, Salaway (2008) examined the additive effects of a supplemental DI addition to a developmental-constructivist preschool curriculum. Outcomes at the end of six months showed children in the

15

experimental group outperformed controls on literacy (mean ES = +0.52) and language (mean ES = +0.46).

***DLM Early Childhood Express* (DLM) supplemented with** *Open Court Reading Pre-K.* DLM was evaluated in conjunction with the *Open Court* literacy-focused curriculum as part of the PCER (2008) project. In the control condition, teachers were provided with the *High/Scope* developmental-constructivist curriculum.

DLM is a comprehensive curriculum, designed to promote children's social, emotional, intellectual, aesthetic, and physical development through the use of hands-on learning experiences. The curriculum has 36 weekly themes that address all learning domains.

The *Open Court* curriculum content is presented in eight thematic units that address children's identity, families, friends, social interactions, transportation, the physical senses, nature, and transitions. Phonological, phonemic, and print-awareness activities are incorporated into each lesson. Controlling for a significant pretest difference, outcomes at the end of preschool showed children in the experimental group outperformed controls on literacy (ES = +0.49) and language (ES = +0.40). Effects for the experimental group were sustained through spring of kindergarten, for an average effect size of +0.47 for literacy outcomes and +0.47 for language outcomes.

Doors to Discovery. *Doors to Discovery* focuses on oral language, phonological awareness, concepts of print, alphabet knowledge, writing, and comprehension. It uses learning centers and shared literacy activities presented in eight thematic units that cover topics such as friendship, communities, and nature. The curriculum components also include family learning activities, professional development support for teachers, and assessment strategies that are integrated into the curriculum units.

Doors to Discovery was one of the curricula evaluated in the PCER (2008) project along with *Let's Begin with the Letter People.* These programs were separately compared to a control group. Effect sizes from *Doors to Discovery* at the end of preschool were +0.10 for literacy and +0.16 for language. Experimental-control differences were non-significant on all measures at the end of kindergarten, with an effect size of -0.09 for literacy and +0.12 for language.

***Early Literacy and Learning Model* (ELLM).** ELLM is a literacy-focused curriculum and support system designed for young children from low-income families. ELLM includes curriculum and literacy building blocks, assessment for instructional improvement, and professional development for literacy coaches and teachers. The ELLM curriculum and support system is designed to enhance existing classroom curricula with a two-hour daily block of language and literacy instruction. They include a set of literacy performance standards; monthly literacy packets; targeted instructional strategies; resource guides for teachers; a book lending library, and literacy calendars. Teachers target instruction in phonological awareness and letter recognition specifically for individual children based on baseline assessments.

16

ELLM was evaluated as part of the PCER (2008) project. Effects on four literacy measures averaged +0.10, and effects on PPVT and TOLD outcomes averaged +0.16. On follow-up into kindergarten, effect sizes averaged +0.11 for literacy and +0.39 for language.

In an extension to the PCER (2008) study, Cosgrove et al. (2006) also evaluated ELLM. As in the PCER evaluation, ELLM was implemented in combination with the existing curricula, while controls used only the existing curricula (generally *Creative Curriculum* or *High/Scope*). The treatment group scored significantly higher than the control groups on five literacy outcomes (mean ES = +0.25).

***Exemplary Model of Early Reading Growth and Excellence* (EMERGE).** EMERGE is a literacy-based program designed to help children from low-income families acquire early literacy skills. The program supports children's development of early literacy skills, using research-based teaching practices, progress monitoring to identify the need for more intensive intervention, provision of a literacy-rich learning environment, and continuous professional development. The curriculum includes interactive shared book reading and theme-based activities. Gettinger & Stoiber (2007) evaluated the model in a quasi-experimental one-year study. *EMERGE* children outperformed those in the control classrooms, with a mean effect size in literacy of +0.33, and in language of +0.13, at the end of preschool.

***Ladders to Literacy*.** *Ladders to Literacy* is a supplementary early literacy and language development curriculum for preschool and kindergarten children. It includes skill-building activities that are organized by print awareness; metalinguistic awareness; and oral language. Teachers are encouraged to select the activities that they want to implement and incorporate those activities into their daily classroom schedule. Teachers are provided with guidance on how to scaffold learning to individualize children's learning of language and literacy skills.

In the PCER (2008) study, *Ladders to Literacy* was implemented as a supplementary curriculum to the *Creative Curriculum*. Classrooms in the control condition implemented *Creative Curriculum* without the supplement. Effects were negative at the end of preschool (ES = -0.08 for literacy and -0.30 for language) and also at the end of kindergarten (ES = -0.25 for literacy and -0.18 for language).

***Let's Begin with the Letter People* (LBLP).** LBLP emphasizes early language and literacy development through play. In addition to classroom teaching emphasizing phonics, the program has a strong home/parent component.

Fischel et al. (2007) carried out a one-year study to evaluate the effectiveness of LBLP and the *Waterford Early Reading Program* separately compared to control groups. Head Start preschool classrooms in six centers were randomly assigned to one of these programs or a control condition, which used *High/Scope*. Children in LBLP scored significantly higher than the control group across 7 literacy measures, averaging an effect size of +0.20. However, there were no differences on PPVT (ES = + 0.06).

LBLP was one of the curricula evaluated in the PCER (2008) project. It was compared to a control group that implemented teacher-developed, non-specific curricula. No impacts on preschool or kindergarten pupil-level outcomes were found. In preschool, effect sizes were +0.04 across four literacy measures and +0.03 for PPVT and TOLD measures. On kindergarten follow-up measures, effect sizes were -0.13 for literacy and -0.06 for language.

Literacy Express. Literacy Express is a preschool curriculum that is designed to promote children's emergent literacy skills with a curriculum that is structured around thematic units. These units include selected children's books that address theme-relevant vocabulary for small- and large-group reading activities. In addition, each thematic unit includes small-group activities, conducted three to four times a week, which provide homogeneous small groups of children with practice in the skills needed to develop oral language, phonological sensitivity, and print awareness. *Literacy Express* and the combined *DLM & Open Court* model were evaluated as part of the PCER project in comparison to *High/Scope*. At the end of preschool there was an effect size of +0.17 across four literacy measures, and +0.07 on PPVT and TOLD. By the end of kindergarten, the literacy effects had faded to +0.03, and language outcomes averaged +0.13. As an extension of the PCER study, Lonigan, et al., (2011) evaluated *Literacy Express* in a study involving random assignment of preschool centers to experimental or control groups. At the end of preschool, there were differences on expressive language of +0.27.

Ready, Set, Leap!. Ready, Set, Leap! is a comprehensive preschool curriculum, published by LeapFrog SchoolHouse, which combines literacy-focused instructional approaches with multisensory technology. The curriculum is structured around 9 thematic units, each with detailed lesson plans for large- and small-group instruction, and ongoing assessment tools. The program stresses the importance of experiential learning, social and emotional development, teacher-child relationships, with a focus on language and literacy. The technology is designed around thematic center-based activities that provide individualized feedback to pupils.

For the PCER (2008) project, *Ready, Set, Leap!* was compared to *High/Scope*. No significant differences were found on the prekindergarten or kindergarten child outcomes, and all kindergarten effect sizes were essentially zero (literacy mean ES = -0.02 and language mean ES = -0.03).

A cluster randomized study of *Ready, Set, Leap!* (Davidson, Fields, & Yang, 2009) found non-significant effects on five posttest measures, with a mean effect size of +0.16 for literacy measures and +0.01 for language measures.

Abt Associates (2007) compared the impacts of three intervention programs on teacher behaviors, classroom environments, and child outcomes— *Ready, Set, Leap!*, BELL, and *Breakthrough to Literacy*. Children in the *Ready, Set, Leap!* group scored significantly higher than control students on all four subscales of the Test of Preschool Emergent Literacy (TOPEL), with a mean effect size of +0.50, and +0.28 for vocabulary.

Research-based Developmentally Informed (REDI). REDI is an enrichment program that is integrated into regular Head Start centers that use *High/Scope* or *Creative Curriculum*. This program is designed to promote academic and social-emotional school readiness by training teachers using program-based strategies and techniques in their classrooms that combine Preschool PATHS (a social-emotional approach), *Dialogic Reading*, a set of phonics "Sound Games," and print center activities for emergent literacy skills. Teachers receive a three-day intensive training prior to the intervention and one-day follow-up training four months into the intervention. In addition, teachers receive weekly mentoring support provided by REDI trainers. Parents are also provided with materials for home activities with their children.

Bierman et al. (2008) recruited two cohorts of four-year-olds over two years to participate in an evaluation of REDI. Significant treatment effects in pre-k were reported for two measures of literacy (mean ES = +0.31) and language (ES = +0.16).

A follow-up of the Bierman et al. (2008) sample into kindergarten was reported by Bierman et al. (2014). A significant difference was found only on the TOWRE phonemic decoding test (ES = +0.25, $p < .05$). Effect sizes were +0.03 for WJ Letter-Word Identification and -0.04 for TOWRE Sight Word Reading, for a kindergarten literacy mean of +0.08. On the EOWPVT vocabulary scale, the kindergarten effect sizes was +0.10 (n.s.).

Tools of the Mind. *Tools of the Mind* is a curriculum based on Vygotsky's theories. It focuses on children's ability to self-regulate, oral language, phonemic awareness, letter knowledge, and conventions of print. The activities emphasize children planning their activities, dramatic play, use of self-regulatory private speech, and use of external aids to facilitate memory and attention. Children learn in structured play, doing partner reading and writing activities, dance, and games.

Barnett et al. (2008) carried out a randomized evaluation of *Tools of the Mind*. Children and teachers were randomly assigned to use *Tools of the Mind* or a control condition in which children experienced a district-created "balanced literacy" method. The focus of the two curricula was described as being equal with regard to literacy, but there was more emphasis in the control condition on teacher direction and less on the development of self-regulation skills. All classes provided full-day (6 hrs/day) programs. Adjusting for pretests, effect sizes were -0.04 for literacy and +0.16 for language (PPVT and EOWPVT).

Farran, Meador, Norvell, and Nesbitt (2015) also evaluated *Tools of the Mind*. Schools were randomly assigned to use *Tools of the Mind* or to keep using their existing programs. *Tools of the Mind* was used for one year, and then children were followed up through first grade. Effects of *Tools of the Mind* were minimal in preschool and at follow-up. On two literacy measures, effect sizes were -0.10 at the end of preschool, -0.21 in kindergarten, and -0.19 in first grade. A passage comprehension measure added in first grade showed an effect size of -0.04. On two language measures, mean effect sizes were -0.06 in preschool, -0.01 in kindergarten, and +0.03 in first grade.

Waterford Early Reading Program. *Waterford* is an integrated learning system that provides 15 minutes of daily computerized one-to-one learning activities for preschool children. It focuses on teaching children their letters, as well as developing phonological and phonemic awareness, story and print concepts, and language concepts. It gives teachers information on children's levels of skill, which they are expected to use to provide appropriate teaching outside of computer time. Developmentally appropriate books and videotapes are introduced in class and then sent home with children.

Fischel et al. (2007) evaluated *Waterford*. Combining across three cohorts, a total of 12 classes were randomly assigned to *Waterford* and 11 to continue with *High/Scope*. An additional 12 classes were randomly assigned to LBLP, described elsewhere in this article. Adjusting for pretests, posttest literacy effect sizes comparing *Waterford* to control averaged +0.08 across 7 literacy measures. On the PPVT language measure, the effect size for *Waterford* was +0.06.

Developmental-constructivist programs. While developmental-constructivist programs such as *Creative Curriculum* and *High/Scope* have long been the most popular programs in U.S. preschools, there were surprisingly few studies of these approaches. In fact, they were most likely to show up as the control condition for other innovations. Four evaluated programs were categorized as developmental-constructivist. They were: *Creative Curriculum, Language-Focused Curriculum, Project Approach,* and *Project Construct.* Characteristics and findings of these studies appear in Table 2

Creative Curriculum. *Creative Curriculum* is a comprehensive approach to education for 3- to 5-year-old children. The curriculum addresses four areas of development: social/emotional, physical, cognitive, and language development. *Creative Curriculum* requires the physical space of the classroom to be structured into 10 interest areas: blocks, dramatic play, toys and games, art, library, discovery, sand and water, music and movement, cooking, and computers. Time is also allotted for outdoor activities. The 10 interest areas are designed to address curriculum content, such as literacy, mathematics, science, social studies, the arts, and technology, in the context of play and discovery emphasizing observing, exploring, and problem solving.

The PCER (2008) project compared *Creative Curriculum*, *Bright Beginnings*, and control treatments. In the control classrooms, teachers used teacher-developed curricula with a focus on basic school readiness. No significant differences between *Creative Curriculum* and control classes were found. Effect sizes across literacy measures were +0.12 at preschool and +0.20 at kindergarten, and +0.15 at preschool and +0.12 at kindergarten for two language measures.

Another PCER (2008) project also evaluated *Creative Curriculum.* In the control condition, teachers used teacher-developed, nonspecific curricula. No significant impacts on child outcomes were found. Effect sizes averaged -0.08 for four literacy measures at preschool and -0.03 on two language measures. At kindergarten follow-up, effect sizes were -0.01 for literacy and -0.01 for language.

In the study cited previously by Lipsey et al. (2009), which overlapped the PCER (2008) evaluation, preschools were randomly assigned to *Creative Curriculum, Bright Beginnings,* or control. On HLM analyses at the end of the pre-k year, children in *Creative Curriculum* classes scored non-significantly lower than controls on 2 literacy measures (mean ES = -0.11) and 5 language measures (mean ES = -0.07), and they scored significantly lower than children in *Bright Beginnings* on Woodcock Johnson Letter-Word ID and Spelling, and on PPVT. However, followed up to third grade, former *Creative Curriculum* children scored non-significantly higher than controls on state reading tests (ES = +0.16).

Averaging across these three evaluations, a weighted mean effect size of -0.02 was found for literacy at pre-k and a mean of +0.06 was found at kindergarten. For language measures, mean effect sizes were +0.02 at pre-k and +0.04 at kindergarten.

Language-Focused Curriculum. *Language-Focused Curriculum* was developed at the University of Kansas (Bunce, 1995) for use with 3- to 5-year-old children with language limitations, including children with language impairment; children from disadvantaged backgrounds; and English-language learners. The curriculum has a thematic organization and focuses on the use of daily dramatic play to teach and use linguistic concepts. There are both teacher-led and child-led activities with explicit attention to oral language development that is enhanced by high-quality teacher-child conversations. Teachers use eight specific language stimulation techniques when interacting with children in the classroom, such as event casts (descriptions of an activity while it is taking place) and expansions (repeating the child's utterance with varied vocabulary) (Justice, Mashburn, Pence & Wiggins, 2008).

In the PCER (2008) study, *Language-Focused Curriculum* was compared with *High/Scope*. No significant impacts on preschool or kindergarten child outcomes were found. There was a modest effect on literacy (ES = +0.18) at the end of preschool which had diminished to ES = +0.05 by kindergarten. On two language measures effect sizes were +0.02 in preschool and -0.08 in kindergarten.

Project Approach. The *Project Approach* is a set of teaching strategies that enables teachers to guide children through in-depth investigations of real world topics. The curriculum is designed to use children's interests as the starting point for organizing and developing classroom learning activities. Three curriculum components address children's learning needs: spontaneous play, systematic instruction, and project work. A *project* is defined as an in-depth study of a real world topic that is worthy of children's attention and effort. Projects can be incorporated into an existing classroom instructional program and can extend over several days or weeks. The structural features of the *Project Approach* include discussion, fieldwork, representation, investigation, and display. During the preliminary planning stage, the teacher selects the topic of study (based primarily on classroom learning goals, children's interests, and the availability of local resources). The class brainstorms experience, knowledge, and ideas and the teacher represents them in a topic web. In PA classrooms, the daily schedule is structured so that children and teachers spend at least 45 to 60 minutes engaged in investigation and discovery.

In the PCER (2008) project, *Project Approach* was compared to control classrooms, where teachers implemented their own teacher-developed, nonspecific curricula. At the end of prekindergarten, there were non-significant effects on literacy (ES = +0.22) and language (ES = +0.16). At the end of kindergarten the effects on literacy dropped to +0.07, but the effects on language (ES = +0.21) maintained.

Project Construct. *Project Construct* is organized around 29 goals for students that are set within a context of four development domains: cognitive, representational, sociomoral, and physical.

For the PCER (2008) project, *Project Construct* was compared to teacher-developed generic curricula. Students were randomly assigned to controls. At preschool, effect sizes averaged -0.03 for literacy and -0.01 for language, and in kindergarten effect sizes were 0.00 for literacy and -0.06 for language.

2.3 Programs and Outcomes by Category

Table 3 summarizes overall literacy and language outcomes at preschool and kindergarten for all 32 studies of 22 programs, sorted into "comprehensive" and "developmental-constructivist" categories. Mean effect sizes are weighted by the smaller of the sample sizes involved (experimental or control).

All of the average effect sizes in the meta-analysis for comprehensive programs were statistically significant (p < .05). Weighted effect sizes for preschool outcomes averaged ES = +0.15 (p < .01) for literacy and ES = +0.08 (p < .01) for language. At kindergarten follow-up, effect sizes were +0.14 (p < .01) for literacy and +0.15 (p < .01) for language.

None of the effect sizes for developmental-constructivist programs were either pragmatically or statistically significant. The weighted means were +0.04 for literacy and +0.03 for language in preschool, and +0.06 for literacy and +0.06 for language at kindergarten follow-up.

There is a great deal of variation among both comprehensive and developmental-constructivist programs, and there are only six studies of developmental-constructivist programs, but there is little evidence to support the outcomes of the developmental-constructivist programs for language and literacy measures, and certainly no evidence that initial gains in comprehensive programs would fade by kindergarten.

4. Discussion and Conclusion

The findings of this systematic review provide evidence in support of those in the field of early childhood education who believe that it is beneficial to provide some direct instruction in language and literacy to children before they begin elementary school, including the observations of the National Early Literacy Panel (2008). This is particularly important for children from

22

disadvantaged communities, where they are less likely to be talked with, and read to, at home. All of the programs included in this review were evaluated in implementations in high poverty communities. For this reason, the results may be more generalizable to those populations. All took place in the U.S., so the applicability of the findings of this review to contexts outside the U.S. must be addressed in the future.

The 32 studies included in this review are of exceptional methodological quality. Twenty-seven of the studies used cluster randomized designs, where schools, classes, or teachers were randomly assigned to treatments and analyses were done at the cluster level, and two more did random assignment at the student level. One small study randomly assigned at the cluster level but analyzed data at the student level. Only two studies used quasi-experimental designs. The use of these rigorous designs, all but two with an element of random assignment, virtually rules out selection bias as an alternative explanation for study findings. Such rigorous designs make it difficult for programs to show strong positive effects, so the modest effect sizes for most programs must be seen in this light.

The most important substantive outcome of the review is a clear positive effect of comprehensive programs on literacy and language measures, both in preschool and on kindergarten follow-up. The findings support the idea that young children learn best in programs that balance skills-focused and developmental activities. Programs that focus on developmental-constructivist, child-initiated activities but do not incorporate teaching of phonemic awareness and phonics skills had lower effect sizes than did those that had a focus on early literacy skills as well as developmental activities. Not surprisingly, the advantage of comprehensive over developmental programs was greatest on literacy at the end of preschool (ES = +0.15 for comprehensive programs, ES = +0.08 for developmental-constructivist programs). This makes sense, because literacy activities are primarily what comprehensive programs add. Such programs allocate up to half of their time to literacy activities. However, comprehensive programs also produced somewhat higher effect sizes on language measures at the end of preschool (ES = +0.08 vs. +0.03). On kindergarten follow-up, comprehensive programs showed better outcomes than developmental-constructivist programs on outcomes for both literacy (ES = +0.14 vs. +0.06) and language (ES = +0.15 vs. +0.06). Preschool literacy outcomes may merely indicate that teaching preschool children skills ordinarily emphasized in kindergarten or later produce immediate effects on those skills. However, given that mean effect sizes on language as well as literacy measures were higher for comprehensive programs than for developmental-constructivist programs in kindergarten certainly does not support any concern that preschool literacy outcomes of comprehensive programs might just be temporary, due to an early focus on literacy.

All ten of the programs with strong evidence of effectiveness on at least one preschool or kindergarten outcome (i.e., effect sizes ≥ +0.25 in studies meeting inclusion criteria) were from the comprehensive category. These programs had clear goals and measures of how children are achieving them with some focus on academic outcomes. It is easier for teachers to monitor the progress of children if they have a clear idea of what they are working toward. They can provide carefully planned experiences designed to move children toward success on literacy and

23

language outcomes, and this gives the children a significant advantage as they enter the elementary grades.

4.3 Limitations

It is important to note several limitations of the present review. First, the review focuses on experimental studies using quantitative measures of child learning outcomes of early childhood interventions. To compare the effectiveness of programs, one needs quantitative evidence that can be evaluated on a common scale, primarily on individually-administered standardized tests. These are useful in assessing the practical outcomes of various programs and are fair to control as well as experimental groups. However, the review does not report on experimenter-made measures of content taught in the experimental group but not the control group, although results on such measures, as well as on outcomes other than achievement, may also be of importance to researchers or educators.

Second, the review focuses on replicable programs used in realistic early childhood settings, excluding unrealistic implementations of programs that could not be replicated as they were implemented. This emphasis is consistent with the review's purpose in providing educators with useful information about the strength of evidence supporting various practical programs, but it does not attend to smaller or shorter, perhaps more theoretically-driven studies that may also provide useful information, especially to researchers.

The review focuses on the main approaches used in preschool classes, with a particular interest in contrasts between main approaches that do or do not include explicit teaching of language and early literacy.

The review does not include social-emotional outcomes. We intended to include the impacts of interventions on children's social and emotional development. However, few studies included these outcomes. When they did, the data usually came from teacher or parent ratings of children's behavior, rather than on unbiased observations of children's actual behavior. Teacher and parent ratings can be influenced by their knowledge of being in a study and of the goals of the particular intervention. For this reason, studies in this domain were not included.

Finally, there were very few qualifying studies on mathematics impacts, so we have excluded math outcomes from the review.

4.4 Conclusion

As programs for young children expand in availability, they must also grow in quality. Further development and evaluation of preschool approaches may discover new ways to help young children further prepare themselves for elementary school. The findings of this review of research since 1990 add to a growing body of evidence that early childhood programs can have an important impact on increasing the school readiness of young children. There is a tremendous need for systematic, large-scale, longitudinal, randomized evaluations of the effectiveness of

preschool interventions in bringing children from high-risk environments to normative levels of academic achievement. This review identifies several promising approaches that could be used today to help children begin elementary school ready to succeed, but more programs and more research are needed to better understand how to provide young children with optimal experiences in preschool.

Author Note

This review was partially supported by a grant from the Spencer Foundation. However, any opinions expressed are those of the authors and do not represent the positions or policies of our funders.

5. References

Abt Associates, Inc. (2007, March). *Findings from Project Upgrade in Miami-Dade County.* Cambridge, MA.

Barnett, W. S., Jung, K., Yarosz, D. J., Thomas, J., Hornbeck, A., Stechuk, R. & Burns, S. (2008). Educational effects of Tools of the Mind curriculum: A randomized trial. *Early Childhood Research Quarterly, 23*(3), 299-313. doi:10.1016/j.ecresq.2008.03.001

Belsky, J. & Melhuish, E. (2007). Impact of Sure Start local programmes on children and families. In J. Belsky, J. Barnes, & E. Melhuish, (Eds.). *The National Evaluation of Sure Start: Does Area-Based Early Intervention Work?* (pp. 133-154). Bristol, UK: The Policy Press.

Bereiter, C., & Engelmann, S. (1966). *Effectiveness of direct verbal instruction on IQ performance and achievement in reading and arithmetic.* Champaign: University of Illinois.

Bierman, K. L., Domitrovich, C., et al. (2008). Promoting academic and social-emotional school readiness: The Head Start REDI Program. *Child Development, 79* (6), 1802-1817. doi: 10.1111/j.1467-8624.2008.01227.x.

Bierman, K., Nix, R., et al. (2014). Effects of Head Start REDI on children's outcomes 1 year later in different kindergarten contexts. *Child Development, 85* (1), 140-159. doi:10.1111/cdev.12117

Brewer, M. Cattan, S., Crawford, C., & Rabe, B. (2014). *The impact of free early education for 3 year olds in England.* London, England: ESRC

Bunce, B.H. (1995). *Building a Language-Focused Curriculum for the Preschool Classroom. Volume II: A Planning Guide.* Baltimore, MD: Brookes Publishing.

Byrne, B. & Fielding-Barnsley, R. (1993). Evaluation of a program to teach phonemic awareness to young children: A 1-year follow-up. *Journal of Educational Psychology, 85*(1), 104-111. doi: 10.1037/0022-0663.85.1.104

Camilli, G, Vargas, S., Ryan, S., & Barnett, S. (2009). Meta-analysis of the effects of early education interventions on cognitive and social development. *Teachers College Record, 112* (3), 579-620.

Carneiro, P. & Heckman, J. (2003). Human capital policy. In J. J. Heckman, A. B. Krueger, and B. M. Friedman (Eds.), *Inequality in America: What role for human capital policies?* Cambridge, MA: MIT Press, 77-239.

Claessens, A., & Garrett, R. (2014). The role of early childhood settings for 4-5 year old children in early academic skills and later achievement in Australia. *Early Childhood Research Quarterly, 29,* (4), 550-561.

Coghlan, M., Bergeron, C., White, K., Sharp, C., Morris, M., & Rutt, S. (2009). *Narrowing the gap in outcomes for young children through effective practices in the early years.* Centre for Excellence and Outcomes in Children and Young People's Services, London, UK.

Cooper, H. (1998). *Synthesizing research (3rd ed.).* Thousand Oaks, CA: Sage.

Cosgrove, M., Fountain, C., Wehry, S. et al. (2006, April). *Randomized field trial of an early literacy curriculum and instructional support system.* Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Darrow, C. L. (2009, March). *Language and literacy effects of curriculum interventions for preschools serving economically disadvantaged children: A meta-analysis.* Paper presented at the annual meeting of the Society for Research on Educational Effectiveness, Alexandria, Virginia.

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

Diamond, K., Justice, L., Siegler, R., & Snyder, P. (2013). *Synthesis of IES research on early intervention and early childhood education (NCSER 2013-3001).* Washington, DC: USDOE.

Engelmann, S. (1968). The effectiveness of direct instruction on IQ performance and achievement in reading and arithmetic. In J. Hellmuth (Ed.), *Disadvantaged child* (vol. 3). New York: Brunner/Mazel.

Farran, D., Meador, D., Norvell, J., & Nesbitt, K. (2015). *Experimental evaluation of the Tools of the Mind curriculum* (Tech. rep.). Nashville, TN: Vanderbilt University, Peabody Research Institute.

Fischel, J. E., Bracken, S. S., Fuchs-Eisenberg, A., Spira, E. G., Katz, S., & Shaller, G. (2007). Evaluation of curricular approaches to enhance preschool early literacy skills. *Journal of Literacy Research, 39*(4), 471. doi: 10.1080/10862960701675333.

Gettinger, M. & Stoiber, K. (2007). Applying a response-to-intervention model for early literacy development in low-income children. *Topics in Early Childhood Special Education, 27*(4), 198-213. doi: 10.1177/0271121407311238

Gorey, K. M. (2001). Early childhood education: A meta-analytic affirmation of the short- and long-term benefits of educational opportunity. *School Psychology* Quarterly*, 16*, 9-30.

Gormley, W., Phillips, D., & Gayer, T. (2008). Preschool programs can boost school readiness. *Science, 320* (5884), 1723-1724.

Harms, T., Clifford, R. M., & Cryer, D. (2014). *Early Childhood Environment Rating Scales (ECERS-3).* New York: Teachers College Press.

Heckman, J.J., & Masterov, D.V. (2007). The productivity argument for investing in young children (April 2007). Retrieved June 23, 2015 from http://ssrn.com/abstract=982117

Jacob, R., Creps, C., & Boulay, B. (2004, July). *Meta-analysis of research and evaluation studies in early childhood education.* Cambridge, MA: Abt Associates.

Justice, L.M., Mashburn, A., Pence, K., and Wiggins, A. (2008). Experimental evaluation of a preschool language curriculum: Influence on children's expressive language skills. *Journal Of Speech, Language, and Hearing Research, 51* (4): 983–1001. doi: 10.1044/1092-4388(2008/072).

Karoly, L. A., Kilburn, M., & Cannon, J. (2005). *Early childhood interventions: Proven results, future promises.* Retrieved from http://www.rand.org/pubs/monographs/2005/RAND_MG341.pdf.

Karweit, N.L. (1994). Can preschools alone prevent early reading failure? In R.E. Slavin, N.L. Karweit, & B.A. Wasik (Eds.), *Preventing early school failure.* Boston: Allyn & Bacon.

Lipsey, M.W., Farran, D.C., Hurley, S. M., Hofer, K.G., & Bilbrey, C. (2009, March). *Effects of a literacy-focused curriculum and a developmental curriculum on school readiness and subsequent state achievement test outcomes in rural prekindergarten classrooms.* Washington, DC: Society for Research on Educational Effectiveness.

Lipsey, M. W., Hofer, K., Dong, N., Farran, D., & Bilbrey, C. (2013a). *Evaluation of the Tennessee Voluntary Kindergarten Program: End of pre-k results from the randomized control design.* Nashville, TN: Peabody Research Institute, Vanderbilt University.

Lipsey, M. W., Hofer, K., Dong, N., Farran, D., & Bilbrey, C. (2013b). *Evaluation of the Tennessee Voluntary Kindergarten Program: Kindergarten and first grade follow-up results from the randomized control design.* Nashville, TN: Peabody Research Institute, Vanderbilt University.

Lipsey, M. W., & Wilson, D. B. (2001). Practical meta-analysis. Thousand Oaks, CA: Sage.

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,* 305-337. doi: 10.1007/s11145-009-9214-6

Mason, J. M., Kerr, B. M., Sinha, S., & McCormick, C. (1990). *Shared book reading in an early start program for at-risk children* (Tech Rep. #504). Champaign-Urbana: University of Illinois, Center for the Study of Reading.

Melhuish, E., Belsky, J., & Leyland, R. (2010). *The impact of Sure Start local programmes on five year olds and their families.* London: Department for Education.

National Early Literacy Panel (2008). *Developing early literacy: A systematic synthesis of early literacy development and implications for intervention.* Washington, DC: National Institute for Literacy.

Nelson, G., Westhues, A., & MacLeod, J. (2003). A meta-analysis of longitudinal research on preschool prevention programs for children. *Prevention & Treatment, 6* (31), 1-35.

Neuman, S. B., Copple, C., & Bredekamp, S. (1999). Learning to read and write: Developmentally appropriate practices for young children. Washington, DC: National Association for the Education of Young Children.

Neuman, S. B., Newman, E. H., & Dwyer, J. (2011). Educational effects of a vocabulary intervention on preschoolers' word knowledge and conceptual development: A cluster-randomized trial. *Reading Research Quarterly, 46*(3), 249-272. doi: 10.2307/41228653

Piaget, J. (1952). *The language and thought of the child.* London, UK: Routledge and Kegan-Paul.

Piasta, S. B., Justice, L. M., McGinty, A. S., & Kaderavek, J. N. (2012). Increasing young children's contact with print during shared reading: Longitudinal effects on literacy achievement. *Child Development, 83*(3), 810-820. doi: 10.1111/j.1467-8624.2012.01754.x

Pollard-Durodola, S. D., et al. (2011). The effects of an intensive shared book-reading intervention for preschool children at risk for vocabulary delay. *Exceptional Children, 77*(2), 161-183. doi: 10.1177/001440291107700202

Powell, D. R., Diamond, K. E., Burchinal, M.R., & Koehler, M. J. (2010). Effects of an early literacy professional development intervention on Head Start teachers and children. *Journal of Educational Psychology, 102* (2), 299-312. doi.org/10.1037/a0017763

Preschool Curriculum Evaluation Research Consortium (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. Washington, DC: U.S. Government Printing Office.

Puma, M., Bell, S., Cook, R., & Heid, C. (2010). *Head Start impact study: Final report.* Washington, DC: U.S. Department of Health and Human Services.

Ramey, C. T., & Ramey, S. L. (1998). Early intervention and early experience. *American Psychologist, 53* (2), 109-120.

Raudenbush, S., & Bryk, A. (2002). *Hierarchical linear models* (2nd Ed.). Thousand Oaks, CA: Sage.

Rothstein, H. R., Sutton, A. J., & Borenstein, M. (Eds.) (2005). *Publication bias in meta-analysis: Prevention assessment, and adjustments.* Chichester, UK: John Wiley.

Salaway, J. L. (2008). *Efficacy of a direct instruction approach to promote early learning*. Unpublished Ph.D. dissertation., Duquesne University.

Schweinhart, L. J., & Weikart, D. P. (1997). *Lasting differences*: *The High/Scope Preschool curriculum comparison study through age 23* (Monographs of the High/Scope Educational Research Foundation No. 12) Ypsilanti, MI: High/Scope Press.

Schweinhart, L. J., Barnes, H. V., & Weikart, D. P. (1993). *Significant benefits: The High/Scope Perry Preschool study through age 27* (Monographs of the High/Scope Educational Research Foundation No. 10) Ypsilanti, MI: High/Scope Press.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton-Mifflin.

Vygotsky, L. S. (1987). *The collected works of L.S. Vygotsky (translated by Rierbero Carion).* New York: Pendulum.

Waldfolgel, J. & Washbrook, E. (2010). *Low income and early cognitive development in the UK*. London, England: Sutton Trust.

Wasik, B. A., & Bond, M. A. (2001). Beyond the pages of a book: Interactive book reading and language development in preschool classrooms. *Journal of Educational Psychology, 93* (2), 243-250. doi:10.1037/0022-0663.93.2.243

Weikart, D. P. (1995). Early childhood education. In J.H. Block, S.T. Everson, & T.R. Guskey (Eds.), *School Improvement Programs* (pp. 289-312). New York: Scholastic, Inc.

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 (2015). *Shared book reading: WWC intervention report.* Washington, DC: Institute of Education Sciences. Retrieved June 18, 2015 from http://ies.ed.gov/ncee/wwc/interventionreport.aspx?sid=458.

Whitehurst, G. J., Zevenbergen, A. A., Crone, D. A., Schultz, M. D., Velting, O. N., & Fischel, J. E. (1999), Outcomes of an emergent literacy intervention from Head Start through second grade. *Journal of Educational Psychology, 91*(2), 261-272. doi: 10.1037/0022-0663.91.2.261.

Whitehurst, G. J., Epstein, J. N., Angell, A. C., Payne, A. C., Crone, D. A., & Fischel, J. E. (1994). Outcomes of an emergent literacy intervention in Head Start. *Journal of Educational Psychology, 86,* 542-555. doi.org/10.1037/0022-0663.86.4.542

Wilson, S., Dickinson, D., & Rowe, D. (2013). Impact of an Early Reading First program on the language and literacy achievement of children from diverse language backgrounds. *Early Childhood Research Quarterly, 28,* 578-592.

Table 1
Designs, Settings, and Outcomes for Balanced Programs

Study	Design/Control	Duration	N	Sample Characteristics	Posttest	Preschool ES	Preschool Mean ES	Kinder ES	Kinder Mean ES	Post Kinder ES
Breakthrough to Literacy										
Abt Associates (2007)	Cluster Randomized / Ordinary preschool	18 months	863 students (354E, 509C)	162 child care centers in Miami-Dade County that had 75% or more students eligible for FRL	Literacy					
					TOPEL Print Knowledge			+0.60	+0.52	
					TOPEL Phonological Awareness			+0.44		
					Language					
					TOPEL Definitional Vocabulary			+0.31	+0.31	
Bright Beginnings										
PCER (2008)	Cluster Randomized / Ordinary preschool	2 years	14 classes 208 students (103E, 105C)	Seven school districts in six counties in TN; 80% White, 18% African American, 11% Hispanic	Literacy					
					TERA	+0.39		-0.07		
					WJ Letter Word ID	+0.35	+0.21	+0.09	+0.02	
					WJ Spelling	+0.18		+0.06		
					Pre-CTOPP/CTOPP	-0.07		+0.01		
					Language					
					PPVT	+0.13	+0.11	+0.07	+0.12	
					TOLD	+0.09		+0.16		
Lipsey et al. (2009)	Cluster Randomized / Ordinary preschool	5 years	36 pre-K classes (116E, 148C)	Seven school districts in six rural middle TN counties. Overlaps PCER sample. 70% White, 19% Afr-American, 4% Hispanic	Literacy					
					WJ Letter-Word ID	+0.20				
					WJ Spelling	+0.15	+0.18			

Reference	Design	Duration	Sample Size	Location/Demographics	Measure	ES	Subtotal	Overall
					Language			
					WJ Picture Vocabulary	-0.02		
					WJ Story Recall	+0.06		
					WJ Directions	+0.06	-0.03	
					WJ Comprehension	-0.17		
					PPVT	+0.05		
					Follow-Up			
					TCAP Reading			Gr. 3: +0.27

Building Early Language and Literacy (BELL)

Reference	Design	Duration	Sample Size	Location/Demographics	Measure	ES	Subtotal	Overall
Abt Associates (2007)	Cluster Randomized Ordinary preschool	18 months	849 students (340E, 509C)	162 child care centers in Miami-Dade County Florida that had 75% or more students eligible for FRL	**Literacy**			
					TOPEL Print Knowledge		+0.07	+0.06
					TOPEL Phonological Awareness		+0.04	
					Language			
					TOPEL Definitional Vocabulary		+0.07	+0.07

Classroom Links to Early Literacy

Reference	Design	Duration	Sample Size	Location/Demographics	Measure	ES	Subtotal	Overall
Powell, Diamond, Burchnal, & Koehler (2010)	Cluster Randomized Mostly Creative Curriculum	1 Semester	73 teachers (42E, 31C) 568 students (310E, 258C)	Head Start centers in a midwest state. 40% African American, 30% White, 23% Hispanic	**Literacy**			
					WJ Letter-Word ID	+0.10		
					FACES Alphabet	+0.29		
					Concepts of Print	+0.22	+0.19	
					Writing	+0.17		
					TOPEL Blending	+0.18		
					Language			
					PPVT	-0.03	-0.03	

Curiosity Corner

Study	Design	Duration	N	Sample	Outcome	ES	Subtotal	ES	Total
PCER (2008)	Cluster Randomized / Mostly Creative Curriculum	2 years	18 pre-K programs / 225 students (105E, 110C)	Preschool programs in FL, KS, and NJ; 28% White, 51% African American, 14% Hispanic, and 8% others	**Literacy**				
					TERA	+0.10		+0.43	
					WJ Letter Word ID	+0.09	+0.10	+0.43	+0.33
					WJ Spelling	+0.04		+0.20	
					Pre-CTOPP/CTOPP	+0.18		+0.25	
					Language				
					PPVT	-0.01	-0.05	+0.14	+0.15
					TOLD	-0.08		+0.15	
Authors (2001)	Matched Control / Ordinary preschool	1 year	316 students (206E, 110C)	3 and 4-year-old children enrolled in child care centers and preschools in 4 high poverty urban school districts in New Jersey. 70% African American, 16% White, 11% Hispanic	**Language**				
					MSEL Expressive Language	+0.24	+0.15		
					MSEL Receptive Language	+0.06			

Dialogic Reading plus Sound Foundations

Study	Design	Duration	N	Sample	Outcome	ES	Subtotal	ES	Total
Whitehurst et al. (1999)	Randomized Quasi-Experiment / Ordinary preschool	4 years	37 classes / 280 students	Head Start centers in Suffolk County, New York. 43% African American, 33% White, 18% Hispanic, & 6% others	**Literacy**				
					DSC Print Concepts	+0.12	+0.12	+0.03	+0.08
					DSC Writing	+0.11		+0.13	
					Language				
					EOWPVT Language	+0.08		+0.14	
					DSC Memory	+0.19	+0.12	+0.12	+0.13
					DSC Auditory	+0.09		+0.13	

				Follow-Up			Gr 1: -0.16; Gr. 2: -0.29
				Stanford Word Reading			Gr 1: -0.16; Gr. 2: -0.29
				WRMT Word Attack			Gr 1: -0.10; Gr. 2: -0.26

Direct Instruction

				Literacy			
Salaway (2008)	Randomized Developmentally appropriate preschool	6 months	61 students (35E, 26C)	A preschool center in an urban, at risk community. 20% White, 69% African American, 2% Hispanic, and 10% others	DIBELS Initial Sounds Fluency	+0.75	
					DIBELS Letter Naming Fluency	+0.50	+0.52
					Letter and Word Skills	+0.32	
				Language			
				K-SEALS Expressive Language	+0.40	+0.46	
				K-SEALS Receptive Language	+0.51		

DLM Express plus Open Court

				Literacy			
PCER (2008)	Cluster Randomized High/Scope	2 years	11 preschool programs 198 students (101E, 97C)	Public preschool programs in FL; 30% White, 59% African American, 6% Hispanic, 5% others	TERA	+0.68	+0.76
					WJ Letter Word ID	+0.51	+0.50
					WJ Spelling	+0.46 +0.49	+0.22 +0.47
					Pre-CTOPP/CTOPP	+0.32	+0.38
				Language			
				PPVT	+0.40	+0.48
				TOLD	+0.40 +0.40	+0.46 +0.47

Doors to Discovery

Reference	Design	Duration	Sample	Context	Outcomes				
PCER (2008)	Cluster Randomized Ordinary preschool	2 years	29 classes 297 students (101E, 96C)	Head Start and public preschool programs in TX; 30% White, 13% African American, 43% Hispanic, 13% others	**Literacy**				
					TERA	+0.06		-0.05	
					WJ Letter Word ID	+0.10		-0.09	
					WJ Spelling	+0.06	+0.10	-0.12	-0.09
					Pre-CTOPP/CTOPP	+0.18		-0.09	
					Language				
					PPVT	+0.15		+0.18	
					TOLD	+0.17	+0.16	+0.06	+0.12

Early Literacy & Learning Model (ELLM)

Reference	Design	Duration	Sample	Context	Outcomes		
Cosgrove et al. (2006)	Cluster Randomized Creative Curriculum or High/Scope	1 year	466 students (222E, 244C)	Head Start, subsidized, faith-based and preschool classrooms from 3 locations in FL; 14% White, 71% African American, 8% Hispanic, 6% others. Overlaps PCER sample.	**Literacy**		
					TERA Reading Quotient	+0.28	
					TERA Alphabet	+0.28	
					TERA Print	+0.17	+0.25
					TERA Meaning	+0.29	
					Alphabet Letter Recognition Inventory	+0.25	
PCER (2008)	Cluster Randomized Creative Curriculum, High/Scope, others	1 year	28 classes 244 students (137E, 107C)	Head Start, subsidized, faith-based and preschool classrooms from 3 locations in FL; 14% White, 71% African American, 8% Hispanic, 6% others	**Literacy**				
					TERA	+0.15		+0.30	
					WJ Letter Word ID	-0.05		0.00	
					WJ Spelling	+0.11	+0.10	+0.04	+0.11
					Pre-CTOPP/CTOPP	+0.18		+0.08	
					Language				
					PPVT	+0.17		+0.34	
					TOLD	+0.15	+0.16	+0.44	+0.39

EMERGE

Study	Design	Duration	Sample	Location	Domain/Measure	ES		
Gettinger & Stoiber (2007)	Matched Control / Ordinary preschool	1 year	342 students (188E, 154C)	Low SES Head Start and preschool centers in Milwaukee, Wisconsin, 90% African American	**Literacy**			
					PALS Pre-K Alphabet Knowledge	+0.47		
					PALS Pre-K Print Awareness	+0.50	+0.33	
					PALS Pre-K Name Writing	+0.01		
					Language			
					PPVT	+0.13	+0.13	

Ladders to Literacy

Study	Design	Duration	Sample	Location	Domain/Measure	ES			
PCER (2008)	Cluster Randomized / Creative Curriculum	2 years	14 classes 123 students (62E, 61C)	Head Start centers in NH 38% White, 11% African American, 30% Hispanic, and 20% others	**Literacy**				
					TERA	-0.30		-0.54	
					WJ Letter Word ID	-0.16		-0.27	
					WJ Spelling	+0.30	-0.08	-0.08	-0.25
					Pre-CTOPP/CTOPP	-0.16		-0.10	
					Language				
					PPVT	-0.38	-0.30	-0.30	-0.18
					TOLD	-0.22		-0.06	

Let's Begin With The Letter People

Study	Design	Duration	Sample	Population	Measure					
PCER (2008)	Cluster Randomized High/Scope	2 years	30 classes 196 students (100E, 96C) Shared same control group with Doors to Discovery	Head Start and public preschool programs in TX 30% White, 13% African American, 43% Hispanic, 13% others	**Literacy**					
					TERA	+0.02		-0.13		
					WJ Letter Word ID	+0.10		-0.18		
					WJ Spelling	+0.17	+0.04	-0.06	-0.13	
					Pre-CTOPP/CTOPP	-0.13		-0.13		
					Language					
					PPVT	-0.03		0.00		
					TOLD	+0.08	+0.03	-0.12	-0.06	
Fischel et al. (2007)	Randomized Quasi-Experiment High/Scope	1 year	35 classes 335 students (185E, 150C)	Six Head Start centers in SE New York State 42% African American, 41% Hispanic, 7% White, 8% multiracial; 14% Spanish language dominant	**Literacy**					
					Get Ready to Read! Screen	+0.32				
					Letters Known	+0.31				
					WJ-R Letter Word ID	+0.29				
					WJ-R Dictation	+0.38	+0.20			
					WJ-R Book Knowledge	+0.12				
					WJ-R Print Conventions	+0.23				
					WJ-R Comprehension	-0.12				
					Language					
					PPVT	+0.06	+0.06			

Literacy Express

Study	Design	Duration	Sample	Population	Measure	Effect Size	Subtotal	Effect Size	Subtotal
Lonigan et al (2011)	Cluster Randomized High/Scope	1 year	48 centers 739 students (497E, 242C)	Low SES Head Start and Title I school district preschools in Tallahassee, FL and LA, CA. 56% African American, 35% Hispanic, 8% White, 1% Other	PLS Expressive language	+0.27			
PCER (2008)	Cluster Randomized High/Scope	2 years	12 preschool programs 195 students (99E, 97C)	Public preschool programs in FL; 30% White, 59% African American, 6% Hispanic, 5% others	**Literacy**				
					TERA	+0.17		-0.11	
					WJ Letter Word ID	+0.30		+0.08	
					WJ Spelling	+0.05	+0.17	+0.06	+0.03
					Pre-CTOPP/CTOPP	+0.14		+0.08	
					Language				
					PPVT	+0.17	+0.07	+0.16	+0.13
					TOLD	-0.04		+0.10	

Ready, Set, Leap!

Study	Design	Duration	Sample	Population	Measure	Effect Size	Subtotal	Effect Size	Subtotal
PCER (2008)	Cluster Randomized High/Scope	2 years	39 classes 286 students (149E, 137C)	Preschools from an urban area in New Jersey 78% African American, 20% Hispanic	**Literacy**				
					TERA	+0.08		+0.01	
					WJ Letter Word ID	+0.01		-0.12	
					WJ Spelling	+0.20	+0.05	+0.04	-0.02
					Pre-CTOPP/CTOPP	-0.09		-0.02	
					Language				
					PPVT	+0.15	+0.02	-0.02	-0.03
					TOLD	-0.11		-0.03	

40

					Literacy			
Davidson. Fields, & Yang (2009)	Cluster Randomized High/Scope	1 year	27 classes 254 students (129E, 125C)	High poverty inner-city Newark public elementary schools. Overlaps PCER sample. 14% White, 43% African American, 37% Hispanic, 5% Other, 83% FRL	CTOPP Blending	+0.35		
					DIBELS Letter Naming	-0.10		
					DIBELS Initial Sound Fluency	+0.21		
					WJ Rhyming	+0.19	+0.16	
					WJ Passage Comprehension	+0.09		
					WJ Letter ID	+0.19		
					Language			
					PPVT	+0.01	+0.01	
Abt Associates (2007)	Cluster Randomized Ordinary preschool	18 months	829 students (320E, 509C)	162 child care centers in Miami-Dade County Florida that had 75% or more students eligible for FRL	Literacy			
					TOPEL Print Knowledge		+0.65	
					TOPEL Phonological Awareness		+0.35	+0.50
					Language			
					TOPEL Definitional Vocabulary		+0.28	+0.28

Research-Based, Developmentally Informed (REDI)

Study	Design	Duration	N	Sample	Outcome	ES 1	ES 2	ES 3	ES 4
Bierman et al. (2008); Bierman et al. (2014)	Cluster Randomized; High/Scope or Creative Curriculum	2 years	356 students	44 Head Start classrooms in three counties in PA. 25% African American, 42% White, 17% Hispanic	**Literacy**				
					TOPEL Print Awareness	+0.18	+0.31		
					TOPEL Blending and Elision	+0.43			
					TOWRE Phonemic Decoding			+0.25	+0.08
					TOWRE Sight Word			-0.04	
					WJ LWID			+0.03	
					Language				
					EOWPVT Picture Vocabulary	+0.16	+0.16	+0.10	+0.10

Tools of the Mind

Study	Design	Duration	N	Sample	Outcome	ES 1	ES 2
Barnett et al. (2008)	Cluster Randomized; District program	1 year	18 classes 218 students (85E, 120C)	High poverty urban school district in NJ; 80% free lunch, 92% Hispanic	**Literacy**		
					Get Ready to Read	+0.03	-0.04
					WJ Letter-Word	-0.11	
					Language		
					PPVT	+0.22	+0.16
					EOWPVT	+0.11	

Study	Design	Duration	N	Sample	Outcome	ES				
Farran et al. (2015)	Cluster Randomized Ordinary preschool	3 years	60 classes (32E, 28C) 877 children (498E, 379C)	Schools in 4 rural and 1 urban district in 2 southern states. 41%White, 25% Hispanic, 23% African-American, 6% Asian, 88% free lunch	**Literacy**					
					WJ Letter-Word ID	-0.15	-0.10	-0.17	-0.21	Gr 1: -0.12
					WJ Spelling	-0.04		-0.24		Gr 1: -0.25
					Language					
					WJ Oral Comprehension	-0.07	-0.06	-0.02	-0.01	Gr 1: 0.00
					WJ Picture Vocabulary	-0.04		+0.01		Gr 1: +0.05
					Follow-Up					
					WJ Passage Comprehension (Gr. 1)					Gr 1: +0.04
Waterford										
Fischel et al. (2007)	Randomized Quasi-Experiment High/Scope	1 year	35 classes 335 students (185E, 150C)	Six Head Start centers in SE New York State 42% African American, 41% Hispanic, 7% White, 8% multiracial; 14% Spanish language dominant	**Literacy**					
					Get Ready to Read! Screen	+0.32				
					Letters Known	+0.12				
					WJ-R Letter Word ID	+0.11				
					WJ-R Dictation	+0.02	+0.08			
					WJ Book Knowledge	+0.00				
					WJ Print Conventions	+0.21				
					Comprehension	-0.21				
					Language					
					PPVT	+0.06	+0.06			

Abbreviations:
CTOPP: Comprehensive Test of Phonological Processing
DIBELS: Dynamic Indicators of Basic Early Literacy Skills
DSC: Developing Skills Checklist
FACES: Family and Child Experiences Survey from Head Start
K-SEALS: Kaufman Survey of Early Academic and Language Skills
PALS-Pre-K: Phonological Awareness and Literacy Screening-PreKindergarten

PLS: Preschool Language Scale
PPVT: Peabody Picture Vocabulary Test
PRE-CTOPP: Preschool Comprehensive Test of Phonological and Print Processing
TCAP: Tennessee Comprehensive Assessment Program
TELD: Test of English Language Development
TERA: Test of Early Reading Ability
TOPEL: Test of Preschool Early Literacy
TOLD: Test of Language Development
WJ: Woodcock-Johnson
WJ: COMP: Woodcock-Johnson Passage Comprehension
WJ: DIRECTIONS: Woodcock-Johnson Understanding Directions
WJ: LWID: Woodcock-Johnson Letter-Word Identification
WRMT: Woodcock Reading Mastery Tests

Table 2
Designs, Settings, and Outcomes for Developmental Programs

Study	Design/Control	Duration	N	Sample Characteristics	Posttest	Preschool ES	Preschool Mean ES	Kinder ES	Kinder Mean ES	Post Kinder ES
Creative Curriculum										
Lipsey et al. (2009)	Cluster Randomized Ordinary preschool	5 years	36 pre-K classes (109E, 148C)	Seven school districts in six rural middle TN counties. Overlaps PCER sample. 70% White, 19% African American, 4% Hispanic	Literacy					
					WJ Letter-Word ID	-0.14	-0.11			
					WJ Spelling	-0.08				
					Language					
					WJ Picture Vocabulary	-0.02				
					WJ Story Recall	+0.04				
					WJ Understanding Directions	-0.09	-0.07			
					WJ Oral Comprehension	-0.16				
					PPVT	-0.14				
					Follow-Up					
					TCAP Reading 3rd Grade					Gr. 3: +0.16
PCER (2008) (Tennessee)	Cluster Randomized Ordinary preschool	2 years	14 classes 206 students (101E, 105C)	Seven school districts in six rural counties in TN; 80% White, 18% African American, 11% Hispanic	Literacy					
					TERA	+0.02		+0.10		
					WJ Letter Word ID	+0.16	+0.12	+0.38	+0.20	
					WJ Spelling	+0.19		+0.25		
					Pre-CTOPP/CTOPP	+0.10		+0.06		
					Language					
					PPVT	+0.23	+0.15	+0.12	+0.12	
					TOLD	+0.07		+0.11		

PCER (2008) (North Carolina and Georgia)	Cluster Randomized Ordinary preschool	2 years	18 classes 194 students (97E, 97C)	Head Start centers in NC and GA; 3% White, 85% African American, 8% Hispanic	**Literacy**				
					TERA	-0.08		-0.04	
					WJ Letter Word ID	-0.08	-0.08	0.00	-0.01
					WJ Spelling	-0.18		-0.05	
					Pre-CTOPP/CTOPP	+0.02		+0.06	
					Language				
					PPVT	+0.08	-0.03	+0.15	-0.01
					TOLD	-0.16		-0.17	

Language-Focused Curriculum

PCER (2008)	Cluster Randomized High/Scope	2 years	14 classes 195 students (97E, 98C)	Head Start and public preschool classrooms in VA; 71% White, 21% African American, 4% Hispanic, 3% others	**Literacy**				
					TERA	+0.16		+0.05	
					WJ Letter Word ID	+0.11	+0.18	+0.02	+0.05
					WJ Spelling	+0.25		+0.11	
					Pre-CTOPP/CTOPP	+0.20		+0.03	
					Language				
					PPVT	+0.02	+0.02	-0.09	-0.08
					TOLD	+0.01		-0.07	

Project Approach

PCER (2008)	Cluster Randomized Ordinary preschool	2 years	13 classes 204 students (114E, 90C)	Public preschool programs in WI; 28% White, 40% African American, 17% Hispanic, 13% others	**Literacy**				
					TERA	+0.14		+0.29	
					WJ Letter Word ID	+0.42	+0.22	+0.03	+0.07
					WJ Spelling	+0.27		+0.14	
					Pre-CTOPP/CTOPP	+0.05		-0.17	
					Language				
					PPVT	+0.16	+0.16	+0.10	+0.21
					TOLD	+0.15		+0.32	

Project Construct

PCER (2008)	Cluster Randomized Ordinary preschool	2 years	21 preschool programs 231 students (123E, 108C)	Preschool centers from urban and rural MO; 65% White, 29% African American, 3% Hispanic, 6% others	Literacy				
					TERA	0.00		-0.03		
					WJ Letter Word ID	-0.05	-0.03	+0.16	0.00
					WJ Spelling	-0.15		0.00	
					Pre-CTOPP/CTOPP	+0.10		-0.12	
					Language				
					PPVT	+0.03	-0.01	+0.10	+0.06
					TOLD	-0.05		+0.01	

Abbreviations:
CTOPP: Comprehensive Test of Phonological Processing
DIBELS: Dynamic Indicators of Basic Early Literacy Skills
DSC: Developing Skills Checklist
FACES: Family and Child Experiences Survey from Head Start
K-SEALS: Kaufman Survey of Early Academic and Language Skills
PALS-Pre-K: Phonological Awareness and Literacy Screening-PreKindergarten
PLS: Preschool Language Scale
PPVT: Peabody Picture Vocabulary Test
PRE-CTOPP: Preschool Comprehensive Test of Phonological and Print Processing
TCAP: Tennessee Comprehensive Assessment Program
TELD: Test of English Language Development
TERA: Test of Early Reading Ability
TOPEL: Test of Preschool Early Literacy
TOLD: Test of Language Development
WJ: Woodcock-Johnson
WJ: COMP: Woodcock-Johnson Passage Comprehension
WJ: DIRECTIONS: Woodcock-Johnson Understanding Directions
WJ: LWID: Woodcock-Johnson Letter-Word Identification
WRMT: Woodcock Reading Mastery Tests

Table 3

Summary of Programs and Outcomes by Category

Balanced	Pre-K		Kindergarten	
	Literacy	Language	Literacy	Language
Breakthrough to Literacy			+0.52	+0.31
BELL			+0.06	+0.07
Bright Beginnings				
PCER	+0.21	+0.11	+0.02	+0.12
Lipsey	+0.18	-0.03		
Classroom Links to Early Literacy	+0.19	-0.03		
Curiosity Corner				
PCER	+0.10	-0.05	+0.33	+0.15
Authors		+0.15		
Dialogic Reading + Sound Foundations	+0.12	+0.12	+0.08	+0.13
Direct Instruction (Supplement)	+0.52	+0.46		
DLM Express + Open Court	+0.49	+0.40	+0.47	+0.47
Doors to Discovery	+0.10	+0.16	-0.09	+0.12
ELLM				
Cosgrove	+0.25			
PCER	+0.10	+0.16	+0.11	+0.39
EMERGE	+0.33	+0.13		
Ladders to Literacy	-0.08	-0.30	-0.25	-0.18
Let's Begin				
PCER	+0.04	+0.03	-0.13	-0.06

48

Authors		+0.15		
Dialogic Reading + Sound Foundations	+0.12	+0.12	+0.08	+0.13
Direct Instruction (Supplement)	+0.52	+0.46		
DLM Express + Open Court	+0.49	+0.40	+0.47	+0.47
Doors to Discovery	+0.10	+0.16	-0.09	+0.12
ELLM				
Cosgrove	+0.25			
PCER	+0.10	+0.16	+0.11	+0.39
EMERGE	+0.33	+0.13		
Ladders to Literacy	-0.08	-0.30	-0.25	-0.18
Let's Begin				
PCER	+0.04	+0.03	-0.13	-0.06
Fischel	+0.20	+0.06		
Literacy Express				
Lonigan		+0.27		
PCER	+0.17	+0.07	+0.03	+0.13
Ready, Set, Leap				
PCER	+0.05	+0.02	-0.02	-0.03
Davidson	+0.16	+0.01		
Abt			+0.50	+0.28
REDI	+0.31	+0.16	+0.08	+0.10
Tools of the Mind				
Barnett	-0.04	+0.16		
Farran	-0.10	-0.06	-0.21	-0.01
Waterford	+0.08	+0.06		
Balanced Mean (weighted)	+0.15	+0.08	+0.14	+0.15

Developmental-Constructivist				
	Pre-K		Kindergarten	
	Literacy	Language	Literacy	Language
Creative Curriculum				
Lipsey	-0.11	-0.07		
PCER-Tenn	+0.12	+0.15	+0.20	+0.12
PCER-NC/GA	-0.08	-0.03	-0.01	-0.01
Language-Focused Curriculum	+0.18	+0.02	+0.05	-0.08
Project Approach	+0.22	+0.16	+0.07	+0.21
Project Construct	-0.03	-0.01	0.00	+0.06
Developmental-Constructivist Mean (weighted)	+0.04	+0.03	+0.06	+0.06