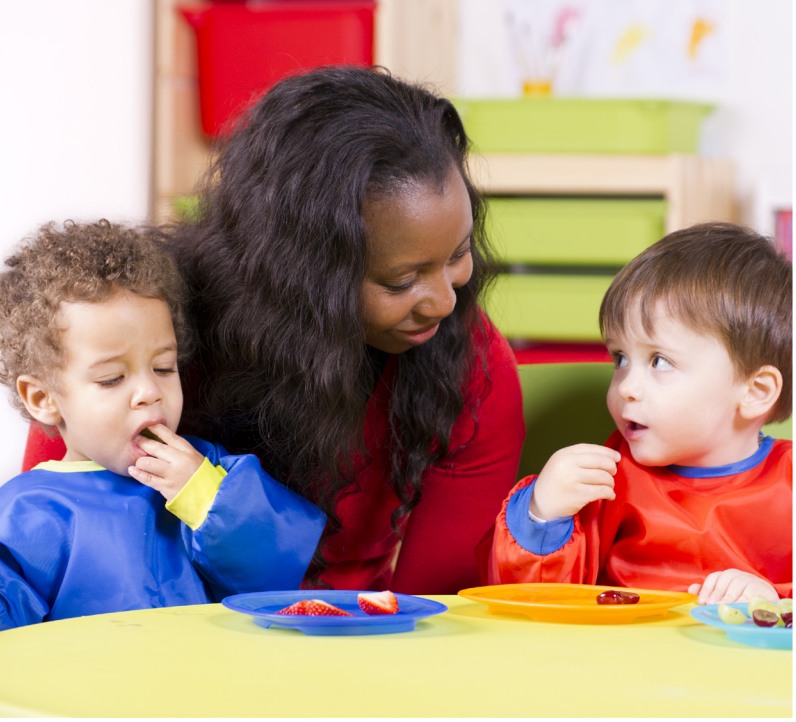
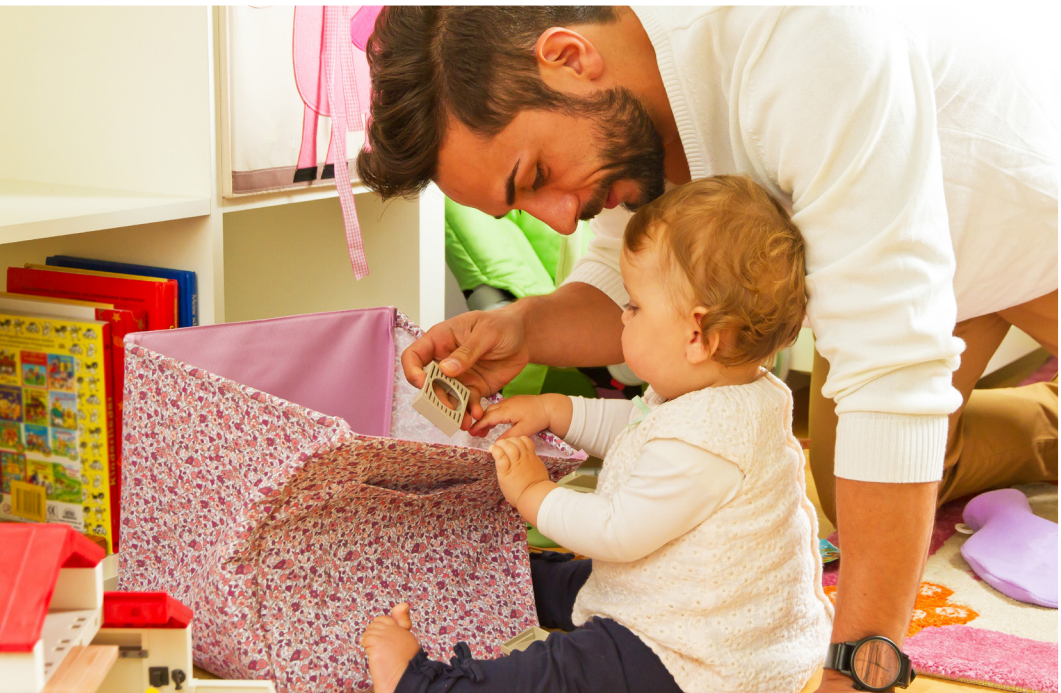


Learning About Infant and Toddler Early Education Services (LITES): A Systematic Review of the Evidence



December 2015

Learning About Infant and Toddler Early Education Services (LITES): A Systematic Review of the Evidence

December 2015

Shannon Monahan, Jaime Thomas, Diane Paulsell, and Lauren Murphy

Contract Number: HHSP23320095642WC/HHSP23337046T

Mathematica Reference Number: 40301

Submitted to:
Office of the Assistant Secretary for Planning and Evaluation
U.S. Department of Health and Human Services
200 Independence Avenue, S.W., Washington, DC 20201
Project Officers: Lisa Trivits and Lindsey Hutchison

Submitted by:
Mathematica Policy Research
P.O. Box 2393
Princeton, NJ 08543-2393
Telephone: (609) 799-3535
Facsimile: (609) 799-0005
Project Director: Diane Paulsell
Reference Number: 40301.126

Suggested citation:

Monahan, S., Thomas, J., Paulsell, D., and Murphy, L. "Learning About Infant and Toddler Early Education Services (LITES): A Systematic Review of the Evidence." Washington, DC: U.S. Department of Health and Human Services, Office of the Assistant Secretary for Planning and Evaluation, 2015.

ACKNOWLEDGMENTS

The authors wish to thank Lindsey Hutchison and Lisa Trivits, our federal project officers at the Office of the Assistant Secretary for Planning and Evaluation (ASPE), as well as Amy Madigan, Meryl Barofsky, and T'Pring Westbrook at the Office of Planning, Research and Evaluation (OPRE) in the Administration for Children and Families (ACF), for their oversight, guidance, and vision. We would also like to thank Jeanne Brooks-Gunn, Margaret Burchinal, and Ellen Kisker for their contributions to this project. In addition, we were fortunate to have the expertise of our expert panel members, whose guidance on program models for infants and toddlers and systematic review methodology was invaluable.

We also gratefully acknowledge our colleagues at Mathematica Policy Research, including Patricia Del Grosso and Christopher Jones, who contributed their time, effort, and expertise to this project. Sally Atkins-Burnett, Emily Sama-Miller, and Sarah Avellar provided thoughtful quality assurance. Maura Butler provided editorial support and Jill Miller formatted this report.

While acknowledging the contributions of these individuals, the authors take responsibility for any errors or omissions that remain. The views expressed in this report do not necessarily reflect those of ASPE or the U.S. Department of Health and Human Services.

This page has been left blank for double-sided copying.

EXPERT PANEL MEMBERS

We would like to thank the following members of our expert panel for their assistance with the LITES Systematic Review. The views expressed in this publication do not necessarily reflect the views of these members.

Clancy Blair
New York University

James Elicker
Purdue University

Diane Horm
University of Oklahoma

Julia Isaacs
Urban Institute

Stephanie Jones
Harvard Graduate School of Education

Brenda Jones Harden
University of Maryland

Jeffrey Valentine
University of Louisville

This page has been left blank for double-sided copying.

CONTENTS

This page has been left blank for double-sided copying.

TABLES

BOXES

This page has been left blank for double-sided copying.

REPORT SUMMARY

A growing body of research indicates that high quality early learning experiences that begin as early as possible in life can promote young children's development and help reduce achievement gaps. Although research is building about programs for preschool children, less is known about effective program models to support infant and toddler early learning in the areas of language, cognitive, and/or social emotional/behavioral development. The Office of the Assistant Secretary for Planning and Evaluation, in partnership with the Administration for Children and Families within the U.S. Department of Health and Human Services, funded Mathematica Policy Research and its partners to conduct the Learning About Infant and Toddler Early Education Services (LITES) project. LITES aimed to identify program models to support infant and toddler early learning in out-of-home early care and education settings to inform future research, policy, and program directions at the federal, state, and local levels. LITES included two main components: (1) a systematic review of the evidence base for program models that aim to support infant and toddler early learning; and (2) a scan of the field for program models that are compelling, but currently lack rigorous research examining impacts on children's outcomes. This report focuses on the LITES systematic review. A second report profiles the 15 models identified in the compelling models scan, including information about model implementation and existing research (Del Grosso et al., 2015).

- The LITES systematic review identified 15 program models with 50 eligible studies that examined the impact of an out-of-home model of early learning services on children's language, cognitive, and/or social emotional/behavioral development. Of these 50 studies, 38 were randomized control trials (RCTs) and 12 were matched comparison group designs (MCGDs).[1] LITES rated 21 of the 50 eligible studies as high- or moderate-quality studies. Of the 15 program models with eligible studies, five models had high- or moderate-quality studies, and four of those models showed evidence of effectiveness on children's outcomes (Abecedarian, Early Head Start, the Infant Health and Development Project, and the Parent-Child Development Centers). Three of the models with evidence of effectiveness measured and demonstrated effects on child outcomes well after the intervention ended (Abecedarian, Early Head Start, and the Infant Health and Development Project).
- The four models that demonstrated evidence of effectiveness for improving child outcomes were all direct multicomponent models that targeted multiple domains of child development. These models began before or soon after birth and continued until at least 36 months of age. All were designed for at-risk children and families and most had a parenting component as part of the intervention. All four models demonstrated favorable end-of-intervention effects in at least one of the following domains: cognitive, language, or social-emotional/behavioral development. Two of the three models with follow-up studies showed at least one favorable long-term impact in one of these domains. However, impacts on health outcomes were varied, with some unfavorable end-of-intervention effects and mixed (favorable and unfavorable) long-term effects. In general, long-term impacts were less consistently favorable than short-term, end-of-intervention impacts.

[1]See the full report for additional information on RCTs and MCGDs.

- The four models that demonstrated evidence of effectiveness on child outcomes were evaluated with randomized controlled trials conducted between 1970 and the late 1990s. No studies were conducted that attempted to replicate the findings of these studies. Further research of contemporary models is needed to build on the foundation established by these studies.
- LITES joins other ongoing federal efforts in promoting broad awareness and use of evidence-based and high quality practices to help children realize their full potential. LITES differs from other federal systematic reviews in that it is a one-time review and is not associated with funding decisions. At this time, only a modest number of program models have studies eligible for review.
- LITES included models in the systematic review if their studies used an eligible research design to estimate impacts on child outcomes (that is, a randomized control trial, matched comparison group, regression discontinuity, or single case design). In contrast, the LITES compelling models scan only examined models nominated by experts and practitioners in the field that lacked rigorous research examining impacts on children's outcomes. An exhaustive scan for all potential programs was beyond the scope of the compelling models report. For example, models in the systematic review whose studies were all rated as low quality were *not* considered for or included in the compelling models report. These models with only low-rated studies are described in Appendices A and B of this report.

EXECUTIVE SUMMARY

A. Purpose and scope of the LITES review

A growing body of research indicates that high quality early learning experiences that begin as early as possible in life can promote young children's development and help reduce achievement gaps (Camilli et al., 2010; Duncan & Magnuson, 2013). Recent research bringing together neuroscience, child development, and economics has made the case that children's early experiences have cognitive, social-emotional, and physical health repercussions that extend into the school years and beyond (Camilli et al., 2010; National Scientific Council on the Developing Child, 2007; Halle et al., 2009; Yoshikawa et al., 2013). Although research is building about programs for preschool children, little is known about effective program models to support infant and toddler early learning. The Office of the Assistant Secretary for Planning and Evaluation (ASPE), in partnership with the Administration for Children and Families (ACF) within the U.S. Department of Health and Human Services, funded Mathematica Policy Research and its partners to conduct the Learning About Infant and Toddler Early Education Services (LITES) project. LITES aimed to identify effective and replicable program models to support infant and toddler early learning in out-of-home early care and education (ECE) settings to inform future research, policy, and program directions at the federal, state, and local levels.

LITES had two main components: (1) a systematic review to identify effective program models to support infant and toddler early learning in out-of-home ECE settings, and (2) a scan of the field for program models that are compelling but lack rigorous research examining impacts on children's developmental outcomes. For both components, we examined infant and toddler early learning models that targeted children's cognitive, language, or social-emotional/behavioral development. For the systematic review, we conducted a comprehensive literature review to identify studies with eligible research designs, rated the quality of the studies, and examined evidence of effectiveness on children's outcomes. In contrast, for the compelling models scan, we identified models through a nomination process and discussion with experts in the field.[2] This report focuses on the systematic review; a second report focuses on findings from the compelling models scan (Del Grosso et al., 2015).

Together, the two components provide a picture of available models to support infant and toddler early learning, including those with rigorous evidence of effectiveness on child outcomes and those considered compelling in the field but lacking rigorous research evidence. LITES, however, does not provide an all-inclusive review of all available infant and toddler early learning models, nor the full range of descriptive research conducted on them. For the compelling models report, in particular, an exhaustive scan for all potential programs was beyond the scope of this report. For example, models in the systematic review whose studies were all rated as low quality were *not* considered for or included in the compelling models report. These models with only low-rated studies are described in Appendices A and B of this report.

[2] We developed the compelling models nomination process to identify models considered compelling by ECE experts. Because it was a nomination process, the compelling models report does not provide a representative or exhaustive list of all possible replicable program models that support infant and toddler early learning in out-of-home ECE settings and lack rigorous research.

Federal agencies have sponsored a growing number of systematic reviews to identify programs and approaches that have the strongest evidence of success. The LITES review differs from other federally funded reviews in that it is a one-time review of the ECE literature. Because of the modest number of program models with studies eligible for review under LITES, subsequent rounds of review are not necessary at this time. This differs from systematic reviews in other fields, such as the What Works Clearinghouse (WWC) or the Clearinghouse for Labor Evaluation and Research (CLEAR), that produce a larger body of eligible studies. In addition, unlike other recurring systematic reviews, such as the Home Visiting Evidence of Effectiveness (HomVEE) systematic review and the Teen Pregnancy Prevention Evidence Review (TPP Evidence Review), LITES is not associated with funding decisions.[3]

B. Methods

To conduct the review, we first carried out a comprehensive literature search for research on replicable program models to support infant and toddler early learning in out-of-home ECE settings in language, cognitive, and/or social-emotional/behavioral development. The search included database searches, web searches, and a call for studies. We included studies of (1) *direct multicomponent models* with out-of-home early learning services for infants and toddlers; (2) *direct enhancement models* that could be layered on another model and focused on improving child outcomes, typically in a single domain; and (3) *indirect enhancement models* that could be layered on another model and focused on improving caregiver practice to support infant and toddler learning.[4] Services had to broadly target infants and toddlers and/or their out-of-home caregivers; programs narrowly targeting children with specific diagnosed disabilities or medical conditions were not included. Studies had to be published in English in 1960 or later. Because the purpose of the review was to identify models with evidence of effectiveness for supporting infant and toddler early learning, eligible studies had to measure at least one child outcome in one of the following domains: cognitive, language, and social-emotional/behavioral development. In addition, studies had to use one of the following four research designs: (1) randomized controlled trial, (2) matched comparison group, (3) single case, or (4) regression discontinuity. Our search did not identify any studies using single case or regression discontinuity designs.

[3] The HomVEE review is linked to the Maternal, Infant, and Early Childhood Home Visiting Program, authorized under the Patient Protection and Affordable Care Act (ACA), which provides states with $1.5 billion over five years to support evidence-based home visiting programs for at-risk pregnant women and children from birth to age 5. The TPP Evidence Review is associated with the Teen Pregnancy Prevention (TPP) program, as part of the Consolidated Appropriations Act, 2010, and the Personal Responsibility Education Program, authorized under the ACA, which together provide more than $175 million annually.

[4] For LITES, *direct multicomponent models* are those composed of more than one replicable component that aims to support the development of infants and toddlers across multiple developmental domains. For instance, model components might include: use of a curriculum; regular child assessment; professional development opportunities for teachers; mental health or health services for children; and/or services for parents to support the healthy development of parents and families. We defined *direct enhancement models* as models that had at least one replicable program component, provided direct early learning services to infants and toddlers in out-of-home ECE settings, and that can be layered on top of an existing model or intervention (for instance, an intervention to promote children's language development that can be delivered in a group care setting). Finally, *indirect enhancement models* have the same definition as direct enhancement models, except that they aim to promote child development by primarily intervening with teachers or caregivers.

After identifying studies eligible for review, a team of trained reviewers assessed the research design and execution of each study using a standard protocol. Each study was assigned a rating of high, moderate, or low to indicate the study's capacity to provide unbiased estimates of program effects. The high rating was reserved for random assignment studies with low attrition of sample members and no reassignment of sample members after initial random assignment. The moderate rating applied to random assignment studies that did not meet the high rating due to flaws in design, execution, or analysis of the data and matched comparison group designs that established baseline equivalence on selected measures and used statistical controls in their analyses. Studies that did not meet criteria for a high or moderate rating were assigned a low rating (details on study ratings appear in Chapter II of this report and the LITES review protocol [Monahan et al., 2015]).

After the team reviewed all studies, we assessed the evidence across studies that received a high or moderate rating for each program model. We considered all child outcomes in the cognitive, language, and social-emotional/behavioral development domains, as well as measures of child health with statistical significance of 0.05 or an effect size of at least 0.20 in absolute value as providing evidence of effectiveness. We examined the evidence on these child outcomes across all studies for each model and assigned each model an effectiveness rating of favorable, mixed, not discernible, or unfavorable effects. We also recorded interim outcomes (such as global child care quality or caregiver knowledge of child development) that met our statistical significance or effect size criteria (see Chapter III). However, because the review aimed to identify models with evidence of effectiveness on child outcomes, we did not consider them as providing evidence of effectiveness.

C. Review results

LITES identified 15 program models with 50 eligible studies (38 randomized control trials and 12 matched comparison group designs), including nine direct multicomponent program models, two direct enhancement models, and four indirect enhancement models. LITES rated 21 of the 50 eligible studies as high- or moderate-quality studies. Five models had at least one high- or moderate-rated study, and four showed evidence of effectiveness on children's outcomes (Table 1).

Table 1. Evidence of effectiveness on child outcomes for program models with high quality research

Program model	Favorable effects on cognitive outcomes	Favorable effects on language outcomes	Favorable effects on social-emotional/ behavioral outcomes	Favorable effects on health outcomes
Abecedarian	Yes	Yes	Yes	Yes
Early Head Start (EHS) (with a center-based component)[a]	Yes	Yes	Yes	No
Infant Health and Development Program (IHDP)	Yes	Yes	Yes	Yes[b]
Parent-Child Development Centers (PCDC)	Yes	Not applicable	Not applicable	Not applicable
Program for Infant/Toddler Care (PITC)	No	No	No	Not applicable

Note: A "Yes" indicates that at least one significant or substantial favorable effect was found at the end of the intervention or at later follow-ups. A significant or substantial effect is statistically significant ($p \leq 0.05$) or has an effect size greater than or equal to 0.2 standard deviations in absolute value. A "No" indicates that there was not at least one significant or substantial favorable effect at the end of the intervention or at later follow-ups. "Not applicable" indicates that no outcomes were measured that rated at least moderate in a study that met eligibility criteria.

[a]EHS included center-based and mixed-approach sites only; home-based sites were excluded.

[b]IHDP demonstrated one unfavorable effect on child health at the end of the intervention, and a mix of one unfavorable effect and several favorable effects in subsequent follow-ups.

The four models with evidence of effectiveness on child outcomes are (1) Abecedarian, (2) Early Head Start (EHS), (3) Infant Health and Development Program (IHDP), and (4) Parent-Child Development Centers (PCDC). All four models demonstrated favorable end-of-intervention effects in at least one of the following domains: cognitive, language, or social-emotional/behavioral development. Two of the three models with follow-up studies showed at least one favorable long-term impact in one of these domains. However, impacts on health outcomes were varied by program model, with some unfavorable end-of-intervention effects and mixed (favorable and unfavorable) long-term effects. In general, long-term impacts were less consistently favorable than short-term, end-of-intervention impacts. We describe these effects in more detail below.

- **Abecedarian** was a program model designed for disadvantaged infants to improve cognitive, language, perceptual-motor, and social development. Children attended a year-round, full-day educational child care program starting at about 3 months of age and ending at about 5 years of age. Abecedarian demonstrated favorable effects in the cognitive domain at the end of the intervention as well as long-term favorable effects in the cognitive, language, social-emotional/behavioral, and child health domains.
- **Early Head Start** is a program serving low-income pregnant women and families with children under 3 years of age. EHS aims to improve child development outcomes, including health, social, cognitive, and language development, as well as to support family, staff, and

community development. LITES reviewed findings for center-based and mixed-approach (combination of center-based and home visiting services) programs only, as the HomVEE review has already reviewed findings for the EHS home-based program approach. EHS demonstrated favorable end-of-intervention effects in the cognitive, language, and social-emotional/behavioral domains. These effects were sustained in the cognitive domain. There were mixed long-term effects in the social-emotional/behavioral domain, and no discernable long-term effects in language development. No effects were detected on child health.

- The **Infant Health and Development Project** focused on premature (gestational age of 37 or fewer weeks) newborns with low birth weights (less than 2,500 grams). The program aimed to support cognitive development, behavioral competence, and children's health. The program began when infants were discharged from the neonatal nursery and continued until children were 3 years old. IHDP demonstrated favorable end-of-intervention effects in the cognitive, language, and social-emotional/behavioral domains. These effects were not sustained. There were also unfavorable end-of-intervention effects on child health and long-term effects were mixed in this domain.
- **Parent-Child Development Centers** were a multisite ECE model for low-income families and children in Birmingham, Houston, and New Orleans. Each site had a slightly different approach and program components, but all sites had programming for mothers and children, and the program lasted until the children were 36 months old. Birmingham was the only site in the LITES review with moderate-rated outcomes, and thus the only one for which we assessed evidence of effectiveness on child outcomes.[5] PCDC demonstrated favorable effects in the cognitive domain at the end of the intervention.

These program models with evidence of effectiveness on child outcomes shared several characteristics. All were direct multicomponent models that began before or soon after birth and continued until at least 36 months of age. All the models focused on supporting children's cognitive and social-emotional development and at least one other developmental domain. Most of the models had a parenting component, in addition to early learning services for children, and all provided individualized early learning, social, and health services tailored to child and family circumstances and needs. All were designed for at-risk children and families. All had an out-of-home ECE service component, but the models varied in terms of their dosage, and two of the four also included home visiting.

Studies of these models were conducted from 1970 to the 2000s. In follow-up studies, Abecedarian, EHS, and IHDP showed long-term favorable effects (Table 2) although these effects varied by model and domain. Specifically, Abecedarian demonstrated long-term favorable effects in the cognitive, language, social-emotional/behavioral, and health domains. Abecedarian also demonstrated significant or substantial effects on long-term risk[6] and economic well-being. EHS demonstrated long-term favorable effects in the cognitive domain, and mixed long-term effects in the social-emotional/behavioral domain. IHDP demonstrated mixed effects on child health in long-term follow-ups. EHS and IHDP also showed effects on interim outcomes. Specifically, EHS had favorable effects in the parent or caregiver knowledge of child

[5] As described in more detail in Chapter II, outcomes within a study often received different ratings. Outcomes in a randomized controlled trial with high attrition, for example, could receive a moderate or low rating, depending on whether baseline equivalence was established and proper statistical controls were used.

[6] The long-term risk domain includes outcomes such cigarette and marijuana use and binge drinking at age 30.

development, global home environment quality, and the parent- or caregiver-child interaction domain. IHDP had two favorable effects on parent-child interaction.

Table 2. Evidence of effectiveness ratings by primary child outcome domain

Program model name	Cognitive development		Language development		Social-emotional/behavioral development		Child health	
	End of intervention effects	Sustained or delayed effects	End of intervention effects	Sustained or delayed effects	End of intervention effects	Sustained or delayed effects	End of intervention effects	Sustained or delayed effects
Abecedarian	Favorable	Favorable	Not applicable	Favorable	Not applicable	Favorable	Not applicable	Favorable
Early Head Start (EHS)[a] (with a center-based component)	Favorable	Favorable	Favorable	No discernible effects	Favorable	Mixed[b]	No discernible effects	Not applicable
Infant Health and Development Program (IHDP)	Favorable	No discernible effects	Favorable	No discernible effects	Favorable	No discernible effects	Unfavorable[c]	Mixed[d]
Parent-Child Development Centers (PCDC)	Favorable	Not applicable	Not applicable	Not applicable	Not applicable	Not applicable	Not applicable	Not applicable
Program for Infant/Toddler Care (PITC)	No discernible effects	Not applicable	No discernible effects	Not applicable	Unfavorable	Not applicable	Not applicable	Not applicable

Notes: Only models that had at least one study rated moderate or high appear in this table. **End of intervention effects** were measured at 36 months and/or at the end of the intervention, if the intervention extended beyond 36 months. **Sustained or delayed effects** were measured one year or more after the end of the intervention. Evidence of effectiveness ratings as described in Chapter II are as follows: favorable (at least one high- or moderate-rated study showed at least one significant or substantial favorable effect, and no high- or moderate-rated study showed any significant or substantial unfavorable effects); mixed (at least one high- or moderate-rated study showed at least one significant or substantial favorable effect, and at least one high- or moderate-rated study showed at least one significant or substantial unfavorable effect); no discernible effects (no high- or moderate-rated study showed any significant or substantial effects, either favorable or unfavorable); and unfavorable (at least one high- or moderate-rated study showed at least one significant or substantial unfavorable effect, and no high- or moderate-rated study showed any significant or substantial favorable effects). **Not applicable** indicates that no outcomes were measured that rated at least moderate in a study that met eligibility criteria.

This table summarizes information from the Domain-Specific Evidence of Effectiveness Ratings table for individual program models in Appendix A.

[a]The EHS home-based approach was not included in LITES. The EHS program model summary in Appendix A presents findings separately for each EHS program approach eligible for LITES (center-based and mixed-approach).

[b]EHS had a mix of one unfavorable effect and one favorable effect in a subsequent follow-up.

[c]This is morbidity during the first three years of life, which is the sum of reported injuries, surgeries, illnesses, and conditions.

[d]IHDP had a mix of one unfavorable effect and several favorable effects in subsequent follow-ups.

None of the four models with evidence of effectiveness on child outcomes had additional studies with non-overlapping samples that attempted to replicate the findings from the original studies. In addition, none of the more recently developed models showed evidence of effectiveness on child outcomes, nor did the direct or indirect enhancement models, primarily because all studies were rated low. One exception was the Program for Infant/Toddler Care (PITC), which had one high-rated study that did not show favorable effects on child outcomes. PITC combines direct caregiver training and on-site coaching or tailored assistance for center-based and family child care providers. PITC aims to indirectly improve children's language, cognitive, and social-emotional development by working with providers. Four of five child outcomes measured were null, and one outcome in the social-emotional/behavioral domain was unfavorable. PITC also had one favorable effect in the parent- or caregiver-child interaction domain. Because this was an interim outcome in the context of this review, rather than a child outcome, it does not demonstrate evidence of effectiveness for this review.

D. Suggestions for future research

LITES identified a number of strengths and gaps in the research base for program models to support infant and toddler early learning in out-of-home ECE settings. Strengths included several models with evidence of effectiveness for child outcomes based on well-implemented RCTs. Three of these models had studies showing long-term favorable effects in the cognitive, language, social-emotional/behavioral, and health domains. The research base is limited, however, and LITES identified a number of gaps. Further research can build on the foundation established by these seminal studies.

In particular, more work is needed to promote development of program models that are replicable and ready for evaluation. Several innovative strategies exist for supporting model development and testing. In recent years, ACF and a number of its partner agencies have launched research networks that bring groups of researchers together around different topics. The Network for Infant/Toddler Researchers (NITR), sponsored by OPRE, could serve as a forum for supporting development of ECE models for infants and toddlers. NITR brings together federal staff and researchers with expertise in developmental science, implementation, professional development, and data use to identify existing research for informing policy; identify research gaps; and build capacity to conduct research that can inform infant and toddler programs. For example, the NITR Program Practices Workgroup is collaborating with the Quality Initiatives Research and Evaluation Consortium, another ACF-sponsored working group, to examine how state QRIS systems can include assessments of curricula for infants and toddlers.

Collaborative innovation and improvement networks (CoIINs) aim to advance breakthrough improvements in specific programs and topics of concern by supporting learning communities of practitioners, researchers, and experts. These networks aim to develop innovative practices and to improve outcomes, informed by practitioner knowledge and current research. The networks engage in Plan-Do-Study-Act cycles to design innovations, test them in practice settings, reflect on the results, and refine them as needed. For example, the HRSA-sponsored Home Visiting Collaborative Improvement and Innovation Network brings together teams of researchers and implementing agencies to develop and test innovations in areas of home visiting practice for which evidence-based practice does not yet exist, such as strategies to increase duration of breastfeeding and early detection of developmental delays.

In the private sector, early learning labs aim to accelerate experimentation and development of scalable early learning interventions by bringing together experts in early childhood, design, and innovation. For example, the Oakland-based Early Learning Lab supports testing of new innovations in several California communities funded by the Packard Foundation-sponsored Starting Smart and Strong Initiative. With support from the Early Learning Lab, these communities will engage in rapid cycle testing of innovations in formal and informal early care and education settings to accelerate the development of scalable interventions that positively impact children's school readiness. All of these strategies involve bringing together groups of experts to collaborate, innovate, and experiment and offer opportunities for model and practice development and evaluation.

To address identified gaps and strengthen the research base, we offer the following recommendations for future evaluation efforts:

- Conduct research using designs with strong internal validity
- Conduct replication studies to confirm findings and assess external validity
- Collect and report data on baseline characteristics for the analytic sample used to assess program effects
- Incorporate strategies to reduce the risk of finding statistically significant findings by chance due to multiple comparisons
- Design studies to incorporate assessment of effects on diverse subgroups of infants and toddlers
- Report effect sizes
- Incorporate assessment of implementation fidelity into studies of program effectiveness
- Conduct planned variation studies to identify which components of program models contribute to program effects

Overall, there is still much to learn about what works for infants and toddlers. The field should continue building the knowledge base to identify a broader range of effective programs that can help young children reach their full potential.

I. PURPOSE AND FOCUS OF THE REVIEW

A growing body of research indicates that high quality early learning experiences that begin as early as possible in life can promote young children's development and help reduce achievement gaps (Camilli et al., 2010; Duncan & Magnuson, 2013). Recent research bringing together neuroscience, child development, and economics has made the case that children's early experiences have cognitive, social-emotional, and physical health repercussions that extend into the school years and beyond (Camilli et al., 2010; National Scientific Council on the Developing Child, 2007; Halle et al., 2009; Yoshikawa et al., 2013). Although research is building about programs for preschool children, less is known about effective program models to support infant and toddler early learning.

The Office of the Assistant Secretary for Planning and Evaluation (ASPE), in partnership with the Office of Planning, Research and Evaluation (OPRE) in the Administration for Children and Families (ACF) within the U.S. Department of Health and Human Services, funded Mathematica Policy Research and its partners to conduct the Learning About Infant and Toddler Early Education Services (LITES) project. LITES aimed to identify effective and replicable program models that support infant and toddler early learning in out-of-home early care and education (ECE) settings to inform future research, policy, and program directions at the federal, state, and local levels. LITES joins other federal efforts, including the Home Visiting Evidence of Effectiveness (HomVEE) project and other federally funded systematic evidence reviews in promoting broad awareness and use of evidence-based and high quality practices to help children realize their full potential.

LITES had two main components: (1) a systematic review to identify effective program models that support infant and toddler early learning; and (2) a scan for early learning program models that are compelling to the field but currently lack rigorous research examining impacts on infant and toddler developmental outcomes. Both project components focused on program models that take place in out-of-home ECE settings. Because we wanted to review models with the potential for replication, LITES focused on well-specified models that included a defined package of components of infant and toddler early learning services or professional development to help caregivers support infant and toddler early learning.[7] This report focuses on the systematic review.

For both components, we examined infant and toddler early learning models that targeted children's cognitive, language, or social-emotional/behavioral development. For the systematic review, we conducted a comprehensive literature review to identify studies with eligible research designs, rated the quality of the studies, and examined evidence of effectiveness on children's outcomes. In contrast, for the compelling models we identified models through a nomination

[7] We defined *well-specified models* as those that had (1) clear inclusion and exclusion criteria that define the population for which the model is intended, (2) a clear description of the model components or features that must be present, and (3) clear practice guidance to promote consistency of service delivery (such as the availability of implementation guides and staff training materials, requirements for staff qualifications, or the availability of ongoing technical assistance; Fixsen et al., 2013).

process and discussion with experts in the field.[8] This report focuses on the systematic review; a second report focuses on findings from the compelling models scan (Del Grosso et al., 2015).

Together, the two components provide a picture of available models to support infant and toddler early learning, including those with rigorous evidence of effectiveness on child outcomes and those considered compelling in the field but lacking rigorous research evidence. LITES, however, does not provide an all-inclusive review of all available infant and toddler early learning models nor the full range of descriptive research conducted on them. For the compelling models report, in particular, an exhaustive scan for all potential programs was beyond the scope of this report. For example, we did not consider models with only studies rated as low quality in the systematic review for the compelling models scan. These models with only low-rated studies are described in Appendixes A and B of this report.

Federal agencies have sponsored a growing number of systematic reviews to identify programs and approaches that have the strongest evidence of success. The LITES review differs from other federally funded reviews in that it is a one-time review of the ECE literature. Because of the modest number of program models with studies eligible for review under LITES, subsequent rounds of review are not necessary at this time. This differs from systematic reviews in other fields, such as the What Works Clearinghouse (WWC) or the Clearinghouse for Labor Evaluation and Research (CLEAR), which produce a larger body of eligible studies. In addition, unlike other recurring systematic reviews, such as HomVEE and the Teen Pregnancy Prevention Evidence Review (TPP Evidence Review), LITES is not associated with funding decisions.[9]

The systematic review included 15 program models to support infant and toddler early learning in out-of-home ECE settings (Box I.1). The review included program models that provided direct early learning services to infants and toddlers in out-of-home ECE settings and indirect early learning services through professional development services to adult out-of-home caregivers. We included program models that focused on directly influencing children's cognitive, language, and/or social-emotional/behavioral development and had multiple program components (direct multicomponent models). We also included enhancement models that could be layered on another model and typically focused on improving child outcomes in a single domain (direct enhancement models) or focused on improving caregiver practice (indirect enhancement models). In this report, we summarize available research for each of these three categories of models, including information on the quality of studies conducted and evidence of effectiveness for eligible models.

[8] We developed the compelling models nomination process was intended to identify models considered compelling by ECE experts. Because it was a nomination process, the compelling models report does not provide a representative or exhaustive list of all possible replicable program models that support infant and toddler early learning in out-of-home ECE settings that lack rigorous research examining impacts on children's developmental outcomes.

[9] The HomVEE review is linked to the Maternal, Infant, and Early Childhood Home Visiting Program, authorized under the Patient Protection and Affordable Care Act (ACA), which provides states with $1.5 billion over five years to support evidence-based home visiting programs for at-risk pregnant women and children from birth to age 5. The TPP Evidence Review is associated with the Teen Pregnancy Prevention (TPP) program, as part of the Consolidated Appropriations Act, 2010, and the Personal Responsibility Education Program, authorized under the ACA, which together provide more than $175 million annually.

In consultation with ASPE, OPRE, and an expert work group of researchers, the LITES team established criteria to identify program models eligible for inclusion in LITES and criteria to identify research studies eligible for review. We describe these criteria in the rest of this chapter. In Chapter II, we describe how we identified research on program models and the criteria we used to assess the quality of that research. Chapter III summarizes the evidence of effectiveness on short- and long-term outcomes across program models. In Chapter IV, we describe the lessons learned from the systematic review, identify research gaps, provide recommendations for future research, and suggest policy implications. Summaries of the program models and their effects are included in Appendix A for direct multicomponent models and in Appendix B for direct and indirect enhancement models.

Box I.1. Infant and toddler ECE models identified as eligible for the LITES systematic review

Direct multicomponent models

- Abecedarian
- Brookline Early Education Project
- Early Head Start
- Infant Health and Development Program
- Milwaukee Infant Stimulation Project
- Parent-Child Development Centers
- Project CARE
- Sure Start Local Programmes
- Yale Child Welfare Research Program

Direct enhancement models

- LearningGames
- Music Education

Indirect enhancement models

- Eager and Able to Learn
- Learning Language and Loving It
- Project Secure Child in Child Care
- Program for Infant/Toddler Care

A. Characteristics of program models eligible for review

The LITES systematic review focused on program models designed to improve outcomes in language, cognition, and/or social emotional/behavioral development for infants and toddlers. To be considered eligible for inclusion in the LITES review, we required program models to meet the following criteria:

- **Replicable components with a focus on supporting early learning.** Eligible models fell into one of three categories:
  - *Direct multicomponent models* provided a defined set of replicable program components, including direct early learning services to infants and toddlers in out-of-home ECE settings.
  - *Direct enhancement models* had at least one replicable program component and provided direct early learning services to infants and toddlers in out-of-home ECE settings.
  - *Indirect enhancement models* consisted of professional development programs with replicable program components focused on helping adult out-of-home caregivers to

support infant and toddler early learning, rather than primarily targeting infants and toddlers.

- **A focus on infants and toddlers.** The target population for the models had to include infants and toddlers, defined as children from birth to age 36 months, or their adult out-of-home caregivers. Models could include children from other age groups as well. For example, models could target children from birth to age 5, or the programs could begin prenatally. However, the primary focus of the models had to be supporting infant and toddler early learning in out-of-home ECE settings.
- **Broad targeting.** Models had to be targeted broadly to infants and toddlers and/or their adult out-of-home caregivers. Models narrowly targeting infants and toddlers with diagnosed disabilities or specific medical conditions were not included in the review.[10] However, models targeting broad groups of at-risk infants and toddlers (for example, children from low-income families or low birth weight children) were eligible for inclusion.[11]
- **Out-of-home delivery.** Services had to be provided outside of the children's homes. Models could be implemented in center-based settings, such as child care centers, or in home-based settings, such as family child care homes or informal caregivers' homes. Program models that provided supplemental home visits were eligible for inclusion in the review, but the primary setting had to be out-of-home care. Similarly, program models that provided supplemental services in areas such as nutrition, health and developmental screening, supports for parents, and referrals to other community resources were considered for inclusion in the review. However, the primary focus of services delivered outside the child's home had to be supporting infant and toddler early learning.
- **Specific criteria for indirect enhancement models.** Professional development programs delivered to adult out-of-home caregivers were eligible for inclusion in the review if the programs involved intervening directly with caregivers, took place in the caregiving or a similar setting, and focused on helping caregivers support infant and toddler early learning.[12]

For an eligible program model to be included in the systematic review, it had to have at least one study that met the LITES study inclusion criteria outlined below.

[10] The federal government currently makes specific investments in special education and to support the development of children with disabilities. This review focused on identifying effective program models for supporting early learning among a broad range of infants and toddlers.

[11] Although the review targets children broadly, subgroups of particular interest include children from low-income families, dual-language learners and immigrants, children from minority racial and ethnic groups, children with special needs, and children in author-defined risk groups. Findings for subgroups that met LITES eligibility criteria appear in the program model summaries in Appendix A.

[12] Other indirect services—such as parenting, family self-sufficiency, or referral services—were not eligible for the review, because they did not target children's early learning in out-of-home care settings. However, outcomes in some of these domains—such as parenting—were recorded as part of LITES, if reported in the original studies. See Table II.3 for a full list of LITES outcome domains.

B. Characteristics of studies eligible for review

To be considered eligible for inclusion in the LITES review, we required that studies meet the following criteria:

- **Study sample.** Study samples had to include children enrolled in the program before 36 months of age (including prenatal enrollment). If the sample contained children older than the target age range, we report on disaggregated results for those enrolled before age 36 months, when possible. If disaggregated study results were not available, we required that 50 percent or more of the sample be younger than 30 months at the time of program enrollment.
- **Outcomes of interest.** We required that the study include at least one outcome in any of the following child outcome domains:[13]
  - **Cognitive development,** including outcomes such as attention, memory, object permanence, concept development and categorization, understanding relationships (for example, cause and effect), spatial reasoning, and problem solving.
  - **Social-emotional/behavioral development,** including outcomes such as emotion regulation, impulse control, sociability, and attachment.
  - **Language development,** including outcomes such as receptive language, expressive language (including gestures), joint attention, and emergent literacy skills (for example, listening comprehension).
- **Language of publication.** The study must have been published in English.
- **Publication time frame.** The study must have been published in 1960 or later.
- **Study design.** Eligible designs for review included randomized controlled trials (RCTs), matched comparison group designs (MCGDs), single case designs (SCDs), and regression discontinuity designs (RDDs).

In the next chapter, we describe the criteria we used to evaluate the quality of the studies in the LITES review.

[13] Child health outcomes such as height, weight, and hospitalizations were also reported if present in a study of a *model* that had at least one study with child outcomes in a cognitive, social-emotional/behavioral, or language domain.

This page has been left blank for double-sided copying.

II. METHODS

To identify and evaluate studies consistently and objectively, it is important to use a systematic, well-specified, and transparent process. Therefore, we developed a search and screening procedure to ensure that we could find potentially relevant literature and efficiently identify the studies and models eligible for review. To evaluate studies, we specified criteria to assess the quality and evidence of effectiveness on child outcomes of the studies prioritized for review.

In this chapter, we summarize the LITES review protocol for (1) identifying studies and models eligible for review, (2) assessing the quality of the research, and (3) assessing the evidence of effectiveness on child outcomes.

A. Identifying relevant literature

We used a multistep process to search, screen, and identify studies eligible for review. First, we conducted a comprehensive search for relevant literature in databases and issued a call to researchers and stakeholders in the ECE field for relevant studies. We then screened the studies to identify citations potentially relevant to the review. From the studies that passed this screening, we identified program models for review, and then conducted a targeted search for studies of the identified models using the model name as a search term.

1. Searching for relevant studies

The search encompassed relevant literature, including unpublished literature that aligned with the LITES review scope and study inclusion criteria. Focused and carefully ordered search terms, developed to align with the inclusion criteria described in Chapter I, optimized our ability to find relevant literature on eligible models without capturing a large volume of irrelevant literature (for more details, see the LITES review protocol [Monahan et al., 2015]). We searched titles, abstracts, subjects, and keywords within numerous databases. Mathematica librarians used advanced searching techniques, such as proximity searches, to help us pinpoint relevant literature. The librarians saved literature search results in an online bibliographic management system that enabled storing, screening, and sorting a customized list of study citations and abstracts.

To ensure a thorough and comprehensive literature search, we compared our results with the studies included in the recent National Institute of Child Health and Development meta-analysis, *The effects of early childhood programs on children: a comprehensive meta-analysis*[14] and earlier literature reviews (Karoly et al., 2005; Leak et al., 2010; Meisels & Shonkoff, 1990; Shonkoff & Meisels, 2000). While LITES builds on this prior work, it is unique in its contribution to the field. The meta-analysis included a broad range of ECE programs that served children prenatally to age 5, and its purpose was to estimate average effects across these programs. In addition, unlike LITES, the meta-analysis and literature reviews did not conduct a detailed assessment of study quality.

[14] Personal communication with Magnuson, November 18, 2013.

Finally, we asked researchers, practitioners, and other stakeholders in the field to recommend or submit research articles and reports to locate studies not found in peer-reviewed journals, such as dissertations and unpublished studies.[15] We distributed a call for studies using a broad range of electronic mailing lists, including research and policy organizations, key early childhood professional associations and practitioner groups, and university-affiliated research centers. We added to our list of citations from the resulting submissions that were not already represented in our records.

2. Screening the literature to identify relevant studies

We used a screening procedure to identify the most relevant citations. We removed duplicate citations and publications that were not studies (such as letters to the editor, book reviews, and press releases) and then screened for the following factors:

- **English publication.** Excluded studies not published in English.
- **Policy relevant.** Excluded studies of models delivered in a developing-world context.
- **Possible to attribute effects solely to the model of interest.** Excluded studies in which it was not possible to attribute effects solely to the model of interest. For example, studies in which a direct multicomponent model of interest was combined with another direct multicomponent intervention were excluded.
- **Published 1960 or later.** Excluded studies published before 1960.
- **Primary study.** Excluded summaries of studies reported elsewhere (for example, literature reviews or meta-analyses).
- **Target population in range.** Excluded studies in which the children or families were not enrolled in the program model before the child reached 36 months of age. To target models for children from birth to 36 months of age, we required results disaggregated for those enrolled before age 36 months. If disaggregated study results were not available, we required that 50 percent or more of the sample be younger than 30 months at the time of program enrollment.[16]
- **Services relevant to the review.** Excluded studies that were not: (1) direct multicomponent models that provided a defined set of replicable program components, including early learning services to infants and toddlers in out-of-home ECE settings; (2) direct enhancement models with at least one replicable program component that provided early learning services to infants and toddlers in out-of-home ECE settings; or (3) indirect enhancement models consisting of professional development programs with replicable program components focused on helping adult out-of-home caregivers to support infant and toddler early learning. For example, models that provided primarily medical, parenting

[15] The initial database search occurred from January 21, 2014, to February 26, 2014, and the call for studies was open from January 28, 2014, to March 25, 2014.

[16] This criterion is similar to several WWC review protocols that use a 50 percent threshold for defining eligible study samples when results are aggregated (such as the Early Childhood Education for Children with a Disability topic area protocol). To exclude ECE services that focused primarily on children 36 months and older, we set the threshold at 30 months.

education, or nutrition services would not be eligible for the review. Also excluded were studies in which services primarily targeted children with specific disabilities or medical conditions.

- **Primary service delivery location is out of the home.** Excluded studies of models in which out-of-home ECE services were not the primary service delivery mechanism—for example, those that primarily delivered services through home visits.[17]
- **Replicable program model.** Excluded studies in which the ECE services under study did not include a defined package of replicable program components.
- **Subgroups out of scope.** Excluded studies that only reported on subgroups that were not the LITES pre-identified subgroups of interest.
- **Eligible outcomes.** Excluded studies that did not measure at least one child outcome in one of the following domains: cognitive, language, or social-emotional/behavioral development.
- **Eligible design.** Excluded studies that did not use one of the eligible designs: RCTs, MCGDs, SCDs, or RDDs. We coded ineligible designs in the database to retain supplemental information about the models prioritized for the systematic review.[18]

3. Identifying models and conducting a targeted search for selected models

After the search and screening phases, we analyzed the remaining citations and identified a list of replicable program models with eligible studies for review. We sorted the resulting list of models into three categories of program models. The first included direct multicomponent models that provided out-of-home early learning services for infants and toddlers. The second included direct enhancement models that could be layered on another model and typically focused on improving child outcomes in a single domain. The third included indirect enhancement models that could be layered on another model and focused on improving caregiver practice. After identifying program models for review, we repeated the search and screening process using model names as key search terms to ensure inclusion of all available studies.[19]

Although our search and call for studies identified more than 9,000 citations, we reviewed a much smaller group after applying screening procedures (Table II.1). We eliminated about 90 percent of the citations at the initial relevance screening stage, mainly because the studies focused on services not relevant to the review and because the target population was out of the specified age range. At the second screening stage, we eliminated citations that had ineligible study designs, examined services not relevant to the review (which was not always apparent at the initial screening stage), or did not measure relevant outcomes. After applying the additional screening criteria for relevance and design, we reduced the list to 50 eligible studies.

[17] Research on hybrid models (such as models that include both home visiting and center-based components) could be included if out-of-home services were the primary service delivery mechanism.

[18] Studies of process, fidelity, cost, sustainability, and implementation, as well as correlational, descriptive, pre-post design, and ethnographic studies, were not eligible for review because these study designs do not allow a researcher to confidently determine that the intervention under study caused changes observed in children's outcomes.

[19] The specific search for the direct multicomponent models occurred the week of May 12, 2014, and the specific search for the direct and indirect enhancement models occurred the week of August 25, 2014.

Table II.1. Study screening disposition

	Total studies
Search for relevant literature	
Database search hits (de-duplicated)	9,073
Call for studies submissions	9
Total	9,082
Initial relevance screening in online bibliographic management system	
Studies screened in	976
More information needed to screen	58
Screened out	8,048
Non-studies	333
Home visiting is the primary program element	203
Non-English	6
Not a primary study (such as a literature review or meta-analysis)	400
Not policy relevant[a]	218
Publication date out of range	0
Study on services not relevant to review[b]	5,522
Target population out of range	1,867
Additional relevance and design screening	
Screened in	50
Direct multicomponent model	44
Direct enhancement model	2
Indirect enhancement model	4
Ineligible for review	20
Handbook or conference proceedings	4
Could not obtain full text	7
Supplemental materials	7
Study is not the most recent and complete version available	2
Screened out	964
Additional source	19
Home visiting is the primary program element	38
Ineligible study design	251
International and not policy relevant	47
Not possible to attribute effects solely to the model of interest	9
No eligible outcomes	179
Non-English	3
Not a primary study (for example, a literature review)	160
Study does not examine a replicable program	82
Publication date out of range	0
Study on services not relevant to review	206
Subgroups out of scope	3
Target population out of range	163

Notes: Some studies were screened out for multiple reasons.

[a]*Not policy relevant* refers to models delivered in a developing world context.

[b]Examples of services *not relevant for review* include medical, parenting education, and nutrition services.

B. Rating study quality

We applied the LITES study quality standards to consistently and objectively assess the quality, or internal validity, of the 50 studies identified for review.[20] We began by considering the criteria from existing evidence reviews and clearinghouses, including the WWC (WWC, 2011; 2014), the HomVEE review (HomVEE, 2014), and the HHS TPP Evidence Review (TPP Evidence Review, 2012). We then adapted the criteria to take into account the distinct features and needs of the infant-toddler ECE research field. For example, unlike other reviews, LITES did not require studies to demonstrate baseline equivalence on pre-test measures of child outcomes. Baseline measures of many outcomes (for example, measures of language development) are not feasible for children enrolled prenatally or soon after birth. In addition, based on a recommendation of the LITES expert panel, we required studies to demonstrate baseline equivalence on child age, unlike HomVEE. Other research suggests that child age is highly correlated with outcomes for infants and toddlers. In this section, we define the study designs eligible for review and describe our criteria for assessing study quality and assigning study ratings.[21]

1. Eligible study designs

The four designs eligible for the LITES review are (1) RCTs, (2) MCGDs, (3) SCDs, and (4) RDDs. Because all the studies we reviewed for LITES were RCTs or MCGDs, we focus on those designs (Box II.1).[22]

RCTs and MCGDs are group designs, meaning that they feature an intervention group that receives program services and a comparison group that does not. When properly executed, these designs allow evaluators to confidently conclude that differences in outcomes between the groups can be attributed to the intervention and not to other factors. This is because the comparison group serves as a counterfactual, or an estimate of what would have happened to the intervention group in the absence of the intervention. For these types of designs to produce reliable estimates of program effects, they must present convincing evidence that the intervention and comparison groups were similar at baseline—that is, at the onset of the study. In contrast, study designs without a comparison group (for example, pre-post designs) offer no way to assess what participants' outcomes would have been in the absence of the intervention, and thus are not eligible for review.

[20] Internal validity is a study's ability to isolate the effects of a program or intervention from other factors that may influence participants' outcomes.

[21] The standards were applied by trained reviewers. All staff and consultants on the LITES project were required to sign a conflict of interest statement.

[22] The LITES study quality criteria for SCDs and RDDs were the same as those for the WWC pilot SCD and RDD standards (WWC, 2011; 2014).

Box II.1. Description of eligible study designs

Randomized Controlled Trials (RCTs) use random assignment to create two or more groups that are, on average, similar to each other at the onset of the study (at baseline).[23] These studies provide strong evidence that differences in the outcomes between the intervention and comparison groups after the implementation of an intervention (at follow-up) can be attributed to the intervention rather than to preexisting differences between the groups (Shadish et al., 2002). Of the 50 studies reviewed, 38 were RCTs.

Matched Comparison Group Design (MCGD) participants are sorted into groups through a process other than random assignment. Even if the intervention and comparison groups are well matched based on observed characteristics at baseline, they may still differ on unmeasured characteristics. We cannot rule out the possibility that the findings are attributable to unmeasured group differences; therefore, we have somewhat less confidence in the conclusions of these types of studies than in RCTs. Of the 50 studies reviewed, 12 were MCGDs.

2. Description of ratings

The study quality standards focused on internal validity—that is, a study's ability to isolate the effects of a program from other factors that may influence participants' outcomes. We used three ratings—high, moderate, and low—to assess a study's internal validity. The high rating was reserved for well-executed RCTs with low attrition, no reassignment, and no confounding factors (Table II.2). The moderate rating applied to MCGDs and to RCTs with high attrition and/or reassignment that demonstrated baseline equivalence, applied statistical controls, and had no confounding factors. Low-rated studies used an eligible design but did not meet the requirements for a high or moderate rating. The decision trees in the separate LITES review protocol (Monahan et al., 2015) depict the study rating process applied by reviewers who were trained to assess study quality.

Table II.2. Summary of study rating criteria for the LITES review

LITES study rating	RCTs	MCGDs
High	Random assignment Low attrition No reassignment No confounding factors	Not applicable
Moderate	If there was reassignment or high attrition, highest possible rating was moderate and MCGD rating criteria applied	Baseline equivalence established on required measures Proper statistical controls used No confounding factors
Low	Studies that did not meet the requirements for a high or moderate rating	

3. Threats to internal validity

Confounding factors. Confounding factors, or "confounds," threaten the internal validity of RCTs and MCGDs because, if a confounding factor is present, a study cannot distinguish between the effect of that factor and the intervention of interest. A confounding factor is often defined as a third variable related to both the independent variable and dependent variable, and that might account for the observed relationship between the two. In many cases, this occurs

[23] If random assignment is applied appropriately, then there are no *systematic* differences between the two groups at baseline; however, there may be chance differences. Chance differences may be more likely with small sample sizes.

when some aspect of the design lines up exactly with either the intervention or comparison group. For example, if there is only one classroom in the intervention group, intervention effects are indistinguishable from classroom effects. Thus, it is impossible to determine whether the intervention or another feature of the classroom, such as the teacher or the composition of the students, caused the observed outcomes.

Attrition. In the context of rating study quality, attrition, or losing participants from the study, is only examined for RCTs.[24] These studies depend on the assumption that randomization results in intervention and comparison groups that are similar at baseline. Attrition can compromise the initial equivalence of the groups and lead to biased estimates of an intervention's effects. The cause of the attrition may be related to the program under study, rendering it difficult to know whether an observed difference is due to the program itself, or to whatever caused the attrition.

Both overall and differential attrition can contribute to bias in the estimated effect (Box II.2). Overall attrition is the total percentage of participants who left the study between the point of random assignment and the follow-up period. Differential attrition is the difference in attrition between the intervention and comparison groups.

Box II.2. Overall and differential attrition

To illustrate overall and differential attrition, consider a hypothetical study that randomly assigned 100 children to the intervention group and 100 to the comparison group. Suppose that, at the end of the intervention, 80 children remained in the intervention group and 70 remained in the comparison group. In this example, the overall attrition rate would equal the total number of children who left the study divided by the total number of children randomly assigned: 50/200, or 25 percent. The differential attrition rate is the absolute value of the difference between the attrition rates in the intervention and comparison groups: |20/100 – 30/100|, or 10 percent.

Study designs without random assignment. In MCGDs, group assignment is nonrandom, and we cannot rule out the possibility that groups differ in unobservable ways at baseline, even if they appear similar on characteristics that were measured. Unmeasured baseline differences can bias estimates of the intervention's impact. For example, if the intervention group contains families who, before the intervention, provided more developmental materials for their children at home than families in the comparison group (and if this difference was not controlled for in impact analyses), researchers might find cognitive development impacts that appear to be due to the intervention but are instead due to this preexisting difference between the study groups. As such, LITES had criteria requiring that MCGDs classified with strong internal validity demonstrate that the intervention and comparison groups were similar on key observable characteristics at baseline and that statistical controls were used to minimize bias in impact estimates.

4. Standards to address threats to internal validity

Studies with confounds receive low ratings. For this review, a low rating is assigned to RCTs or MCGDs with only one unit (for example, one child, one classroom, one center) in the

[24] Attrition is not a factor examined for MCGDs because for this study design type, only the *analytic* sample is considered when determining study quality.

intervention and/or comparison condition or other confounding factors such as systematic differences in data collection procedures between the intervention and comparison groups.

Attrition standards set cutoffs for acceptable levels of attrition.[25] The cut-off for an acceptable—in other words, low—level of sample attrition is tied to the extent of overall and differential attrition and to a combination of the two. There is a trade-off between overall and differential attrition: studies with a small amount of overall attrition can have a higher level of differential attrition, and vice versa. For example, an RCT with overall attrition of 10 percent could have up to approximately 6 percent differential attrition and have an acceptable level of attrition, whereas a study with 50 percent overall attrition would need less than 2 percent differential attrition to be acceptable. Monahan et al. (2015) contains a depiction of acceptable attrition ranges. In LITES, RCTs with combinations of overall and differential attrition that fall in the acceptable range were deemed to have low attrition and received a high rating (barring other issues, such as reassignment). RCTs with high attrition were reviewed with the same criteria as MCGDs. The highest possible rating for RCTs with high attrition was moderate.[26] Monahan et al. (2015) provides additional detail on LITES attrition standards.

Studies had to establish baseline equivalence. For MCGDs and RCTs with high attrition or reassignment, baseline equivalence of intervention and comparison groups is a key concern. MCGDs and RCTs with high attrition had to demonstrate baseline equivalence to receive a moderate rating (the highest possible rating for studies of these types). Demonstrating baseline equivalence means showing that the intervention and comparison groups have similar observable characteristics at baseline. This supports conclusions that the intervention—rather than preexisting differences—led to the observed outcome (Shadish et al., 2002). For this review, equivalence had to be established on the final analytic sample used in the analysis of follow-up outcomes (not the baseline sample).[27]

In LITES, baseline equivalence was established if there were no statistically significant differences on specified variables (described below) for the analytic sample at baseline.[28] The

[25] See Monahan et al. (2015) for LITES standards on attrition in clustered RCTs.

[26] The attrition standards do not apply to MCGDs. These studies were evaluated on the basis of the final analysis sample, from which there is no attrition.

[27] It is important to establish baseline equivalence on key variables rather than merely adjusting for these variables by including them as covariates in a regression, because establishing baseline equivalence provides some assurance that intervention and comparison groups overlap enough with respect to these characteristics to enable a reasonable estimation of the program effect. If there is little overlap, the regression-based approach depends heavily on the model's functional form assumptions—that is, how accurately the model captures the true relationship between the covariates and the outcome. In this case, impact estimates rely heavily on extrapolation (Stuart, 2010), and such extrapolations can be highly sensitive to functional form (Foster, 2003).

[28] Variables upon which baseline equivalence must be established vary by evidence review, but typically include demographic information and pre-intervention outcomes. LITES only required studies to establish baseline equivalence on demographic characteristics and *not* child outcome measures. Child outcome measures were not required because, for infants and toddlers, these measures are not necessarily predictive of future outcomes, and the same measures are not always available for assessment at baseline and follow-up (for example, if a family enrolls in a study prenatally, there will not be child outcome baseline variables).

LITES review used author-reported baseline equivalence calculations, if available, and preferred two-tailed tests with $\alpha = 0.05$. When necessary, the LITES team calculated baseline equivalence and used a *p*-value from a chi-squared test for categorical variables (including dichotomous variables).

LITES required that baseline equivalence must be established on the following characteristics:

- Race/ethnicity
- Socioeconomic status (SES)
- Child age

Studies had to use statistical controls. In addition to establishing baseline equivalence, to receive a moderate rating, MCGDs and RCTs with high attrition were required to do at least one of the following:[29]

- Use some type of covariate adjustment when estimating impacts. To meet this requirement, a study could control for any or all of the required baseline characteristics (that is, race/ethnicity, SES, and/or age) or use different controls that could help reduce bias.[30]
- Demonstrate that results are not sensitive to the statistical controls selected. For example, a study could present a table of results from different models that included different sets of control variables, or state that impacts were estimated using models with different control variables but results were similar in sign, magnitude, and significance levels regardless of model.

5. Outcome- and study-level ratings

Outcomes within a study often receive different ratings. For example, some outcomes in an RCT might have low attrition and therefore receive a high rating. Other outcomes in the same RCT, however, might have high attrition, and therefore could receive only a moderate or low rating, depending on whether baseline equivalence was established and proper statistical controls were used.

Taking into account the possibility that outcomes within a study could receive different ratings, LITES reported study-level ratings as follows:

- **High:** The study had at least one high-rated outcome.
- **Moderate:** The study had at least one moderate-rated outcome and no high-rated outcomes.

[29] Although including statistical controls (such as pretests or sociodemographic characteristics) can improve the precision of impact estimates (Deke et al., 2010), the LITES review did not require statistical controls or covariate adjustment for RCTs with low attrition and no reassignment.

[30] Studies used a wide variety of control variables. If, for example, a study established baseline equivalence on all required variables but used other important variables as controls, we would not downgrade it. Endogenous covariates, or variables that are assessed after baseline and may have been influenced by the intervention, were not eligible to be used as control variables.

- **Low:** The study had no moderate- or high-rated outcomes.

C. Assessing evidence of effectiveness on child outcomes

In consultation with ASPE, ACF, and an expert work group, we adapted criteria for assessing evidence of effectiveness from the WWC, the HomVEE review, and the TPP Evidence Review. The LITES team customized these criteria for evaluations of out-of-home early learning programs for infants and toddlers. We examined eligible child outcomes from all high- and moderate-rated studies to determine the strength of the evidence of effectiveness for each program model. All child outcomes within the cognitive, social-emotional/behavioral, language development, and child health domains that met our criteria for a high or moderate rating were deemed eligible for providing credible evidence of program effects.[31]

We also recorded information on outcomes about long-term risk (for example, cigarette use at age 30) and economic well-being domains (for example, employment, annual income, or use of public assistance) and in interim outcome domains (for example, parent- or caregiver-child interaction), but these outcomes did not influence a program model's evidence of effectiveness rating. Table II.3 contains the primary, long-term risk and economic well-being, and interim domains reported in LITES.

[31] When a study followed participants from childhood through adolescence or adulthood, we continued to consider outcomes within these domains as eligible to provide evidence of effectiveness.

Table II.3. LITES outcome domains

Outcome domain	Examples
Primary child outcomes	
Cognitive development	Attention, memory, object permanence, concept development, categorization, understanding relationships (for example, cause and effect, part to whole), visual-motor integration, spatial reasoning, representational play, and problem solving
Social-emotional/behavioral development	Emotion regulation, impulse control, sociability, empathy, social problem solving, peer interaction, attachment, and adaptive behaviors (for example, self-help skills)
Language development	Receptive language, expressive language (including gestures), joint attention, and pre-literacy skills (for example, listening comprehension)
Child health[a]	Height, weight, cortisol levels, body mass index, parental ratings of general health, and fine and gross motor skills
Long-term risk and economic well-being outcomes	
Long-term risk behaviors	Substance abuse, dropping out of high school, and teen pregnancy
Long-term economic well-being	Employment and home-ownership in adulthood
Interim outcomes	
Global child care quality	Scores on the Infant/Toddler Environment Rating Scale (ITERS) or Family Child Care Environment Rating Scales (FCCERS)
Structural features of care	Child-to-staff ratios; group size; caregiver qualifications; professional development; the physical environment and furnishings; schedules; personal care routines; and health, safety, and nutrition practices
Parent- or caregiver-child interaction	Sensitivity/responsiveness, learning and language supports/instruction and cognitive stimulation, positive regard/warmth, behavior guidance, support for peer interaction, and areas of concern in interactions
Parent or caregiver knowledge of child development	Ability to identify developmental milestones
Global home environment	Home Observation for Measurement of the Environment (HOME) scores, language environment, cognitive stimulation, organization of the home, and safety

[a]Child health outcomes alone did not make a model eligible for inclusion in LITES, but child health outcomes were assessed for evidence of effectiveness.

1. Extracting and documenting data

We extracted basic information on primary child outcomes, long-term risk and economic well-being, and interim outcomes that were reported in a study. For outcomes rated high or moderate, we recorded the impact estimates reported by authors and whether the impacts were favorable, unfavorable, or neutral to the intervention.[32] We also recorded the statistical significance of the impact estimates and their effect sizes or the information necessary to calculate them, when the information was available.[33]

The review team documented all this information as the study reported it, including composite, scale-level scores and subscale scores of a measure, if reported separately. We based the evidence of effectiveness rating on subscales when they were the only measures available and on composite, scale-level measures when they were the only measures available. When both types were available, we based the evidence of effectiveness rating on subscales *and* composite measures, as long as the composite measure provided additional information beyond that contained in the subscales. If the composite measure overlapped entirely with the subscales, we reported the subscales only.

2. Assessing the evidence of effectiveness on child outcomes

Based on the information about eligible child outcomes, the review team assessed the extent of evidence for each program model.[34] We assigned a domain-specific evidence of effectiveness rating for each of the primary child outcome domains (cognitive, language, or social-emotional/behavioral development) and child health, if reported. We applied these ratings to end of intervention outcomes and to sustained or delayed outcomes—that is, outcomes measured one year or more after the end of the intervention.[35] We did not apply any multiple comparison corrections when assessing domain-specific evidence of effectiveness.[36] Table II.4 provides an overview of these ratings.

[32] An impact estimate with a positive sign is not necessarily favorable (for example, measures of problem behaviors).

[33] We recorded information on magnitudes and standard errors as presented by study authors. If authors did not report effect sizes, LITES attempted to compute them in a uniform manner (using Hedges' *g*, as in the WWC) when the necessary information was available (namely, intervention and comparison group outcome measure means, standard deviations, and sample sizes).

[34] We used categorizations similar to those for the evidence of effectiveness ratings developed by the WWC but tailored the terminology for the LITES literature. For example, the WWC refers to positive and negative effects; LITES uses "favorable" and "unfavorable."

[35] End of intervention outcomes included those measured at 36 months and/or those measured at the end of the intervention. These ratings could also apply to replicated outcomes—that is, outcomes measured in two or more non-overlapping study samples—but none of the reviewed program models had any replicated effects.

[36] Mathematica's experience conducting the HomVEE systematic review taught us that authors do not commonly provide all of the information necessary to make multiple comparison adjustments (namely, exact *p*-values). To avoid overburdening study authors with excessive author queries, we chose not to query them for this information. To provide some indication of whether a significant effect was due to chance, we report the number of significant effects as well as the number of null effects for each outcome domain.

Based on the domain-specific ratings, we assessed whether a program model exhibited evidence of effectiveness on child outcomes. If a program model exhibited "favorable" effects for end of intervention or sustained or delayed outcomes within any of the four primary child outcome domains, we deemed that model as exhibiting evidence of effectiveness.

LITES defined favorable and unfavorable effects as those that were statistically significant ($p \leq 0.05$) or that had an effect size greater than or equal to 0.2 standard deviations in absolute value. That is, results satisfying either of these two criteria counted toward an evidence of effectiveness rating. This decision was made because small studies would be less likely than large studies to demonstrate significant effects since smaller sample sizes are associated with larger *p*-values. Therefore, if statistical significance had been the only criterion for demonstrating an effect, there would have been a bias towards studies with larger sample sizes.

Table II.4. LITES evidence of effectiveness ratings

Domain rating	Outcome evaluation criteria
Favorable effects: evidence of a favorable effect with no overriding contrary evidence	At least one high- or moderate-rated study shows at least one significant or substantial favorable effect [a] AND No high- or moderate-rated study shows any significant or substantial unfavorable effects
Mixed effects: evidence of inconsistent effects	At least one high-or moderate rated study shows at least one significant or substantial favorable effect AND At least one high- or moderate-rated study shows at least one significant or substantial unfavorable effect
No discernible effects: no affirmative evidence of effects	No study shows any significant or substantial effects, either favorable or unfavorable
Unfavorable effects: evidence of an unfavorable effect with no overriding contrary evidence	At least one high- or moderate-rated study shows at least one significant or substantial unfavorable effect AND No high- or moderate-rated study shows any significant or substantial favorable effects

[a]A **significant effect** is statistically significant ($p \leq 0.05$). A **substantial effect** has an effect size greater than or equal to 0.2 standard deviations in absolute value.

3. Reporting of subgroup-specific outcomes

Child outcomes reported for a study's full sample contributed to a program model's overall evidence of effectiveness rating. We also rated a program model's subgroup-specific evidence of

effectiveness on outcomes in the domains of interest to LITES if they were reported separately (Appendix A). Subgroups of particular interest for this review were:[37]

- Children from low-income families
- Dual-language learners and/or immigrants
- Children from minority racial and ethnic groups
- Children with special needs
- Children in study-defined risk groups

In the next chapter, we summarize the short- and long-term outcomes for the program models included in the LITES systematic review.

[37] The full sample of a study might coincide with one of these subgroups of interest—for example, if a program model targets low-income families, a study's sample might consist entirely of low-income children. In this case, the results for the full sample would contribute to the program model's overall evidence of effectiveness rating, and to the program model's effectiveness rating for the low-income subgroup. If a study presented results for a broad sample—for example, children from low-, middle-, and high-income families—and for the subgroup of low-income children separately, the results reported for the full sample would contribute to the program model's overall evidence of effectiveness rating, and the results reported separately for the low-income subgroup would contribute to the program model's low-income subgroup effectiveness rating. Some subgroups of interest described above were not reported in any of the eligible studies. See Chapter IV for a discussion of the subgroups examined by studies of the models eligible for the LITES review.

III. SYSTEMATIC REVIEW FINDINGS

In this chapter, we present information on the studies included in the systematic review and findings from those studies. We first present the findings for the direct multicomponent models and then for the direct and indirect enhancement models.

A. Direct multicomponent models for infants and toddlers

For the LITES systematic review, direct multicomponent models needed to have a defined set of replicable program components and provide direct early learning services to infants and toddlers in out-of-home ECE settings. Nine direct multicomponent models had studies with causal designs eligible for review. Table III.1 lists these models, shows the number of studies that we identified as eligible for review, and summarizes the study quality ratings we assigned after examining these studies for each model. Additional details about the direct multicomponent models, studies, and findings appear in Appendix A.

1. Evidence base for direct multicomponent models

Most eligible studies for the nine direct multicomponent models were published before 2000. More recent studies typically report findings from longitudinal follow-ups of studies published earlier. For example, the four Abecedarian studies published after 2000 are all longitudinal follow-ups of a sample enrolled from 1972 to 1977. Only EHS and Sure Start have no studies published before 2000.

Of the studies reviewed, most rated low (N = 24) according to the LITES review criteria. Most of these studies did not establish baseline equivalence (N = 14), or they established baseline equivalence but did not use statistical controls (N = 10). Five models had studies that only rated low; therefore, we were unable to assess evidence of effectiveness for those models.

Four models had moderate- or high-quality research studies: Abecedarian, EHS, IHDP, and PCDC. Of the studies for these four models, 14 studies rated moderate, and 6 studies rated high.

2. Models with moderate- or high-quality research

There was some variation in the outcome domains and populations that each of the four models targeted (Table III.2). All focused on cognitive and social-emotional development and at least one other domain and were designed for at-risk children and families. All the models began before or soon after birth and continued until at least 36 months of age. IHDP and EHS also included home visiting in addition to out-of-home ECE services.[38]

[38] EHS programs that implemented only the home-based option were excluded from the LITES systematic review because evaluation findings on these programs have been reviewed under the HomVEE systematic review.

Table III.1. Extent of research on direct multicomponent models

	Number of studies eligible for review	Publication date of eligible studies		Quality ratings of eligible studies		
		1960–1999	2000 or later	High	Moderate	Low
Abecedarian	10	6	4	1	5	4
Brookline Early Education Project	4	3	1	0	0	4
Early Head Start (EHS)[b] (with a center-based component)	5	0	5	3[c]	1	1
Infant Health and Development Program (IHDP)	11	9	2	2[c]	7	2
Milwaukee Infant Stimulation Project	3	3	0	0	0	3
Parent-Child Development Center (PCDC)	6	5	1	0	1	5
Project CARE	2	2	0	0	0	2
Sure Start Local Programmes	2	0	2	0	0	2
Yale Child Welfare Research Program	1	1	0	0	0	1

Notes: Study inclusion criteria as described in Chapter I were as follows: a study was eligible for review if the model was (1) a *direct multicomponent model* that provided a defined set of replicable program components, including early learning services to infants and toddlers in out-of-home ECE settings; (2) a *direct enhancement model* with at least one replicable program component that provided early learning services to infants and toddlers in out-of-home ECE settings; or (3) an *indirect enhancement model* consisting of professional development programs with replicable program components focused on helping adult out-of-home caregivers support infant and toddler early learning. ECE services had to target infants and toddlers broadly. The program could serve children who were enrolled after age 36 months, as long as it primarily targeted infants and toddlers. The study had to include at least one outcome in any of the primary child outcome domains identified for this review (specifically, language, cognition, social-emotional/behavioral).The study also had to be in a policy-relevant nondeveloping-world context. The study design had to be an RCT, MCGD, SCD, or RDD. The study had to be a primary study published in English in 1960 or later. RCTs with low attrition and no reassignment received a high rating. RCTs with high attrition or reassignment and MCGDs received a moderate rating if they established baseline equivalence on required measures and used statistical controls. Studies that did not meet the requirements for a high or moderate rating received a low rating. Studies that had at least one high-rated outcome received a high rating. We further define study quality ratings in Chapter II.

This table summarizes information from the Extent of Research tables for individual program models in Appendix A.

[a]An additional model, Even Start, appears in Appendix C. It was ultimately deemed ineligible, as the Even Start studies identified did not meet the LITES age criteria.

[b]The EHS home-based approach was not included in LITES. The EHS program model summary in Appendix A presents findings separately for each EHS program approach eligible for LITES (mixed-approach and center-based).

[c]These studies contained both high- and moderate-rated outcomes.

Table III.2. Program model characteristics

Program model name	Child outcomes targeted by the models				Targeted additional outcomes	Sample characteristics	Ages of children served	Home visiting provided
	Cognitive development	Language development	Social-emotional/ behavioral development	Child health				
Abecedarian	Yes	Yes	Yes	No	Yes	Disadvantaged children[a]	3 to 60 months[b]	No
Early Head Start (EHS)[c] (with a center-based component)	Yes	Yes	Yes	Yes	Yes	Low-income families	Prenatal to 36 months	Yes
Infant Health and Development Program (IHDP)	Yes	No	Yes	Yes	No	Low birth weight premature infants	Discharge from hospital to 36 months	Yes
Parent-Child Development Centers (PCDC)	Yes[d]	Yes	Yes	No	No	Low-income families	3 to 36 months	No[e]

[a] Disadvantage was determined based on maternal intelligence, parent education, income, and other social factors.

[b] Abecedarian also included a school-age intervention for a subset of the intervention group. This school-age intervention was not the focus of the LITES review.

[c] The EHS home-based approach was not included in LITES. The EHS program model summary in Appendix A presents findings separately for each EHS program approach eligible for LITES (mixed-approach and center-based).

[d] PCDC targeted and measured outcomes in multiple domains, but only outcomes in the cognitive domain were rated moderate and are therefore reported below.

[e] The PCDC Birmingham site was center-based only.

3. Models with evidence of effectiveness on child outcomes

All four models with moderate- or high-quality research also demonstrated evidence of effectiveness on child outcomes (Table III.3).[39] Abecedarian, EHS, and IHDP each measured all the primary child outcome domains identified for LITES. Studies of all four models included longitudinal follow-ups, which provide the opportunity to observe sustained or delayed effects; however, all the follow-up studies for PCDC rated low, so potential sustained or delayed effects could not be examined for this model. The one moderate-rated study for PCDC measured effects on cognitive development, but only at the end of the intervention. A summary of each model and its effects is as follows:

Abecedarian was a program model designed for disadvantaged infants to improve cognitive, language, perceptual-motor, and social development. Children attended a year-round, full-day educational child care program starting at about 3 months of age and ending at about 5 years of age. The studies examined four cohorts of children, who were enrolled in infancy from 1972 to 1977, and were followed until they were 35 years old. Abecedarian demonstrated favorable effects in the cognitive domain at the end of the intervention,[40] and these effects were sustained in subsequent follow-ups. Abecedarian also measured and demonstrated favorable effects in the language, social-emotional/behavioral, and child health domains in follow-ups after the intervention ended. Additional details on the number of favorable effects in studies rating moderate or high are in Table III.4.

EHS is a program serving low-income families. EHS aims to improve child development outcomes, including health, social, cognitive, and language development, as well as to support family, staff, and community development. EHS offers a center-based approach, a home-based care approach, or a mixed approach in which programs offer a mix of both types of services. The focus of this review was on the center-based and mixed-approach programs.[41] The studies examined families who were enrolled at pregnancy or with a child younger than 12 months old between 1996 and 1998 and followed the families until the children were in 5th grade. There were favorable effects at the end the intervention and at later follow-ups in the cognitive domain. There were favorable effects at the end of the intervention for the language and social-emotional/behavioral domains. At later follow-ups, there were mixed effects in the social-emotional/behavioral domain with one substantial favorable effect and one substantial unfavorable effect during the 5th-grade follow-up.

[39] As described in Chapter II, if a model exhibited "favorable" effects measured at the end of the intervention or at some follow-up period ("sustained or delayed effects") within the cognitive, language, social emotional/behavioral, or child health domains, we deemed that model as exhibiting evidence of effectiveness.

[40] Abecedarian ended when the children were 5 years old. We report on outcomes measured between 3 years of age and 5 years of age and refer to that as "end of the intervention" for Abecedarian.

[41] The EHS home-based approach was out of scope for LITES but is included in the HomVEE review.

Table III.3. Evidence of effectiveness ratings by primary child outcome domain

Program model name	Cognitive development		Language development		Social-emotional/behavioral development		Child health	
	End of intervention effects	Sustained or delayed effects	End of intervention effects	Sustained or delayed effects	End of intervention effects	Sustained or delayed effects	End of intervention effects	Sustained or delayed effects
Abecedarian	Favorable	Favorable	Not applicable	Favorable	Not applicable	Favorable	Not applicable	Favorable
Early Head Start (EHS)[a] (with a center-based component)	Favorable	Favorable	Favorable	No discernible effects	Favorable	Mixed[b]	No discernible effects	Not applicable
Infant Health and Development Program (IHDP)	Favorable	No discernible effects	Favorable	No discernible effects	Favorable	No discernible effects	Unfavorable[c]	Mixed[d]
Parent-Child Development Centers (PCDC)	Favorable	Not applicable	Not applicable	Not applicable	Not applicable	Not applicable	Not applicable	Not applicable

Notes: Only models that had at least one study rated moderate or high appear in this table. **End of intervention effects** were measured at 36 months and/or at the end of the intervention, if the intervention extended beyond 36 months. **Sustained or delayed effects** were measured one year or more after the end of the intervention. Evidence of effectiveness ratings as described in Chapter II are as follows: favorable (at least one high- or moderate-rated study showed at least one significant or substantial favorable effect, and no high- or moderate-rated study showed any significant or substantial unfavorable effects); mixed (at least one high- or moderate-rated study showed at least one significant or substantial favorable effect, and at least one high- or moderate-rated study showed at least one significant or substantial unfavorable effect); no discernible effects (no high- or moderate-rated study showed any significant or substantial effects, either favorable or unfavorable); and unfavorable (at least one high- or moderate-rated study showed at least one significant or substantial unfavorable effect, and no high- or moderate-rated study showed any significant or substantial favorable effects). **Not applicable** indicates that no outcomes were measured that rated at least moderate in a study that met eligibility criteria.

This table summarizes information from the Domain-Specific Evidence of Effectiveness Ratings table for individual program models in Appendix A.

[a]The EHS home-based approach was not included in LITES. The EHS program model summary in Appendix A presents findings separately for each EHS program approach eligible for LITES (center-based and mixed-approach).

[b]EHS had a mix of one unfavorable effect and one favorable effect in a subsequent follow-up.

[c]This is morbidity during the first three years of life, which is the sum of reported injuries, surgeries, illnesses, and conditions.

[d]IHDP had a mix of one unfavorable effect and several favorable effects in subsequent follow-ups.

Table III.4. Number of significant or substantial favorable effects from studies rating moderate or high, by outcome domain

Program model name	Primary child outcomes				Long-term risk and economic well-being	Interim outcomes
	Cognitive development	Language development	Social-emotional/ behavioral development	Child health		
Abecedarian	17	2	1	1	11	Not applicable
Early Head Start (EHS)[a] (with a center-based component)	2	1	3	0	Not applicable	12
Infant Health and Development Program (IHDP)	3	3	2	4	Not applicable	2
Parent-Child Development Centers (PCDC)	2	Not applicable	Not applicable	Not applicable	Not applicable	Not applicable

Notes: The table includes counts of outcomes for the full samples collected at the end of the intervention and during follow-up periods. Only those models that had at least one study rating moderate or high appear in this table. A **significant or substantial effect** is one that is statistically significant ($p \leq 0.05$) or has an effect size greater than or equal to 0.2 standard deviations in absolute value. Interim and long-term risk and economic well-being outcomes are recorded only for studies that rated moderate or high based on primary child outcomes. **Long-term risk and economic well-being** domains include outcomes such as cigarette use at age 30 and annual income. **Interim outcomes** include the following domains: global child care quality, structural features of care, parent or caregiver-child interaction, parent or caregiver knowledge of child development, and global home environment. **Not applicable** indicates that no outcomes were measured that rated at least moderate in a study that met eligibility criteria.

This table summarizes information from the Summary Findings tables for individual program models in Appendix A.

[a]The EHS home-based approach was not included in LITES. The EHS program model summary in Appendix A presents findings separately for each EHS program approach eligible for LITES (mixed-approach and center-based).

IHDP focused on premature (gestational age of 37 or fewer weeks) newborns with low birth weights (less than 2,500 grams). The program aimed to support cognitive development, behavioral competence, and children's health. The program began when infants were discharged from the neonatal nursery and continued until children were 36 months old. The model had three main components: (1) home visits, (2) child development centers, and (3) parent groups.

Children in the program also received pediatric follow-up that included medical, developmental, and social assessments; referrals for pediatric care; and other services. The IHDP studies followed a cohort of newborns from a single multisite RCT that began in 1985 until the participants were 18 years old. For the overall sample, IHDP demonstrated favorable effects in the cognitive and language domains and favorable effects in the social-emotional/behavioral domain at the end of the intervention. These effects were not sustained in long-term follow-ups. At the end of the intervention, IHDP demonstrated an unfavorable effect on child health (specifically on the sum of reported child injuries, surgeries, illnesses, and conditions during the first three years of life). This effect may have been the result of increased access to services for the intervention group. Effects on child health were rated as mixed during long-term follow-ups, with one unfavorable effect and four favorable effects in subsequent follow-ups.

PCDC was a multisite ECE model for low-income families and children in Birmingham, Houston, and New Orleans. Each site had a slightly different approach and program components, but all sites had programming for mothers and children, and the program lasted until the children were 36 months old. Birmingham was the only site in the LITES review with moderate-rated outcomes,[42] and thus the only one for which we assessed evidence of effectiveness. The Birmingham site included one cohort of families that entered the program after 1972. PCDC demonstrated favorable effects in the cognitive domain at the end of the intervention for the Birmingham site.

Across models, almost all full-sample effects[43] on child outcomes were favorable or null (Table III.5). Statistical significance ($p \leq 0.05$) and substantial magnitude (effect size greater than or equal to 0.2 standard deviations in absolute value) were our criteria for assessing whether an effect was favorable or unfavorable versus being null. In particular, EHS had a large number of null effects, although EHS also measured a much larger number of outcomes than the other three models.

4. Favorable effects in the interim, long-term risk, and economic well-being domains

In addition to the effects already described, EHS had significant favorable effects on interim domains (Table III.4). Specifically, EHS had favorable effects in the parent or caregiver knowledge of child development, global home environment quality, and the parent- or caregiver-child interaction domain. IHDP had two favorable effects on parent-child interaction. Abecedarian demonstrated significant or substantial effects on long-term risk and economic well-being. The other models did not measure or did not have significant or substantial favorable

[42] As described in more detail in the next section, outcomes within a study often received different ratings. Outcomes in an RCT with high attrition, for example, could receive a moderate or low rating, depending on whether baseline equivalence was established and proper statistical controls were used.

[43] Only IHDP had subgroup analyses that were eligible for review and had outcomes that rated moderate or above. These subgroup effects are described in the IHDP program model summary in Appendix A.

effects on interim, long-term risk, and economic well-being in the primary analytic samples at the end of each intervention or during subsequent follow-ups.

Table III.5. Summary of findings on LITES program models from studies rating moderate or higher

	Favorable effects on child outcomes			Unfavorable effects on child outcomes			Null effects on child outcomes
Program model name	**During intervention[a]**	**End of intervention**	**Sustained or delayed**	**During intervention**	**End of intervention**	**Sustained or delayed**	
Abecedarian	4	3	18	2	0	0	8
Early Head Start (EHS)[b] (with a center-based component)	4	4	2	0	0	1[c]	107
Infant Health and Development Program (IHDP)	8	8	4	0	1[d]	1[e]	24
Parent-Child Development Centers (PCDC)	1	2	Not applicable	0	0	Not applicable	0

Notes: The table includes counts of outcomes in the cognitive, social-emotional/behavioral, language, and child health domains for the models' full, primary analytic samples during the intervention, at the end of the intervention, and during follow-up periods. Only those models that had at least one study rating moderate or higher appear in this table. A **significant or substantial effect** is one that is statistically significant ($p \leq 0.05$) or has an effect size greater than or equal to 0.2 standard deviations in absolute value. A **null effect** is an effect that is neither a significant or substantial favorable effect nor a significant or substantial unfavorable effect. **End of intervention effects** were measured at 36 months and/or at the end of the intervention, if the intervention extended beyond 36 months. A **sustained or delayed effect** is one that is statistically significant ($p \leq 0.05$) or has an effect size greater than or equal to 0.2 standard deviations and was measured one year or more after the end of the intervention. **Not applicable** indicates that no outcomes were measured that rated at least moderate in a study that met eligibility criteria.

This table summarizes information from the Summary Findings tables for individual program models in Appendix A.

[a]Effects measured during the intervention were not used to determine evidence of effectiveness.

[b]The EHS home-based approach was not included in LITES. The EHS program model summary in Appendix A presents findings separately for each EHS program approach eligible for LITES (mixed-approach and center-based).

[c]This is social problems in 5th grade.

[d]This is morbidity during the first three years of life, which is the sum of reported injuries, surgeries, illnesses, and conditions.

[e]This is a measure of the performance of physical activities such as playing sports, walking up stairs, bending, lifting, and caring for oneself at 96 months.

B. Direct and indirect enhancement models for infants and toddlers

In addition to the review of direct multicomponent models, the LITES systematic review included direct enhancement models that could be layered on another model and typically focused on improving child outcomes in a single domain, as well as indirect enhancement models that could be layered on another model and focused on improving caregiver practice. Two direct enhancement models and four indirect enhancement models had studies with eligible designs that also measured child outcomes (Table III.6). Similar models may measure interim outcomes, such as caregiver-child interaction. LITES only included models with studies that measured child outcomes. If a study measured child outcomes, we also reported interim outcomes. Additional details about the direct and indirect enhancement models, studies, and findings appear in Appendix B.

Table III.6. Extent of research on LITES direct and indirect enhancement models

Program model name[a]	Number of studies eligible for review	Publication date of eligible studies		Ratings of study quality of methodology for eligible studies		
		1980–1999	2000 or later	High	Moderate	Low
Direct enhancement models						
LearningGames	1	0	1	0	0	1
Music Education	1	0	1	0	0	1
Indirect enhancement models						
Eager and Able to Learn	1	0	1	0	0	1
Learning Language and Loving It	1	0	1	0	0	1
Project Secure Child in Child Care	1	0	1	0	0	1
Program for Infant/ Toddler Care (PITC)	1	0	1	1[b]	0	0

Notes: See study eligibility criteria in Chapter I. See definitions of ratings of study quality in Chapter II.

This table summarizes information from the Extent of Research table for each program model summary in Appendix B.

[a]Additional direct and indirect enhancement models, the Child Care Expulsion Prevention and the Responsive Early Childhood Curriculum, appear in Appendix C. These studies were deemed ineligible as they did not meet the LITES age criteria.

[b]This study contains both high- and moderate-rated outcomes.

1. Evidence base for direct and indirect enhancement models

The evidence base for the direct and indirect enhancement models with causal designs was small. Each model had only one study eligible for review. All six studies were published in 2000 or later, and five of the six studies rated low. Four rated low because they did not establish baseline equivalence, and one rated low due to a confounding factor. Therefore, we were unable to assess evidence of effectiveness for those five models. Both of the direct enhancement studies rated low.

PITC, an indirect enhancement model, had one eligible study. PITC combines direct caregiver training and on-site coaching or tailored assistance for center-based and family child care providers. PITC aims to indirectly improve children's language, cognitive, and social-emotional development by working with providers. PITC works with caregivers to implement six strategies for supporting infant and toddler early learning: (1) primary care (assignment of a primary caregiver to each child), (2) small groups, (3) continuity of care, (4) individualized schedules and routines, (5) inclusion of children with special needs, and (6) cultural sensitivity.

PITC's one eligible study received a high rating. PITC, however, did not meet the criteria for demonstrating evidence of effectiveness on child outcomes. Most of the effects (four of five) that were measured in the cognitive, language, and social emotional/behavioral domains were null.[44] Sustained or delayed effects were not measured. PITC received an unfavorable effect rating at the end of the intervention based on one significant unfavorable effect in the social-emotional/behavioral development domain. PITC also had one substantial favorable effect in an interim outcome domain, parent- or caregiver-child interaction. Counts of the individual outcomes from the PITC study that rated high are presented in Appendix B.

In the next chapter, we summarize lessons learned from the LITES systematic review, identify gaps in the research base, make recommendations for future research, and note policy implications.

[44] Child health was not measured.

This page has been left blank for double-sided copying.

IV. RESEARCH GAPS, LESSONS LEARNED, AND IMPLICATIONS

LITES aimed to conduct a systematic, thorough, and transparent review to identify replicable program models that have evidence of effectiveness for supporting infant and toddler early learning in out-of-home ECE settings. As the first systematic evidence review in this field, LITES sheds new light on this body of evidence and points to future research needs. In addition, the review team identified lessons learned about reviewing the evidence for these interventions. In this chapter, we describe gaps in the research on program models for supporting infant and toddler early learning in out-of-home ECE settings and lessons learned from the review process. We also offer suggestions for a future research agenda to continue building the evidence base, and highlight implications for policy.

A. Research gaps

1. Few replicable models for supporting infant and toddler early learning in out-of-home ECE settings have been evaluated using designs with strong internal validity[45]

Designs with strong internal validity use intervention and comparison groups that allow researchers to make causal inferences that the program model, rather than other differences between the two groups, *caused* the observed effect on children's outcomes. LITES review standards included four types of evaluation designs with potential for strong internal validity: (1) RCTs, (2) MCGDs, (3) SCDs, and (4) RDDs.

Based on a review of the literature published between 1960 and 2014, LITES identified 15 replicable program models for supporting infant and toddler early learning in out-of-home ECE settings that were evaluated using one of the four eligible designs. Of these models, five had at least one study that met LITES standards for study quality, and four demonstrated evidence of effectiveness on children's language, cognition, social-emotional/behavioral, and/or health outcomes (Abecedarian, EHS, IHDP, and PCDC).

2. Studies of effective models for supporting infant and toddler early learning in out-of-home ECE settings are dated, conducted between 1972 and the late 1990s, and do not shed light on several areas of current policy focus

The four program models with evidence of effectiveness on child outcomes were implemented between 1972 and the late 1990s. EHS, which enrolled children and families into its research sample in the late 1990s, is the most recently developed program model with evidence of effectiveness. Research on models implemented nearly two decades ago or earlier may not reflect the potential effectiveness of these models today. The social context and services available to comparison group members may have changed.

[45] LITES, like other systematic evidence reviews, emphasizes the importance of internal validity—a study's ability to produce unbiased estimates of program effects—in its review standards. While external validity—the extent to which a study's findings would replicate with a different sample—is also important, there are no well-established standards for assessing the strength of external validity in systematic reviews. Therefore, LITES did not assess external validity.

There are several potential explanations for the dearth in recent years of rigorous research on models to support infant and toddler early learning in out-of-home care. First, model development has not been a significant focus in the infant and toddler field. LITES' scan of the field for program models that are compelling but lack rigorous research yielded only 15 models at various stages of development (Del Grosso et al, 2015). Of those, only two are comprehensive, multicomponent models similar to those with evidence of effectiveness described in this report. One of those models, Educare, has a rigorous evaluation underway. The majority, however, focus on professional development support for caregivers rather than direct early learning services. Most research on those models measures interim outcomes such as environmental quality, caregiver-child interaction, and caregiver skills and credentials. Second, the field has invested in several system-building strategies, rather than model development, to improve the quality of out-of-home care provided to infants and toddlers. For example, the Quality Rating and Improvement System (QRIS) movement has swept the country in the last decade, with 38 states operating a standards-based system to improve quality and assist parents in selecting care (Boller & Maxwell, 2015). ACF has launched the Early Head Start-Child Care Partnership initiative as another system-building strategy to improve the quality of out-of-home infant and toddler early learning services. Finally, the field has engaged in substantial research and evaluation on models that support infant and toddler early learning through home visiting, not out-of-home, services (Avellar & Supplee, 2013).

3. Dated research literature does not reflect advances in measurement of infant and toddler outcomes

Since the 1990s, the field has made advances in understanding infant and toddler development and how to measure it. For example, executive functioning—which underlies self-regulation and involves higher order cognitive processes such as working memory, attentional flexibility, and inhibitory control—is a construct of growing interest because of its predictive power for later development (Carlson, 2005; Diamond et al., 2007). Several measures of executive function are available, such as the Minnesota Executive Function Scale for Early Childhood (Carlson, 2005: Carlson et al., 2015). Advances in understanding and measurement of language development have also occurred, especially in the area of dual language acquisition and strategies and measures for assessment of dual language learners. In addition, research has shown that higher adult word counts and higher numbers of conversational turns are associated with young children's language abilities (Gilkerson & Richards, 2009). Technological advances such as the LENA Pro software package facilitate recording of child and adult utterances using wearable devices for children as young as 2 months old. Because the research we identified as high quality was conducted decades ago, it does not reflect these advances in measurement.

4. None of the effective models for supporting infant and toddler early learning in out-of-home ECE settings has been rigorously evaluated through replication studies

Replication studies attempt to demonstrate a program effect in a second research sample that does not overlap with the sample used for the first evaluation. Researchers increasingly recognize replication as an essential component of scientific inquiry (Valentine et al., 2011). In addition, replication of findings in new service delivery contexts and samples with different characteristics increase confidence in the generalizability of the findings beyond the setting and population included in the initial study.

None of the program models with evidence of effectiveness (Abecedarian, EHS, IHDP, PCDC) assessed the effectiveness in more than one study sample. The Abecedarian study was based on a convenience sample in a single location, and the favorable finding for PCDC was based on a single site. The studies of EHS and IHDP used moderate to large samples in multiple sites, but results eligible for LITES were based on analyses that pooled the samples across sites.

As new models to support infant and toddler early learning in out-of-home care are developed and evaluated, replication studies should be conducted to verify the findings from the initial efficacy studies and to assess the external validity of the findings. For example, models should be rigorously evaluated with diverse samples of children and families that represent the range of families likely to enroll in such programs. Likewise, studies should be designed to test new models in a range of state policy contexts, such as states with different configurations of QRIS, child care licensing requirements, and child care subsidy policies.

5. Conducting MCGD studies is challenging; none of the MCGD studies reviewed by LITES met standards

MCGD studies use a nonrandom process for group assignment. This purposeful selection process can compromise the internal validity of the study. If the two groups differ at baseline, the comparison group does not provide a good indication of what would have happened to the intervention group in the absence of the program. Establishing baseline equivalence increases confidence that the two groups are balanced; however, it can never completely rule out differences in unmeasured characteristics. LITES required baseline equivalence on child race/ethnicity, baseline SES, and child age.

LITES would have assigned a moderate quality rating to any MCGDs that established baseline equivalence on selected characteristics. However, none of the MCGDs reviewed by LITES demonstrated baseline equivalence on required characteristics.[46] Studies failed to establish baseline equivalence for two main reasons: (1) authors did not measure the required characteristics at baseline; or (2) intervention and comparison groups were formed retrospectively, such that required characteristics could not be measured at baseline.

One reason that the studies we reviewed may have lacked measures of baseline equivalence is that the studies tended to be older and conducted before the present-day emphasis on reporting baseline information. Awareness of the importance of establishing baseline equivalence has ameliorated this deficiency in more recent studies. A factor in this change toward reporting baseline information may be the growth in the number of systematic reviews in the last several years. Most of these reviews have baseline equivalence standards reinforcing the importance of measuring key characteristics at baseline.

6. LITES did not identify any SCD or RDD studies of models for supporting infant and toddler early learning in out-of-home ECE settings

Well-executed SCDs establish a causal link between an intervention and outcomes. In an SCD, a case (individual or group) serves as its own control. This design differs from a pre-post

[46] One MCGD was rated low because of a confounding factor, and baseline equivalence was not assessed for this study.

approach because multiple measures of the outcome are taken before, during, and after the introduction of the intervention so that a performance trend can be established. In addition, a demonstration of an effect can be replicated in various ways, such as if an intervention is introduced at different times to different children.

An RDD is another design that can establish a strong causal link between a program model and outcomes. In an RDD, the sample is assigned to intervention and comparison conditions based on the value of a "scoring" variable. For example, a program might be offered to families with infants who were below a specific birth weight and not offered to families with children above the cutoff birth weight. The selection process is known and measurable. Thus, unlike an MCGD with comparison groups formed in some other way, the analysis can adjust for differences in selection to produce an unbiased estimate (Shadish et al., 2002).

LITES did not identify any SCD or RDD studies of models for supporting infant and toddler early learning in out-of-home ECE settings. The use of RDDs in social science has increased in recent years, including the use of RDDs to evaluate preschool programs. A RDD could be a viable design for contemporary infant and toddler programs when program assignment involves a scoring variable. SCDs tend to be most common in the clinical child psychology, school psychology, and special education literature (for example, SCDs are often used to investigate research questions that focus on low-incidence disorders or behaviors). It is likely that we did not encounter SCDs in LITES because we focused on interventions for infants and toddlers broadly. It is possible to use an SCD to investigate the effects of an intervention on children or classrooms, but QEDs and RCTs are more common in the ECE literature.

7. High attrition was a challenge for at least some studies of all models with RCTs reviewed by LITES

All four program models with evidence of effectiveness had studies that used RCT designs. Each of these RCTs experienced high attrition in at least one follow-up, but study authors were able to demonstrate baseline equivalence of the children remaining in the research samples.

8. Most of the studies reviewed by LITES did not correct for multiple comparisons

Most program models reviewed by LITES aimed to affect multiple outcomes in multiple domains; LITES assessed models' evidence of effectiveness based on child outcomes in the domains of cognitive, language, and social-emotional/behavioral development and child health. Some studies measured child outcomes in other domains, as well as interim outcomes. The number of outcomes tested was further increased through multiple longitudinal follow-ups to determine whether effects were sustained, or new effects manifested later. Most of the studies reviewed for LITES did not correct for multiple comparisons.

LITES selected an alpha level, or cutoff, of 0.05 for statistical significance, which means that there is a no more than 5 percentage point chance of finding a false positive for one outcome. A false positive occurs when a significant effect is detected, even though the groups are not truly different. The probability of a false positive increases, however, when multiple outcome tests are conducted. For example, if separate *t*-tests are conducted, each with an alpha level of 0.05, then when five outcomes are tested, the probability of a false positive is 23 percent; when 20 outcomes are tested, the probability rises to 64 percent (Schochet, 2009).

Corrections can be made to address the increased risk of a false positive. For example, the Bonferroni procedures divide the alpha level by the number of outcomes, so that the total probability across all outcomes is 5 percent. Less conservative corrections are also available, including the Benjamini-Hochberg method used by the WWC (Schochet, 2009; WWC, 2014).

9. Few studies reviewed by LITES included analyses of subgroups of interest

Subgroup analyses can highlight differential impacts of an intervention for children with different characteristics, help identify subsets of participants for whom the intervention worked, and indicate areas where model adaptation may be needed.

In consultation with ASPE and OPRE, the LITES review identified the following subgroups of interest:

- Children from low-income families
- Dual-language learners and immigrants
- Children from minority racial and ethnic groups
- Children with special needs[47]
- Children in author-defined risk groups

Of the models reviewed, only IHDP examined subgroups that met the LITES inclusion criteria and were rated at least moderate quality.[48] In one study, authors reported on impacts by family income (Liaw & Brooks-Gunn, 1994), and another study looked separately at the impacts of the intervention for children of white and black mothers bifurcated by education level (Brooks-Gunn et al., 1992). Seven subgroups met LITES criteria, with six demonstrating evidence of effectiveness on child outcomes (Appendix A).

In studies of other models, however, one or more characteristics of interest applied to the full study sample. Thus, the studies provide indications of the effects of these models for populations within ASPE's subgroups of interest. For example, EHS, PCDC, and Abecedarian specifically targeted low-income families. The sample for Abecedarian was also almost entirely African American.

[47] Criteria for this review focused on ECE services that broadly targeted infants and toddlers, but if children with special needs were a subgroup in a study of eligible services, that subgroup would be reported.

[48] Other models examined subgroups, but these did not meet the LITES inclusion criteria or were not rated at least moderate quality. For example, EHS examined subgroups of interest but subgroup effects were not reported separately by program approach (center-based, mixed-approach, and home-based; see Chazan-Cohen et al., 2013; Vogel et al., 2010; ACF, 2001, 2002). LITES only examined effectiveness of mixed and center-based approaches. In another example, Abecedarian examined effects separately for males and females (Campbell et al., 2014), but these were not pre-determined subgroups of interest for the LITES review

10. Impact estimates were difficult to compare across studies; most studies did not include the information needed to calculate effect sizes

An effect size shows the size of the impact (or the difference between the intervention and comparison group) relative to the standard deviation of the measure. Effect sizes can be useful for comparing or summarizing results from different studies (Hedges, 2008). A benefit of using the effect size is that it allows for comparisons of impacts across outcomes that may have been measured using different units. Unlike statistical significance, the effect size is not affected by the sample size but represents the size or strength of the relationship without regard to how precisely the size of that relationship was estimated (Hedges, 2008).

In studies reviewed by LITES, authors used a variety of analysis methods to calculate program impacts but did not always report impacts as effect sizes. Although formulas exist to convert impact estimates (for example, from a multivariate linear regression) to effect sizes, these formulas do not cover the range of estimation methods used by study authors. Aggregating study-specific impact estimates to produce an intervention-level impact estimate is difficult when studies use different estimation methods.

Even when effect sizes are available, there are questions about what constitutes a "large" effect. Commonly cited guidance is that 0.20 represents a "small" effect, 0.5 a "moderate" effect, and 0.8 a "large" effect (Cohen, 1988; Lipsey, 1990). LITES focused on multiple outcome domains. Implications for effect sizes, however, could differ across domains. It is unclear whether an effect size of a similar magnitude in a different outcome domain represents a similarly important effect.

B. Lessons learned from the LITES systematic review

Conducting a systematic review involves making decisions about program model and study eligibility criteria, as well as criteria for assessing study quality and evidence of effectiveness. The LITES team, ASPE, and OPRE consulted closely on many of these issues with a panel of experts on infant and toddler research and systematic review methodology. We learned several lessons from this process that are especially relevant to the literature on early learning models for infants and toddlers.

1. Including dated yet seminal studies in a systematic review of evidence has advantages and disadvantages

We considered the merits of including dated studies in the review, as well as whether to give priority to more recent studies, such as those published after 2000. Some experts felt that more recent studies better reflect the current, post-welfare reform context,[49] in which parents must manage challenging work schedules and transport their children to out-of-home programs. Another consideration was that outcome measures used by studies of infant and toddler program models have changed. In particular, dated studies do not measure some outcomes now considered to be important, such as joint attention or executive function. Several practical issues are also relevant. Dated studies are less likely to include information on baseline equivalence or

[49] Welfare reform refers to changes to the welfare system enacted by the Personal Responsibility and Work Opportunity Reconciliation Act of 1996.

to adjust for multiple comparisons. Locating study authors to request missing information to include in the systematic review is also challenging, and the missing information may no longer be accessible.

Other experts felt that effective approaches to supporting infant and toddler early learning may be less affected by contextual changes compared with programs for older children. Moreover, a strong advantage of including more dated studies in the review is the opportunity to include longitudinal findings, allowing the review to examine whether program effects are sustained, as well as to detect effects that may appear only after an intervention has ended. For example, IHDP conducted follow-up assessments of children at 60 months, 86 months, 96 months, and 18 years. The longitudinal follow-up showed that positive effects on health outcomes emerged only after the end of the intervention.

2. There are trade-offs in using statistical significance or effect sizes to determine whether a program model has evidence of effectiveness

LITES considered the trade-offs to using statistical significance or effect sizes to determine whether a program model has evidence of effectiveness. Some experts suggested that effects from studies with large samples might be statistically significant but not large enough to be meaningful. In other situations, effects might not appear to be replicated in a second study because the study is underpowered to detect the statistical significance of the effect. Others noted that authors may use different methods for calculating effect sizes and that effect sizes considered to be meaningful differ across outcome domains. Moreover, for some domains, there is no consensus about what constitutes a meaningful effect size. At the recommendations of the experts, LITES evidence of effectiveness was based on statistical significance and/or crossing an effect size threshold. Ultimately, however, LITES relied most often on statistical significance, because for many studies, effect size information was not presented and the team was unable to obtain information necessary for calculating effect sizes.

3. Defining the term "program model" for interventions that support infant and toddler early learning in out-of-home ECE settings is challenging

Defining the scope of the review, including the characteristics of program models eligible for inclusion, was an important foundational step in developing the LITES review protocol. ASPE and ACF focused LITES on out-of-home early learning services for infants and toddlers, in part because home visiting services are currently reviewed by ACF/OPRE's HomVEE systematic review. Initially, the review scope labelled the eligible early learning service models as "comprehensive" but that term led to questions. For example, does comprehensive signify an intervention that targets more than one domain or all child domains? Does comprehensive imply a multigenerational focus? Is it the program or its effect that needs to be comprehensive? To clarify, the LITES team instead defined eligible models as varying along two dimensions. Models were (1) either multicomponent or enhancement models and (2) either supported early learning directly through work with children or indirectly through work with caregivers. This resulted in the three categories of models summarized in this report: direct multicomponent models, direct enhancement models, and indirect enhancement models.

Experts advising the LITES team emphasized that the boundaries between "models," "approaches," and "practices" were open to interpretation. A goal of the review was to identify

effective program models that could be replicated and scaled up. LITES sought program models that, at a minimum, provided a defined package of infant and toddler early learning service components or professional development services to help caregivers support infant and toddler early learning. Out-of-home early learning services for infants and toddlers often draw on combinations of more than one theoretical approach, such as Reggio Emilia,[50] and practices, such as continuity of care and primary caregiving. These approaches and practices are implemented in a range of configurations and intensities across settings, making them difficult to replicate consistently without further specification. Therefore, LITES did not review the research evidence on these approaches and practices, even though they are prevalent in the field.

C. Suggestions for future research

LITES identified a number of strengths and gaps in the research base for program models that support infant and toddler early learning in out-of-home ECE settings. Strengths included several program models with evidence of effectiveness based on well-implemented RCTs. Three of these models (Abecedarian, EHS, and IHDP) have studies showing long-term favorable effects beyond the end of the intervention. Abecedarian showed long-term favorable effects in multiple domains: cognitive, language, and social-emotional/ behavioral development; child health; long-term economic well-being; and long-term risk behaviors such as cigarette smoking. EHS showed long-term favorable effects in cognitive development and social-emotional/behavioral development. IHDP showed mixed long-term effects in child health. The research base is limited, however, with a number of gaps as described earlier in the chapter. Further research can build on the foundation established by these seminal studies. To address identified gaps and strengthen the research base, we offer the following recommendations.

1. Conduct research using designs with strong internal validity

Very few studies of eligible program models used research designs with strong internal validity, and, of these, fewer executed the designs well. These designs can include RCTs, MCGDs, RDDs, and SCDs. LITES found that, although several RCTs have been conducted, the infant and toddler field has not yet adopted the use of RDDs and SCDs. These designs, however, may be less expensive and more feasible to implement and should be considered for future research. In particular, they do not require that eligible children and families be assigned to a no-services comparison group.

2. Conduct replication studies to confirm efficacy findings and assess external validity

When a program model demonstrates evidence of effectiveness, replication of findings in a second study increases confidence in the program model's effectiveness. Replication shows that the findings are not limited to a specific population or location, or the result of close involvement of the program developer during the evaluation period. As new models are developed and evaluated, replication studies should be conducted to verify the findings from the initial efficacy studies and to assess their effectiveness with diverse samples of children and families and in a

[50] Reggio Emilia is an educational philosophy based on the idea that children possess strong potential for development and they learn and grow in relationships with others (Hewitt, 2001). It was developed by Loris Malaguzzi from Reggio Emilia, Italy. The program is based on the principles of respect, responsibility, and community through exploration and discovery in a supportive and enriching environment based on the interests of the children through a self-guided curriculum (Edwards, 2002).

range of state policy and service delivery contexts. Ideally, replication studies should be conducted by a research team that is independent of the original team. To the extent feasible, replication studies should use the same outcome measures as the original study to facilitate comparison of results.

3. Collect and report data on baseline characteristics for the analytic sample used to assess program effects

Study authors should collect and present evidence showing that intervention and comparison groups are similar on key baseline characteristics at each follow-up analysis. These key characteristics should be expected to relate to participant outcomes. Even though authors may report on sample members' characteristics at baseline, they should demonstrate that the groups remain similar at follow-up, accounting for any attrition from either group.

4. Incorporate strategies to reduce the risk of finding statistically significant findings by chance due to multiple comparisons

Studies reviewed by LITES measured multiple outcomes in multiple domains, which can lead to finding statistically significant findings by chance. Steps should be taken to reduce this risk in future research. As noted earlier in the chapter, statistical corrections for multiple comparisons for outcomes within a domain can be made using the Bonferroni or other approaches.

5. Design studies to incorporate assessment of effects on diverse subgroups of infants and toddlers

Few of the studies reviewed by LITES included assessment of effects on subgroups of infants and toddlers.[51] Studies with diverse samples of children and families that include assessment of subgroup effects will enhance the generalizability of the findings. Similarly, studies of models implemented in different community contexts can also enhance generalizability. Evaluators should plan for subgroup analyses during the study design stage, including determining the sample sizes needed to detect statistically significant effects and potentially oversampling subgroups of interest.

6. Report effect sizes

Effect sizes facilitate comparisons of results across outcomes and studies, regardless of estimation method. Study authors should, whenever possible, attempt to present impact estimates in terms of effect sizes, in addition to presenting results from their chosen estimation methods. Interpreting the meaning of effect sizes (their clinical or policy relevance) can be challenging. The size of a meaningful effect varies across outcomes domains, and expert consensus about what constitutes a meaningful effect size is lacking. The infant and toddler field should consider developing effect size benchmarks or thresholds for determining the importance of an effect in different outcome domains.

[51] Several models examined subgroups, but only IHDP had subgroups of interest that met the LITES inclusion criteria and were rated at least moderate quality.

7. Incorporate assessment of implementation fidelity into studies of program effectiveness

Experts advising the LITES team recommended assessing information about implementation when feasible, but acknowledged that this information is not typically included in published articles about program effectiveness. Some experts noted that poor implementation of a well-designed program could cause the program to appear less effective than a well-implemented program with a weaker design. Thus, studies that do not measure fidelity of implementation run the risk of drawing incorrect conclusions about the effectiveness of a program model if it was not implemented as specified by the developer. Especially in initial efficacy studies, examination of fidelity can help determine the feasibility of implementing the program model as specified by the developer and contribute to the development of fidelity standards and measures.

8. Conduct planned variation studies to identify which components of program models contribute to program effects

Planned variation studies can identify which program components (or combinations of components) are most effective by testing the relative impact of different combinations (such as different combinations of classroom interventions or teacher supports). These types of studies can also assess the relative effectiveness of different levels of dosage of an intervention. Moreover, planned variation studies can help tease apart which components seem most related to sustained, long-term effects, and whether the long-term effectiveness of particular components varies by outcome domain.

D. Implications for policy

Despite the documented importance of early experiences, less is known about effective program models that support infant and toddler early learning in out-of-home ECE settings. LITES was intended to begin filling this gap by identifying program models with evidence of effectiveness on child outcomes. This project joins other ongoing federal efforts in promoting awareness and use of evidence-based and high quality practices to help children realize their full potential. The federal government's strong interest in this issue was highlighted during the White House Summit on Early Childhood Education in December 2014, which drew national attention to the importance of ECE. There has also been growing interest in the policy sector in establishing a continuum of high quality early learning opportunities for children from birth through age 5, including a focus on growing the supply of high-quality ECE opportunities for infants and toddlers through initiatives such as the Early Head Start-Child Care Partnership grants program.

LITES included findings from several seminal ECE studies that showed that out-of-home early learning programs for infants and toddlers can have positive effects on children's development that for some models persist well beyond the end of the intervention, even into adulthood. All four models with evidence of effectiveness had favorable end-of-intervention effects in cognitive development; EHS and IHDP also had favorable effects in language and social-emotional/behavioral development at the end of the intervention. Moreover, Abecedarian had long-term favorable effects in all child outcomes domains as well as long-term economic well-being and risk behaviors. EHS had long-term favorable effects in cognitive development and long-term mixed effects in social-emotional/behavioral development; IHDP had mixed long-term effects in child health.

These program models with evidence of effectiveness on child outcomes shared several characteristics. All were direct multicomponent models that began before or soon after birth and continued until at least 36 months of age. All the models focused on supporting children's cognitive and social-emotional development and at least one other developmental domain. Most of the models had a focus on parents in addition to children, and all provided individualized early learning, social, and health services tailored to child and family circumstances and needs. All were designed for at-risk children and families. All had an out-of-home ECE service component, but the models varied in terms of their dosage, and two of the four also included home visiting.

Overall, LITES identified very few eligible studies of enhancement models that examined child outcomes, and only one of these models had a study with strong internal validity. This study of a professional development model demonstrated a favorable effect on caregiver-child interaction but did not demonstrate evidence of effectiveness on child outcomes. Several additional enhancement models that measure only outcomes classified as "interim" by LITES are featured in the LITES report on compelling models (Del Grosso et al., 2015).

Given the small number of out-of-home ECE models that have studies with strong internal validity and the lack of replication studies for these models, more research is needed to fully inform the field about what works for infants and toddlers, especially considering current policy interest in expanding and improving services for this population. Only15 models were identified as eligible for LITES, and only 5 had studies with strong internal validity and were therefore eligible for assessing evidence of effectiveness on children's outcomes. All of these studies were RCTs; none used regression discontinuity or single case designs. Most of the studies were conducted nearly two decades ago.

This points to the need for policymakers to support the development and specification of additional ECE models for infants and toddlers,[52] as well as supporting continued research and refinement of existing models that demonstrate evidence of effectiveness. Several innovative strategies exist for supporting model development and testing. In recent years, ACF and a number of its partner agencies have launched research networks that bring groups of researchers together around different topics. The Network for Infant/Toddler Researchers (NITR), sponsored by OPRE, could serve as a forum for supporting development of ECE models for infants and toddlers. NITR brings together federal staff and researchers with expertise in developmental science, implementation, professional development, and data use to identify existing research for informing policy; identify research gaps; and build capacity to conduct research that can inform infant and toddler programs. For example, the NITR Program Practices Workgroup is collaborating with the Quality Initiatives Research and Evaluation Consortium, another ACF-sponsored working group, to examine how state QRIS systems can include assessments of curricula for infants and toddlers.

Collaborative innovation and improvement networks (CoIINs) aim to advance breakthrough improvements in specific programs and topics of concern by supporting learning communities of practitioners, researchers, and experts. These networks aim to develop innovative practices and to improve outcomes, informed by practitioner knowledge and current research. The networks

[52] See Del Grosso et al., 2015 for a description of compelling models and their current level of specification and research.

engage in Plan-Do-Study-Act cycles to design innovations, test them in practice settings, reflect on the results, and refine them as needed. For example, the HRSA-sponsored Home Visiting Collaborative Improvement and Innovation Network brings together teams of researchers and implementing agencies to develop and test innovations in areas of home visiting practice for which evidence-based practice does not yet exist, such as strategies to increase duration of breastfeeding and early detection of developmental delays.

In the private sector, early learning labs aim to accelerate experimentation and development of scalable early learning interventions by bringing together experts in early childhood, design, and innovation. For example, the Oakland-based Early Learning Lab supports testing of new innovations in several California communities funded by the Packard Foundation-sponsored Starting Smart and Strong Initiative. With support from the Early Learning Lab, these communities will engage in rapid cycle testing of innovations in formal and informal early care and education settings to accelerate the development of scalable interventions that positively impact children's school readiness. All of these strategies involve bringing together groups of experts to collaborate, innovate, and experiment and offer opportunities for model and practice development and evaluation. Overall, there is still much to learn about what works for infants and toddlers, and the field will need to continue to build the knowledge base to identify a broader range of effective programs that can help young children reach their full potential.

REFERENCES

Avellar, S. A., & Supplee, L. H. (2013). Effectiveness of home visiting in improving child health and reducing child maltreatment. *Pediatrics, 132*(Supplement 2), S90–S99.

Boller, K., & Maxwell, K. (2015). QRIS research: Looking back and looking forward. *Early Childhood Research Quarterly, 30*, 339–342.

Bradley, R. H., & Corwyn, R. F. (2002). Socioeconomic status and child development. *Annual review of psychology, 53*(1), 371–399.

Brooks-Gunn, J., Gross, R. T., Kraemer, H. C., Spiker, D., & Shapiro, S. (1992). Enhancing the cognitive outcomes of low birth weight, premature infants: For whom is the intervention most effective? *Pediatrics, 89*(6), 1209–1215.

Campbell, F. A., Gunn, E., & Pungello, E. P. (2014). Carolina Abecedarian project and the Carolina approach to responsive education (CARE), age 21 follow up study. *Inter-university Consortium for Political and Social Research.*

Camilli, G., Vargas, S., Ryan, S., & Barnett, W. S. (2010). Meta-analysis of the effects of early education interventions on cognitive and social development. *Teachers College Record, 112*, 579–620.

Carlson, S. M. (2005). Developmentally sensitive measures of executive function in preschool children. *Developmental Neuropsychology, 28*(2), 595–616.

Carlson, S. M., Claxton, L. J., & Moses, L. J. (2015). The relation between executive function and theory of mind is more than skin deep. *Journal of Cognition and Development, 16*(1), 186–197.

Chazan-Cohen, R., Raikes, H. H., & Vogel, C. (2013). Program subgroups: Patterns of impacts for home-based, center-based, and mixed-approach programs. In J. M. Love, R. Chazan-Cohen, H. Raikes, & J. Brooks-Gunn (Eds.), *What Makes a Difference: Early Head Start Evaluation Findings in a Developmental Context*. Boston: Wiley.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Lawrence Erlbaum.

Deke, J., Dragoset, L., & Moore, R. (2010). *Precision gains from publicly available school proficiency measures compared to study-collected test scores in education cluster-randomized trials.* (NCEE 2010-4003). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance.

Del Grosso, P., Jones, C., Paulsell, D., & Monahan. S. (2015). *Learning About Infant and Toddler Early Education Services (LITES): Summarizing the research and gaps on compelling models.* Washington, DC: U.S. Department of Health and Human Services, Office of the Assistant Secretary for Planning and Evaluation.

Diamond, A., Barnett, W. S., Thomas, J., & Munro, S. (2007). Preschool program improves cognitive control. *Science, 318*(5855), 1387.

Duncan, G., & Magnuson, K. (2013). Investing in preschool programs. *Journal of Economic Perspectives, 27*, 109–132.

Edwards, C.P. (2002). "Three Approaches from Europe: Waldorf, Montessori and Reggio Emilia." *Early Childhood Research and Practice*, *4*(1). Available at http://ecrp.uiuc.edu/v4n1/edwards.html. Accessed November 24, 2014.

Fixsen, D., Blase, K., Metz, A., & Van Dyke, M. (2013). Statewide implementation of evidence-based programs. *Exceptional Children*, *79*(2), 213–230.

Foster, E. M. (2003). Propensity score matching: An Illustrative analysis of dose response. *Medical Care, 41*(1), 1183-1192.

Gilkerson, J. & Richards, J. A. (2009). The power of talk: Impact of adult talk, conversational turns, and TV during the critical 0–4 years of child development. *LENA Research Foundation.*

Halle, T., Forry, N., Hair, E., Perper, K., Wandner, L., Wessel, J., & Vick, J. (2009). *Disparities in early learning and development: Lessons from the Early Childhood Longitudinal Study – Birth Cohort (ECLS-B): Executive Summary*. Retrieved from http://www.childtrends.org/wp-content/uploads/2013/05/2009-52DisparitiesELExecSumm.pdf.

Hedges, L. V. (2008). What are effect sizes and why do we need them? *Child Development Perspectives, 2*(3), 167–177.

Hewitt, V. (2001). "Examining the Reggio Emilia Approach to Early Childhood Education". *Early Childhood Education Journal*, *29*(2), 95–100.

Home Visiting Evidence of Effectiveness. (2014). *HomVEE review protocol.* Retrieved from http://homvee.acf.hhs.gov/document.aspx?rid=4&sid=19&mid=5.

Karoly, L. A., Kilburn, M. R., & Cannon, J. S. (2005). *Early childhood interventions: Proven results, future promise.* Report to the PNC Financial Services Group, Inc. Santa Monica, CA: Rand Corporation. Available at http://www.rand.org/pubs/monographs/2005/RAND_MG341.pdf.

Leak, J., Duncan, G. J., Li, W., Magnuson, K., Schindler, H., & Yoshikawa, H. (2010). *Is timing everything? How early childhood education program impacts vary by starting age, program duration and time since the end of the program.* Paper presented at the Biennial Meeting for the Society for Research on Child Development, Montreal, Quebec, Canada, March 31–April 2, 2011.

Liaw, F., & Brooks-Gunn, J. (1994). Cumulative familial risks and low-birthweight children's cognitive and behavioral development. *Journal of Clinical Child Psychology, 23*(4), 360–372.

Lipsey, M. W. (1990). Design sensitivity: statistical power for experimental research. Newbury Park, CA: Sage.

McCormick, M. C., McCarton, C., Tonascia, J., & Brooks-Gunn, J. (1993). Early educational intervention for very low birth weight infants: Results from the Infant Health and Development Program. *Journal of Pediatrics, 123*(4), 527–533.

Meisels, S. J., & Shonkoff, J. P. (Eds.). (1990). *Handbook of early childhood intervention*. New York: Cambridge University Press.

Monahan, S., Thomas, J., Murphy, L., & Paulsell, D. (2015). *LITES review protocol.* Washington, DC: U.S. Department of Health and Human Services, Office of the Assistant Secretary for Planning and Evaluation.

National Scientific Council on the Developing Child. (2007). The timing and quality of early experiences shape brain architecture. Working paper no. 5. Cambridge, MA: Harvard University. Retrieved from http://developingchild.net/pubs/wp-abstracts/wp5.html.

Schochet, P. Z. (2009). An approach for addressing the multiple testing problem in social policy impact evaluations. *Evaluation Review, 33*, 539–567.

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

Shonkoff, J. P., & Meisels, S. J. (Eds.). (2000). *Handbook of early childhood intervention (second edition)*. New York: Cambridge University Press.

Stuart, E. A. (2010). Matching methods for causal inference: A review and a look forward. *Statistical Science, 25*(1), 1-21.

Teen Pregnancy Prevention Evidence Review. (2012). *Identifying programs that impact teen pregnancy, sexually transmitted infections, and associated sexual risk behaviors review protocol, version 2.0.* Retrieved from http://www.hhs.gov/ash/oah/oah-initiatives/teen_pregnancy/db/eb-programs-review-v2.pdf.

U.S. Administration for Children and Families. (2001). *Building their futures: How Early Head Start programs are enhancing the lives of infants and toddlers in low-income families.* Washington, DC: U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research and Evaluation

U.S. Administration for Children and Families. (2002). *Making a difference in the lives of infants and toddlers and their families: The impacts of Early Head Start: Vol. I. Final technical rep*ort. Washington, DC: U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research and Evaluation.

Valentine, J. C., Biglan, A., Boruch, R. F., González Castro, F., Collins, L. M., Flay, B. R., Kellam, S., Moscicki, E., & Schinke, S. P. (2011). Replication in prevention science. *Prevention Science, 12*, 103–117.

Vogel, C. A., Xue Y., Moiduddin, E. M., Carlson, B. L., & Kisker E. E. (2010). *Early Head Start children in grade 5: Long-term follow-up of the Early Head Start Research and Evaluation Study sample. OPRE Report #2011-8.* Washington, DC: U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research and Evaluation

What Works Clearinghouse. (2011). *What Works Clearinghouse procedures and standards handbook version 2.1*. Retrieved from http://ies.ed.gov/ncee/wwc/pdf/reference_resources/wwc_procedures_v2_1_standards_handbook.pdf.

What Works Clearinghouse. (2014). *What Works Clearinghouse procedures and standards handbook version 3.0*. Retrieved from http://ies.ed.gov/ncee/wwc/pdf/reference_resources/wwc_procedures_v3_0_standards_handbook.pdf.

Yoshikawa, H., Weiland, C., Brooks-Gunn, J., Burchinal, M. R., Espinosa, L. M., Gormley, Ludwig, J., Magnuson, K., Phillips, D., & Zaslow, M. J. (2013). Investing in our future: The evidence base on preschool education. New York: Foundation for Child Development.

APPENDIX A

DIRECT MULTICOMPONENT PROGRAM MODEL SUMMARIES

This page has been left blank for double-sided copying.

ABECEDARIAN

Description

The **Abecedarian Project** was an intervention targeted at disadvantaged infants, determined to be disadvantaged based on an index score of maternal intelligence, family income, parents' education, intactness of family, and nine other social factors. The program focused on cognitive, language, perceptual-motor, and social development.

Children attended a year-round, full-day educational child care program starting at about age 3 months and ending at about age 5 years.[53] Although the child care program was the main component of the model, Abecedarian also provided families with pediatric care and family support as requested.

The child care program was open five days a week, 50 weeks a year, and children attended the program from 6 to 10 hours a day. Children up to the age of 15 months were a part of the infant program, which had a teacher-to-infant ratio of 1:3. Children ages 15 months to 3 years were in the toddler program, which had a teacher-to-toddler ratio of 1:4 or 1:5. The children ages 3 to 5 years were in the preschool program, which had a teacher-to-preschooler ratio of 1:6. There were two teachers per class, with the lead teacher having training in early childhood education and previous teaching experience. The teacher's aides were working toward a Child Development Associates certificate. Infants and toddlers received a specially developed curriculum, LearningGames, consisting of educational activities that targeted cognitive, language, motor, and social development (see Appendix B).

We identified 10 studies of the Abecedarian Project as eligible for review (Ramey & Haskins, 1981; Campbell et al., 1998; Campbell et al., 2012; Campbell & Ramey, 1994, 1995; Campbell et al., 2002; Muennig et al., 2011; Campbell et al., 2001; Horacek et al., 1987; Ramey & Campbell, 1984). The studies examined four cohorts of children who were enrolled in infancy from 1972 to 1977 and were followed until they were 35 years old.[54] Children were randomly assigned to either the Abecedarian Project (N = 57) or to the comparison group (N = 54). Children in the comparison group received pediatric care and dietary supplements (iron-fortified infant formula), and their families received family-support social services by request.

This program model summary reports only full sample results. No eligible subgroups were reported in studies of this program model.[55]

[53] The Abecedarian Project had a two-stage intervention: the first was the early childhood stage (ages birth to 5 years), and the second was a school-age stage (ages 6 to 8 years). At kindergarten entry, children from the intervention and comparison groups were randomized to a school-age intervention or to a school-age comparison group. This review focused only on the initial random assignment to the early childhood portion of the intervention.

[54] One study in the review included only the first two cohorts of children.

[55] One study presented results separately by gender, which is not a subgroup of interest for this review.

Domain-specific evidence of effectiveness ratings

Table A.1. Abecedarian, full sample

Outcome domain	End of intervention effects	Sustained or delayed effects
Cognitive development	Favorable effects	Favorable effects
Language development	Not applicable	Favorable effects
Social-emotional/ behavioral development	Not applicable	Favorable effects
Child health	Not applicable	Favorable effects

Notes: Possible ratings are as follows: **favorable effects** (at least one high- or moderate-rated study shows at least one significant or substantial favorable effect, no high-rated study shows any significant or substantial favorable effects, and no high- or moderate-rated study shows any significant or substantial unfavorable effects); **mixed effects** (at least one high-or moderate rated study shows at least one significant or substantial favorable effect and at least one high- or moderate-rated study shows at least one significant or substantial unfavorable effect); **no discernible effects** (no high- or moderate-rated study shows any significant or substantial effects, either favorable or unfavorable); and **unfavorable effects** (at least one high-or moderate-rated study shows at least one significant or substantial unfavorable effect, no high-rated study shows any significant or substantial unfavorable effects, and no high- or moderate-rated study shows any significant or substantial favorable effects). **End of intervention effects** are measured at 36 months and/or at the end of the intervention, if the intervention extends beyond 36 months. **Sustained or delayed effects** are measured one year or more after the end of the intervention. **Not applicable** indicates that no outcomes were measured that rated at least moderate in a study that met eligibility criteria.

Information from this table is summarized in the main report in Table III.3.

Summary of findings

Table A.2. Abecedarian, full sample, primary domains and child health

Outcome domain	Favorable effects			Unfavorable effects			Null effects[a]
	During intervention	End of intervention	Sustained or delayed	During intervention	End of intervention	Sustained or delayed	
Cognitive development	3	3	14	0	0	0	2
Language development	Not applicable	Not applicable	2	Not applicable	Not applicable	0	1
Social-emotional/ behavioral development	Not applicable	Not applicable	1	Not applicable	Not applicable	0	1
Child health	1	Not applicable	1	2	Not applicable	0	4

Notes: A **significant or substantial effect** is one that is statistically significant ($p \leq 0.05$) or has an effect size greater than or equal to 0.2 standard deviations in absolute value. A **null effect** is an effect that is neither a significant or substantial favorable effect nor a significant or substantial unfavorable effect. **End of intervention effects** are measured at 36 months and/or at the end of the intervention, if the intervention extends beyond 36 months. **Sustained or delayed effects** are measured one year or more after the end of the intervention. **Not applicable** indicates that no outcomes were measured that rated at least moderate in a study that met eligibility criteria.

Information from this table is summarized in the main report in Tables III.4 and III.5.

[a]Includes effects measured during and immediately after the intervention, as well as in longer-term follow-ups.

Table A.3. Abecedarian, full sample, children's long-term risk and economic well-being domains

Outcome domain	Significant or substantial favorable effects	Significant or substantial unfavorable effects	Null effects
Long-term economic well-being	8	0	1
Long-term risk behaviors	3	0	12

Notes: A **significant or substantial effect** is one that is statistically significant ($p \leq 0.05$) or has an effect size greater than or equal to 0.2 standard deviations in absolute value. A **null effect** is an effect that is neither a significant or substantial favorable effect nor a significant or substantial unfavorable effect. This table includes effects measured during and immediately after the intervention, as well as in longer-term follow-ups. The **long-term economic well-being domain** includes outcomes such as annual income, job prestige, and use of public assistance. The **long-term risk behaviors domain** includes outcomes such cigarette and marijuana use and binge drinking at age 30.

Table A.4. Abecedarian, extent of research

	Number of studies
Eligible for review	10
Publication date of eligible studies	
1960 to 1979	0
1980 to 1999	6
2000 or later	4
Rating of eligible studies	
Rated high	1[a]
Rated moderate	5
Rated low	4

Notes: Study inclusion criteria as described in Chapter I of the main report are as follows: a study was eligible for review if the model was (1) a direct multicomponent model that provided a defined set of replicable program components including direct early learning services to infants and toddlers in out-of-home ECE settings; (2) a direct enhancement model with at least one replicable program component and provided direct early learning services to infants and toddlers in out-of-home ECE settings; or (3) an indirect enhancement model consisting of professional development programs with replicable program components focused on helping adult out-of-home caregivers to support infant and toddler early learning. ECE services had to be targeted to support learning for infants and toddlers broadly. The program could serve children who were enrolled after age 36 months, as long as it primarily targeted infants and toddlers. The study had to include at least one outcome in any of the primary child outcome domains identified for this review (language, cognition, social-emotional/behavioral). The study also had to be in a policy-relevant nondeveloping-world context. The study design had to be an RCT, SCD, MCGD, or RDD. The study had to be a primary study published in English by 1960 or later. RCTs with low attrition and no reassignment receive a high rating.

RCTs with high attrition or reassignment and MCGDs receive a moderate rating if they established baseline equivalence on required measures and used statistical controls. Studies that did not meet the requirements for a high or moderate rating received a low rating. Studies that had at least one high-rated outcome received a high rating. We further define study quality ratings in Chapter II of the main report.

Information from this table is summarized in the main report in Table III.1.

[a]The high-rated study reported on the first two Abecedarian cohorts. Later studies reported on the full sample of four cohorts.

This page has been left blank for double-sided copying.

EARLY HEAD START (EHS)

Description

Early Head Start (EHS) is a program that targets low-income pregnant women and families with children from birth through age 3. To be eligible, EHS families must have incomes at or below the federal poverty level. EHS aims to improve child development outcomes including health, social, cognitive, and language development. EHS also targets family, staff, and community development.

EHS programs offer a center-based approach, a home-based care approach, or a mixed approach offering both types of services. Children attend EHS until they are 3 years old. Because ACF/OPRE funds a separate review that includes the EHS home visiting approach (known as HomVEE), this review focuses only on the center-based and mixed-approach programs.

Center-based EHS programs provide services to families through center-based child care and parenting education. Centers are required to provide a minimum of 20 hours per week of child development services. Center-based programs also must provide a minimum of two home visits per family per year. Mixed-approach programs deliver the different types of services to meet the families' needs; some families receive the center-based option, some receive the home-based option, and others receive a combination of both options.

We identified five studies of EHS as eligible for review (Chazan-Cohen et al., 2013; Vogel et al., 2010, Whiteside-Mansell et al., 2009; ACF, 2001, 2002). These studies examined the impact of EHS in 17 sites as part of the Early Head Start Research and Evaluation Project (EHSREP); 4 sites provided care through the center-based approach, and 6 used the mixed approach.[56] The studies examined families who were enrolled at pregnancy or with a child younger than 12 months old in 1996 and followed them until the children were in 5th grade. Within each site, families were randomly assigned to either EHS or to a comparison group. The center-based sites enrolled 306 EHS and 306 comparison families; the mixed-approach sites enrolled 500 intervention and 504 comparison families. The comparison group families could not receive EHS services but were eligible for other services within the community.

This program model summary reports full sample results for center-based and mixed-approach programs separately. (Home-based programs were not eligible for review.) Study authors reported on additional subgroups that were not eligible for this review, because results for these subgroups were not reported separately by program service approach.

[56] One study in the review focused only on the center-based sites.

Domain-specific evidence of effectiveness ratings

Table A.5. EHS, center-based programs

Outcome domain	End of intervention effects	Sustained or delayed effects
Cognitive development	No discernible effects	Favorable effects
Language development	No discernible effects	No discernible effects
Social-emotional/ behavioral development	Favorable effects	Mixed effects[a]
Child health	No discernible effects	Not applicable

Notes: Possible ratings are as follows: **favorable effects** (at least one high- or moderate-rated study shows at least one significant or substantial favorable effect, no high-rated study shows any significant or substantial favorable effects, and no high- or moderate-rated study shows any significant or substantial unfavorable effects); **mixed effects** (at least one high-or moderate rated study shows at least one significant or substantial favorable effect and at least one high- or moderate-rated study shows at least one significant or substantial unfavorable effect); **no discernible effects** (no high- or moderate-rated study shows any significant or substantial effects, either favorable or unfavorable); and **unfavorable effects** (at least one high-or moderate-rated study shows at least one significant or substantial unfavorable effect, no high-rated study shows any significant or substantial unfavorable effects, and no high- or moderate-rated study shows any significant or substantial favorable effects). **End of intervention effects** are measured at 36 months and/or at the end of the intervention, if the intervention extends beyond 36 months. **Sustained or delayed effects** are measured one year or more after the end of the intervention. **Not applicable** indicates that no outcomes were measured that rated at least moderate in a study that met eligibility criteria.

Information from this table is summarized in the main report in Table III.3.

[a] EHS center-based programs had a mixed rating with one favorable effect and one unfavorable effect during the 5th grade follow-up.

Table A.6. EHS, mixed-approach programs

Outcome domain	End of intervention effects	Sustained or delayed effects
Cognitive development	Favorable effects	No discernible effects
Language development	Favorable effects	No discernible effects
Social-emotional/ behavioral development	Favorable effects	No discernible effects
Child health	No discernible effects	Not applicable

Notes: Possible ratings are as follows: **favorable effects** (at least one high- or moderate-rated study shows at least one significant or substantial favorable effect, no high-rated study shows any significant or substantial favorable effects, and no high- or moderate-rated study shows any significant or substantial unfavorable effects); **mixed effects** (at least one high-or moderate rated study shows at least one significant or substantial favorable effect and at least one high- or moderate-rated study shows at least one significant or substantial unfavorable effect); **no discernible effects** (no high- or moderate-rated study shows any significant or substantial effects, either favorable or unfavorable); and **unfavorable effects** (at least one high-or moderate-rated study shows at least one significant or substantial unfavorable effect, no high-rated study shows any significant or substantial unfavorable effects, and no high- or moderate-rated study shows any significant or substantial favorable effects). **End of intervention effects** are measured at 36 months and/or at the end of the intervention, if the intervention extends beyond 36 months. **Sustained or delayed effects** are measured one year or more after the end of the intervention. **Not applicable** indicates that no outcomes were measured that rated at least moderate in a study that met eligibility criteria.

Information from this table is summarized in the main report in Table III.3.

Summary of findings

Table A.7. EHS, center-based programs, primary domains and child health

Outcome domain	Significant or substantial favorable effects			Significant or substantial unfavorable effects			Null effects[a]
	During intervention	End of intervention	Sustained or delayed	During intervention	End of intervention	Sustained or delayed	
Cognitive development	0	0	1	0	0	0	15
Language development	0	0	0	0	0	0	8
Social-emotional/ behavioral development	0	1	1	0	0	1	34
Child health	0	0	Not applicable	0	0	Not applicable	3

Notes: A **significant or substantial effect** is one that is statistically significant ($p \leq 0.05$) or has an effect size greater than or equal to 0.2 standard deviations in absolute value A **sustained or delayed effect** is one that is statistically significant ($p \leq 0.05$) or has an effect size greater than or equal to 0.2 standard deviations and is measured one year or more after the end of the intervention. A **null effect** is an effect that is neither a significant or substantial favorable effect nor a significant or substantial unfavorable effect. **Not applicable** indicates that no outcomes were measured that rated at least moderate in a study that met eligibility criteria.

Information from this table is summarized in the main report in Tables III.4 and III.5.

[a]Includes effects measured during and immediately after the intervention, as well as in longer-term follow-ups.

Table A.8. EHS, center-based programs, interim domains

Outcome domain	Significant or substantial favorable effects	Significant or substantial unfavorable effects	Null effects
Parent or caregiver knowledge of child development	0	0	1
Global home environment	0	0	15
Parent- or caregiver-child interaction	0	2	15

Notes: A **significant or substantial effect** is one that is statistically significant ($p \leq 0.05$) or has an effect size greater than or equal to 0.2 standard deviations in absolute value. A **null effect** is an effect that is neither a significant or substantial favorable effect nor a significant or substantial unfavorable effect. This table includes effects measured during and immediately after the intervention, as well as in longer-term follow-ups.

Table A.9. EHS, mixed-approach programs, primary domains and child health

Outcome domain	Significant or substantial favorable effects			Significant or substantial unfavorable effects			Null effects[a]
	During intervention	End of intervention	Sustained or delayed	During intervention	End of intervention	Sustained or delayed	
Cognitive development	0	1	0	0	0	0	13
Language development	2	1	0	0	0	0	2
Social-emotional/ behavioral development	2	1	0	0	0	0	29
Child health	0	0	Not applicable	0	0	Not applicable	3

Notes: A **significant or substantial effect** is one that is statistically significant ($p \leq 0.05$) or has an effect size greater than or equal to 0.2 standard deviations in absolute value. A **sustained or delayed effect** is one that is statistically significant ($p \leq 0.05$) or has an effect size greater than or equal to 0.2 standard deviations and is measured one year or more after the end of the intervention. A **null effect** is an effect that is neither a significant or substantial favorable effect nor a significant or substantial unfavorable effect. **Not applicable** indicates that no outcomes were measured that rated at least moderate in a study that met eligibility criteria.

Information from this table is summarized in the main report in Tables III.4 and III.5.

[a]Includes effects measured during and immediately after the intervention, as well as in longer-term follow-ups.

Table A.10. EHS, mixed-approach programs, interim domains

Outcome domain	Significant or substantial favorable effects	Significant or substantial unfavorable effects	Null effects
Parent or caregiver knowledge of child development	1	0	0
Global home environment	1	0	14
Parent- or caregiver-child interaction	10	0	9

Notes: A **significant or substantial effect** is one that is statistically significant ($p \leq 0.05$) or has an effect size greater than or equal to 0.2 standard deviations in absolute value. A **null effect** is an effect that is neither a significant or substantial favorable effect nor a significant or substantial unfavorable effect. This table includes effects measured during and immediately after the intervention, as well as in longer-term follow-ups.

Table A.11. EHS, extent of research

	Number of studies
Eligible for review	5
Publication date of eligible studies	
1960 to 1979	0
1980 to 1999	0
2000 or later	5
Rating of eligible studies	
Rated high	3[a]
Rated moderate	1
Rated low	1

Notes: Study inclusion criteria as described in Chapter I of the main report are as follows: a study was eligible for review if the model was (1) a direct multicomponent model that provided a defined set of replicable program components including direct early learning services to infants and toddlers in out-of-home ECE settings; (2) a direct enhancement model with at least one replicable program component and provided direct early learning services to infants and toddlers in out-of-home ECE settings; or (3) an indirect enhancement model consisting of professional development programs with replicable program components focused on helping adult out-of-home caregivers to support infant and toddler early learning. ECE services had to be targeted to support learning for infants and toddlers broadly. The program could serve children who were enrolled after age 36 months, as long as it primarily targeted infants and toddlers. The study had to include at least one outcome in any of the primary child outcome domains identified for this review (language, cognition, social-emotional/behavioral). The study also had to be in a policy-relevant nondeveloping-world context. The study design had to be an RCT, SCD, MCGD, or RDD. The study had to be a primary study published in English by 1960 or later. RCTs with low attrition and no reassignment receive a high rating.

RCTs with high attrition or reassignment and MCGDs receive a moderate rating if they established baseline equivalence on required measures and used statistical controls. Studies that did not meet the requirements for a high or moderate rating received a low rating. Studies that had at least one high-rated outcome received a high rating. We further define study quality ratings in Chapter II of the main report.

Information from this table is summarized in the main report in Table III.1.

[a]These studies contain both high and moderate rated outcomes.

INFANT HEALTH AND DEVELOPMENT PROGRAM (IHDP)

Description

The **Infant Health and Development Program (IHDP)** intervention targeted premature (37 or fewer weeks gestational age) newborns with low birth weights (fewer than 2,500 grams) from eight medical institutions. The program targeted cognitive development, behavioral competence, and health status.

The intervention began when infants were discharged from the neonatal nursery and continued until children were 36 months old. Infants in the program received pediatric follow-up that included medical, developmental, and social assessments; referrals for pediatric care; and other services. The intervention also had three other main components: (1) home visits, (2) child development centers, and (3) parent groups.

During the child's first year, the family received weekly home visits. The frequency of visits was reduced to biweekly visits during the child's second and third years. During home visits, families received health and development information as well as family support services. The home visitor also implemented two curricula for the parents, one focused on cognitive, linguistic, and social development curricula, and the other focused on managing problems. From ages 12 to 36 months, children attended center-based care five days a week. Teachers in the child development centers continued to implement the cognitive, linguistic, and social development curricula and tailored the program to the children's needs. Classrooms with children ages 12 to 23 months had six children; classrooms with children ages 24 to 36 months had eight children. Each classroom had two teachers; therefore, the ratios were 1:3 and 1:4, respectively. The program also provided transportation to the center, if needed. Finally, parents with children at least 12 months old could attend bimonthly parent groups that discussed child rearing, health and safety, and other parenting topics.

We identified 11 studies of the IHDP intervention as eligible for review (Brooks-Gunn et al., 1993; Casey et al., 2009; Brooks-Gunn, Gross et al., 1992; Brooks-Gunn, Liaw et al., 1992; Brooks-Gunn et al., 1994; IHDP 1990; Liaw & Brooks-Gunn, 1994; McCarton et al., 1997; Spiker et al., 1993; Blair, 2002; Holloman & Scott, 1998). These studies followed a cohort of newborns from a single multisite randomized controlled trial that began in 1985 until the participants were 18 years old. Initially, 1,028 children from eight sites were randomly assigned to either IHDP or the comparison group using an adaptive randomization method in an attempt to achieve a 2:1 balance between the comparison and intervention groups for each site. Early in the study, 43 children withdrew and additional children had left during subsequent follow-ups. Children in the comparison group received the same pediatric follow-up as the children in IHDP, but they did not receive home visits, attend the child development centers, or have parents who participated in the parent groups. The effects for the full sample are presented here and summarized in Chapter III of the main report.

In addition to reporting on the full sample for IHDP, eligible subgroups reported here include the following:

- The IHDP full sample when children with Cerebral Palsy are removed from the analysis
- "Poor" Infants. A family was classified as "poor" if the annual income fell below the poverty threshold by an income –needs ratio of 1.5
- "Non-Poor" Infants. A family was classified as "non-poor" if the annual income exceeded the poverty threshold by an income–needs ratio of 1.5
- Children of Black Mothers
  - Mothers who attended college
  - Mothers who competed high school or less
- Children of White Mothers
  - Mothers who attended college
  - Mothers who competed high school or less

Six of the subgroups demonstrated evidence of effectiveness, primarily with favorable effects in the cognitive domain at the end of the intervention. A series of tables depict the evidence of effectiveness ratings for the full sample of IHDP followed by subgroup evidence of effectiveness ratings by domain. These tables are then followed by a series of tables depicting the findings for the full sample of IHDP followed by subgroup findings by domain.

Domain-specific evidence of effectiveness ratings

Table A.12. IHDP, full sample

Outcome domain	End of intervention effects	Sustained or delayed effects
Cognitive development	Favorable effects	No discernible effects
Language development	Favorable effects	No discernible effects
Social-emotional/ behavioral development	Favorable effects	No discernible effects
Child health	Unfavorable effects[a]	Mixed effects[b]

Notes: Possible ratings are as follows: **favorable effects** (at least one high- or moderate-rated study shows at least one significant or substantial favorable effect, no high-rated study shows any significant or substantial favorable effects, and no high- or moderate-rated study shows any significant or substantial unfavorable effects); **mixed effects** (at least one high-or moderate rated study shows at least one significant or substantial favorable effect and at least one high- or moderate-rated study shows at least one significant or substantial unfavorable effect); **no discernible effects** (no high- or moderate-rated study shows any significant or substantial effects, either favorable or unfavorable); and **unfavorable effects** (at least one high-or moderate-rated study shows at least one significant or substantial unfavorable effect, no high-rated study shows any significant or substantial unfavorable effects, and no high- or moderate-rated study shows any significant or substantial favorable effects). **End of intervention effects** are measured at 36 months and/or at the end of the intervention, if the intervention extends beyond 36 months. **Sustained or delayed effects** are measured one year or more after the end of the intervention. **Not applicable** indicates that no outcomes were measured that rated at least moderate in a study that met eligibility criteria.

Information from this table is summarized in the main report in Table III.3.

[a]This is morbidity during first three years of life which is the sum of reported injuries, surgeries, illnesses, and conditions.

[b]IHDP had a mix of one unfavorable effect and several favorable effect in child health during follow-ups.

Summary of findings

Table A.13. IHDP, full sample, primary domains and child health

Outcome domain	Significant or substantial favorable effects			Significant or substantial unfavorable effects			Null effects[a]
	During intervention	End of intervention	Sustained or delayed	During intervention	End of intervention	Sustained or delayed	
Cognitive development	5	3	0	0	0	0	6
Language development	2	3	0	0	0	0	3
Social-emotional/ behavioral development	1	2	0	0	0	0	7
Child health	Not applicable	0	4	Not applicable	1[b]	1[c]	8

Notes: A **significant or substantial effect** is one that is statistically significant ($p \leq 0.05$) or has an effect size greater than or equal to 0.2 standard deviations in absolute value. A **sustained or delayed effect** is one that is statistically significant ($p \leq 0.05$) or has an effect size greater than or equal to 0.2 standard deviations and is measured one year or more after the end of the intervention. A **null effect** is an effect that is neither a significant or substantial favorable effect nor a significant or substantial unfavorable effect. **Not applicable** indicates that no outcomes were measured that rated at least moderate in a study that met eligibility criteria.

Information from this table is summarized in the main report in Tables III.4 and III.5.

[a]This includes effects measured during and immediately after the intervention, as well as in longer-term follow-ups.

[b]This is morbidity during first three years of life which is the sum of reported injuries, surgeries, illnesses, and conditions.

[c]This is a measure of the performance of physical activities such as playing sports, walking up stairs, bending, lifting and caring for oneself at 96 months.

Table A.14. IHDP, full sample, interim domains

Outcome domain	Significant or substantial favorable effects	Significant or substantial unfavorable effects	Null effects
Parent- or caregiver-child interaction	2	0	2

Notes: A **significant or substantial effect** is one that is statistically significant ($p \leq 0.05$) or has an effect size greater than or equal to 0.2 standard deviations in absolute value. A **null effect** is an effect that is neither a significant or substantial favorable effect nor a significant or substantial unfavorable effect. This table includes effects measured during and immediately after the intervention, as well as in longer-term follow-ups.

Domain-specific evidence of effectiveness ratings, subgroups

Table A.15. IHDP, ratings of cognitive development, language, social-emotional/behavioral, and child health for subgroups

	Cognitive development		Language		Social-emotional/behavioral		Child health	
Subgroup	**End of intervention effects**	**Sustained or delayed effects**	**End of intervention effects**	**Sustained or delayed effects**	**End of intervention effects**	**Sustained or delayed effects**	**End of intervention effects**	**Sustained or delayed effects**
Full sample without Cerebral Palsy	Favorable effects	No discernible effects	Favorable effects	Favorable effects	Favorable effects	No discernible effects	Unfavorable effects[b]	No discernible effects
'Poor' infants	Favorable effects	Not applicable	Not applicable	Not applicable	Favorable effects	Not applicable	Not applicable	Not applicable
'Non-poor' infants	Favorable effects	Not applicable	Not applicable	Not applicable	Not applicable[a]	Not applicable	Not applicable	Not applicable
Children of black mothers who attended college	No discernible effects	Not applicable	Not applicable	Not applicable	Not applicable	Not applicable	Not applicable	Not applicable
Children of black mothers who completed high school or less education	Favorable effects	Not applicable	Not applicable	Not applicable	Not applicable	Not applicable	Not applicable	Not applicable
Children of white mothers who attended college	Favorable effects	Not applicable	Not applicable	Not applicable	Not applicable	Not applicable	Not applicable	Not applicable
Children of white mothers who completed high school or less education	Favorable effect	Not applicable	Not applicable	Not applicable	Not applicable	Not applicable	Not applicable	Not applicable

Notes: Possible ratings are as follows: **favorable effects** (at least one high- or moderate-rated study shows at least one significant or substantial favorable effect, no high-rated study shows any significant or substantial favorable effects, and no high- or moderate-rated study shows any significant or substantial unfavorable effects); **mixed effects** (at least one high-or moderate rated study shows at least one significant or substantial favorable effect and at least one high- or moderate-rated study shows at least one significant or substantial unfavorable effect); **no discernible effects** (no high- or moderate-rated study shows any significant or substantial effects, either favorable or unfavorable); and **unfavorable effects** (at least one high-or moderate-rated study shows at least one significant or substantial unfavorable effect, no high-rated study shows any significant or substantial unfavorable effects, and no high- or moderate-rated study shows any significant or substantial favorable effects). **End of intervention effects** are measured at 36 months and/or at the end of the intervention, if the intervention extends beyond 36 months. **Sustained or delayed effects** are measured one year or more after the end of the intervention. **Not applicable** indicates that no outcomes were measured that rated at least moderate in a study that met eligibility criteria.

[a]For the Social-Emotional/Behavioral Development domain, one outcome was measured but impact estimates and significance levels were not reported.

[b]This is morbidity during first three years of life which is the sum of reported injuries, surgeries, illnesses, and conditions.

Table A.16. IHDP, extent of research

	Number of studies
Eligible for review	11
Publication date of eligible studies	
1960 to 1979	0
1980 to 1999	9
2000 or later	2
Rating of eligible studies	
Rated high	2[a]
Rated moderate	7[b]
Rated low	2

Notes: Study inclusion criteria as described in Chapter I of the main report are as follows: a study was eligible for review if the model was (1) a direct multicomponent model that provided a defined set of replicable program components including direct early learning services to infants and toddlers in out-of-home ECE settings; (2) a direct enhancement model with at least one replicable program component and provided direct early learning services to infants and toddlers in out-of-home ECE settings; or (3) an indirect enhancement model consisting of professional development programs with replicable program components focused on helping adult out-of-home caregivers to support infant and toddler early learning. ECE services had to be targeted to support learning for infants and toddlers broadly. The program could serve children who were enrolled after age 36 months, as long as it primarily targeted infants and toddlers. The study had to include at least one outcome in any of the primary child outcome domains identified for this review (language, cognition, social-emotional/behavioral). The study also had to be in a policy-relevant nondeveloping-world context. The study design had to be an RCT, SCD, MCGD, or RDD. The study had to be a primary study published in English by 1960 or later. RCTs with low attrition and no reassignment receive a high rating.

RCTs with high attrition or reassignment and MCGDs receive a moderate rating if they established baseline equivalence on required measures and used statistical controls. Studies that did not meet the requirements for a high or moderate rating received a low rating. Studies that had at least one high-rated outcome received a high rating. We further define study quality ratings in Chapter II of the main report.

Information from this table is summarized in the main report in Table III.1.

[a]These studies contained both high and moderate rated outcomes.

[b]For two of these studies, the moderate rating was based on subgroup-only results.

PARENT-CHILD DEVELOPMENT CENTER (PCDC)

Description

The **Parent-Child Development Center (PCDC)** program was a multisite intervention for low-income families and children. The program targeted cognitive, language, and behavioral development.

Three sites, Birmingham, Houston, and New Orleans, were part of the program. Each site had a slightly different approach and program components, but there were some core similarities. All sites had programming for both mothers and children and the program lasted until the children were 36 months old. All programs provided transportation, some meals, health and social services, and a stipend. The parenting component emphasized children's social-emotional, intellectual, and personal and physical development, home management, health, and community resources.

The Birmingham site was center-based only and children entered the program at 3 months of age. Children in the Birmingham site were grouped in classes according to age. At the Houston site, the first year of the program was home-based and the second year was center-based; the children began programming at age 12 months. In the first year, there were about 30 home visits, each lasting 1.5 hours. During the second year, children attended a nursery school four days per week. New Orleans used a mixed approach in which some families received home-based support and others received center-based support; the children began programming at age 2 months. In the center-based care, children attended a laboratory school six hours per week and were grouped by age.

We identified six PCDC studies as eligible for review (Bridgeman, 1981; Andres et al., 1982; Johnson, 1976, 2006; Johnson & Breckenridge, 1982; Johnson & Walker, 1991). Because each site had different programming, outcomes were evaluated separately by site. However, all sites used an RCT design and followed children until they were 16 years old. Families assigned to the comparison conditions received no services. The Birmingham site included one cohort of families that entered the program after 1972 and graduated before fall 1980. The randomized initial sample had 267 families in the intervention group and 183 in the comparison group. In Houston, there was a four-cohort design in which 185 families were assigned to the intervention group and 188 to the comparison group. In New Orleans, there was a four-cohort design in which 67 families were assigned to the intervention group and 59 to the comparison group. All RCTs of PCDC had high attrition.

Full sample results for the Birmingham sample only are presented in here and in Chapter III. One study established baseline equivalence on the analytic sample for three cognitive outcomes from the Birmingham sample and received a moderate rating. The studies based on the Houston and New Orleans sites rated low. No eligible subgroups were reported in studies of this program model.

Domain-specific evidence of effectiveness ratings

Table A.17. PCDC, full Birmingham sample

Outcome domain	End of intervention effects	Sustained or delayed effects
Cognitive development	Favorable effects	Not applicable
Language development	Not applicable	Not applicable
Social-emotional/behavioral development	Not applicable	Not applicable
Child health	Not applicable	Not applicable

Notes: Possible ratings are as follows: **favorable effects** (at least one high- or moderate-rated study shows at least one significant or substantial favorable effect, no high-rated study shows any significant or substantial favorable effects, and no high- or moderate-rated study shows any significant or substantial unfavorable effects); **mixed effects** (at least one high-or moderate rated study shows at least one significant or substantial favorable effect and at least one high- or moderate-rated study shows at least one significant or substantial unfavorable effect); **no discernible effects** (no high- or moderate-rated study shows any significant or substantial effects, either favorable or unfavorable); and **unfavorable effects** (at least one high-or moderate-rated study shows at least one significant or substantial unfavorable effect, no high-rated study shows any significant or substantial unfavorable effects, and no high- or moderate-rated study shows any significant or substantial favorable effects). **End of intervention effects** are measured at 36 months and/or at the end of the intervention, if the intervention extends beyond 36 months. **Sustained or delayed effects** are measured one year or more after the end of the intervention. **Not applicable** indicates that no outcomes were measured that rated at least moderate in a study that met eligibility criteria.

Information from this table is summarized in the main report in Table III.3.

Summary of findings

Table A.18. PCDC, full Birmingham sample, primary domains and child health

Outcome domain	Significant or substantial favorable effects			Significant or substantial unfavorable effects			Null effects[a]
	During intervention	End of intervention	Sustained or delayed	During intervention	End of intervention	Sustained or delayed	
Cognitive development	1	2	Not applicable	0	0	Not applicable	0
Language development	Not applicable	Not applicable	Not applicable	Not applicable	Not applicable	Not applicable	Not applicable
Social-emotional/ behavioral development	Not applicable	Not applicable	Not applicable	Not applicable	Not applicable	Not applicable	Not applicable
Child health	Not applicable	Not applicable	Not applicable	Not applicable	Not applicable	Not applicable	Not applicable

Notes: A **significant or substantial effect** is one that is statistically significant ($p \leq 0.05$) or has an effect size greater than or equal to 0.2 standard deviations in absolute value. A **sustained or delayed effect** is one that is statistically significant ($p \leq 0.05$) or has an effect size greater than or equal to 0.2 standard deviations and is measured one year or more after the end of the intervention. A **null effect** is an effect that is neither a significant or substantial favorable effect nor a significant or substantial unfavorable effect. **Not applicable** indicates that no outcomes were measured that rated at least moderate in a study that met eligibility criteria.

Information from this table is summarized in the main report in Tables III.4 and III.5.

[a]Includes effects measured during and immediately after the intervention, as well as in longer-term follow-ups.

Table A.19. PCDC, extent of research

	Number of studies
Eligible for review	6
Publication date of eligible studies	
1960 to 1979	1
1980 to 1999	4
2000 or later	1
Rating of eligible studies	
Rated high	0
Rated moderate	1
Rated low	5

Notes: Study inclusion criteria as described in Chapter I of the main report are as follows: a study was eligible for review if the model was (1) a direct multicomponent model that provided a defined set of replicable program components including direct early learning services to infants and toddlers in out-of-home ECE settings; (2) a direct enhancement model with at least one replicable program component and provided direct early learning services to infants and toddlers in out-of-home ECE settings; or (3) an indirect enhancement model consisting of professional development programs with replicable program components focused on helping adult out-of-home caregivers to support infant and toddler early learning. ECE services had to be targeted to support learning for infants and toddlers broadly. The program could serve children who were enrolled after age 36 months, as long as it primarily targeted infants and toddlers. The study had to include at least one outcome in any of the primary child outcome domains identified for this review (language, cognition, social-emotional/behavioral). The study also had to be in a policy-relevant nondeveloping-world context. The study design had to be an RCT, SCD, MCGD, or RDD. The study had to be a primary study published in English by 1960 or later. RCTs with low attrition and no reassignment receive a high rating.

RCTs with high attrition or reassignment and MCGDs receive a moderate rating if they established baseline equivalence on required measures and used statistical controls. Studies that did not meet the requirements for a high or moderate rating received a low rating. Studies that had at least one high-rated outcome received a high rating. We further define study quality ratings in Chapter II of the main report.

Information from this table is summarized in the main report in Table III.1.

BROOKLINE EARLY EDUCATION PROJECT (BEEP)

Description

The **Brookline Early Education Project (BEEP)** was a community-based research and education program for children from birth to age 5 and their families living in Brookline, Massachusetts. The project ran from 1972 to 1979. It emphasized early detection of conditions that could impair children's health and/or function as they approach school age. BEEP was also designed to support parents as teachers of their children. The program aimed to serve a diverse population of children and families. It consisted of assessments or diagnostic services, and education programs for children and parents.

The assessment and diagnostic services consisted of providing comprehensive health and developmental examinations to all children enrolled in BEEP at ages 2 weeks, 3 months, 6 months, 14 months, 24 months, 30 months, and 42 months, and at kindergarten entry. The parent education components included home visits with families from the children's birth until they were 2 years old. Parents also had the opportunity to attend group meetings, parent-teacher conferences, and classroom visits as their children progressed into center care. When children were 2 years old, they attended weekly playgroups at BEEP centers. When children were 3 and 4 years old, they attended a daily morning pre-K program in local elementary schools that used a curriculum focused on building social skills and competencies in school.

We identified four BEEP studies as eligible for review (Bronson et al., 1994; Palfrey et al., 2005; Palfrey et al., 1987; Tivnan & Pierson, 1982). They followed children born between 1973 and 1974 who were enrolled in BEEP from birth to kindergarten entry until age 25. BEEP originally enrolled 282 children. Each of the four studies had comparison groups that were formed retrospectively during follow-ups. Comparison group children did not receive any specific intervention.

All eligible studies of BEEP received a low rating. They were MCGDs in which the comparison group was formed retrospectively. Thus, this review could not assess baseline equivalence on socioeconomic status, one of the measures required for MCGDs to receive a moderate rating.

Table A.20. BEEP, extent of research

	Number of studies
Eligible for review	4
Publication date of eligible studies	
1960 to 1979	0
1980 to 1999	3
2000 or later	1
Rating of eligible studies	
Rated high	0
Rated moderate	0
Rated low	4

Notes: Study inclusion criteria as described in Chapter I of the main report are as follows: a study was eligible for review if the model was (1) a direct multicomponent model that provided a defined set of replicable program components including direct early learning services to infants and toddlers in out-of-home ECE settings; (2) a direct enhancement model with at least one replicable program component and provided direct early learning services to infants and toddlers in out-of-home ECE settings; or (3) an indirect enhancement model consisting of professional development programs with replicable program components focused on helping adult out-of-home caregivers to support infant and toddler early learning. ECE services had to be targeted to support learning for infants and toddlers broadly. The program could serve children who were enrolled after age 36 months, as long as it primarily targeted infants and toddlers. The study had to include at least one outcome in any of the primary child outcome domains identified for this review (language, cognition, social-emotional/behavioral). The study also had to be in a policy-relevant nondeveloping-world context. The study design had to be an RCT, SCD, MCGD, or RDD. The study had to be a primary study published in English by 1960 or later. RCTs with low attrition and no reassignment receive a high rating.

RCTs with high attrition or reassignment and MCGDs receive a moderate rating if they established baseline equivalence on required measures and used statistical controls. Studies that did not meet the requirements for a high or moderate rating received a low rating. Studies that had at least one high-rated outcome received a high rating. We further define study quality ratings in Chapter II of the main report.

Information from this table is summarized in the main report in Table III.1.

MILWAUKEE INFANT STIMULATION PROJECT

Description

The **Milwaukee Infant Stimulation Project** was a comprehensive intervention program designed to modify the natural environments of high-risk infants to improve cognitive development. It targeted African-American mothers with an IQ score below 75 from a low-income area of Milwaukee. The program had two components—a mother intervention and an infant intervention. It was provided to children from age 3 months to 6 years.

The infant intervention was divided into two programs: (1) the infancy program for children ages 3 to 24 months and (2) the preschool program for children ages 24 months to 6 years. The infancy program was intended to support perceptual-motor, cognitive-language, and social-emotional development. For 3- to 12-month-old children, the teacher-to-child ratio was 1:1; for 12- to 18-month-old children, the teacher-to-child ratio was 1:2; and for 18- to 24-month-old children, the teacher-to-child ratio was 1:3. The preschool program was intended to support language, reading, and mathematics/problem solving. In the preschool program, children of the same age were grouped in a classroom with three teachers. The full-day program was delivered five days a week, 12 months per year, and transportation to the center was provided.

The mother intervention, which aimed to prepare mothers for employment and improve their homemaking and childrearing skills, had two phases: (1) a one-month rehabilitation program that provided adult education classes on life skills and job readiness skills four days per week, and (2) a 26-week occupational training program, delivered three days per week, on vocational skills in laundry, housekeeping, food service, or nursing. Group counseling sessions were held at the end of each day of training.

We identified three Milwaukee Infant Stimulation Project studies as eligible for review (Garber, 1988; Heber, 1972; McBride, 1989). The studies followed a cohort of children from a randomized controlled trial (RCT) that began in 1965 until the children were in 10th grade. Fifty-five mothers were randomly assigned on a rolling basis over 24 months into either the intervention (N = 28) or comparison condition (N = 27). Due to attrition, the final sample included 20 children in the intervention group and 20 children in the comparison group. Those in the comparison group did not receive any intervention.

All eligible studies of the Milwaukee Infant Stimulation Project received a low rating. They were RCTs with high attrition or reassignment in which no controls or required statistical adjustments were used in analyses of outcomes.

Table A.21. Milwaukee, extent of research

	Number of studies
Eligible for review	3
Publication date of eligible studies	
1960 to 1979	1
1980 to 1999	2
2000 or later	0
Rating of eligible studies	
Rated high	0
Rated moderate	0
Rated low	3

Notes: Study inclusion criteria as described in Chapter I of the main report are as follows: a study was eligible for review if the model was (1) a direct multicomponent model that provided a defined set of replicable program components including direct early learning services to infants and toddlers in out-of-home ECE settings; (2) a direct enhancement model with at least one replicable program component and provided direct early learning services to infants and toddlers in out-of-home ECE settings; or (3) an indirect enhancement model consisting of professional development programs with replicable program components focused on helping adult out-of-home caregivers to support infant and toddler early learning. ECE services had to be targeted to support learning for infants and toddlers broadly. The program could serve children who were enrolled after age 36 months, as long as it primarily targeted infants and toddlers. The study had to include at least one outcome in any of the primary child outcome domains identified for this review (language, cognition, social-emotional/behavioral). The study also had to be in a policy-relevant nondeveloping-world context. The study design had to be an RCT, SCD, MCGD, or RDD. The study had to be a primary study published in English by 1960 or later. RCTs with low attrition and no reassignment receive a high rating.

RCTs with high attrition or reassignment and MCGDs receive a moderate rating if they established baseline equivalence on required measures and used statistical controls. Studies that did not meet the requirements for a high or moderate rating received a low rating. Studies that had at least one high-rated outcome received a high rating. We further define study quality ratings in Chapter II of the main report.

Information from this table is summarized in the main report in Table III.1.

PROJECT CAROLINA APPROACH TO RESPONSIVE EDUCATION (PROJECT CARE)

Description

Project Carolina Approach to Responsive Education (Project CARE) was an intervention for infants at risk for developmental delay based on their families' low income, low educational attainment, and other family challenges. Children participated in Project CARE from age 6 weeks to 5 years. The program focused on improving cognitive, language, social development, and health. It had a child care component similar to the Abecedarian program and a family education component which involved home visiting. LITES considers Project CARE a distinct intervention from Abecedarian because of the home visiting component.

Children began attending the full-day child care program as young as age 6 weeks. Groups in the child care program ranged from four to seven children; the teacher-to-child ratio was 1:3 for infants and toddlers, 1:4 for 2-year-olds, and 1:6 for 3- to 5-year-olds. Project CARE staff had ethnic or cultural backgrounds similar to those of the families. All staff had previous training in caregiving and access to opportunities for continued staff training. Each program employed at least one male teacher and/or assistant teacher. Volunteers were also a major component of the program. Project CARE used a specified curriculum, LearningGames, which is divided into game-like activities that are applied in cycles and integrated into all services (see Appendix B). For children older than age 2, the program offered regular field trips out of the center. The family education component consisted of a toy-lending library, home visits, and parent group meetings. The program center also had a medical program, in which medical staff provided routine examinations, screenings, immunization, and care for illnesses. Project CARE provided nutritious meals to the children in group-like settings.

We identified two Project CARE studies as eligible for review (Ramey et al., 1985; Wasik et al., 1990). The studies followed one cohort of children who were enrolled in infancy from 1978 to 1980 until they were 4.5 years old. Families were randomly assigned to one of three groups: the child care plus family education group (N = 17 families), the family education-only group (N = 25 families), or the comparison group (N = 23 families). Children in the comparison group received no services other than receiving iron-fortified formula until age 15 months; children in the other groups also received the formula.

Only the comparisons with the child care plus family education group were eligible for review. Both eligible studies of Project CARE received a low rating. They were RCTs with high attrition in which no controls or required statistical adjustments were used in analyses of outcomes.

Table A.22. Project CARE, extent of research

	Number of studies
Eligible for review	2
Publication date of eligible studies	
1960 to 1979	0
1980 to 1999	2
2000 or later	0
Rating of eligible studies	
Rated high	0
Rated moderate	0
Rated low	2

Notes: Study inclusion criteria as described in Chapter I of the main report are as follows: a study was eligible for review if the model was (1) a direct multicomponent model that provided a defined set of replicable program components including direct early learning services to infants and toddlers in out-of-home ECE settings; (2) a direct enhancement model with at least one replicable program component and provided direct early learning services to infants and toddlers in out-of-home ECE settings; or (3) an indirect enhancement model consisting of professional development programs with replicable program components focused on helping adult out-of-home caregivers to support infant and toddler early learning. ECE services had to be targeted to support learning for infants and toddlers broadly. The program could serve children who were enrolled after age 36 months, as long as it primarily targeted infants and toddlers. The study had to include at least one outcome in any of the primary child outcome domains identified for this review (language, cognition, social-emotional/behavioral). The study also had to be in a policy-relevant nondeveloping-world context. The study design had to be an RCT, SCD, MCGD, or RDD. The study had to be a primary study published in English by 1960 or later. RCTs with low attrition and no reassignment receive a high rating.

RCTs with high attrition or reassignment and MCGDs receive a moderate rating if they established baseline equivalence on required measures and used statistical controls. Studies that did not meet the requirements for a high or moderate rating received a low rating. Studies that had at least one high-rated outcome received a high rating. We further define study quality ratings in Chapter II of the main report.

Information from this table is summarized in the main report in Table III.1.

SURE START LOCAL PROGRAMMES (SSLPS)

Description

Sure Start Local Programmes (SSLPs) was an area-based intervention in the United Kingdom that aimed to improve the health and development of children younger than 4 years old. The intervention targeted all children from birth to age 3 and their families who lived in small socially disadvantaged communities. SSLPs did not have a prescribed set of services, and each local program worked within the community to improve existing services while providing at least the core services. Core services included outreach or home visiting; family support for good-quality play, learning, and child care experiences; primary and community health care; advice about child and family health and development; and support for people with special needs. Programs were established between 1999 and 2003. From 2004 to 2006, SSLPs became Sure Start Children's Centers, which had a more specified set of services.

We identified two studies of SSLPs as eligible for review (Belsky et al., 2006; Melhuish et al., 2008). Both studies used a matched group comparison design and the intervention group included families living in SSLP communities rather than only families who participated in SSLPs. The first study was conducted in 2003 to 2004. The intervention group included 16,502 children who were either 9 months old or 36 months old from randomly selected families from 150 randomly selected communities that implemented SSLPs. The comparison group consisted of 2,610 children who were 9 or 36 months old from families from 50 areas waiting to become SSLP areas. The second study was conducted about two years later and drew its sample from children who were 9 months old in the first study. Therefore, the sample group members lived in areas in which SSLPs became Sure Start Children's Centers. The intervention group included 5,883 3-year-old children and their families from 93 SSLP communities. The matched-sample comparison group consisted of 1,879 3-year-old children and their families from 72 similarly disadvantaged areas.

Both eligible studies of SSLPs received a low rating. They were MCGDs that lacked a true baseline; that is, baseline variables were not measured prior to the start of the intervention. Thus, this review could not assess baseline equivalence on socioeconomic status, one of the measures required for MCGDs to receive a moderate rating.

Table A.23. SSLPs, extent of research

	Number of studies
Eligible for review	2
Publication date of eligible studies	
1960 to 1979	0
1980 to 1999	0
2000 or later	2
Rating of eligible studies	
Rated high	0
Rated moderate	0
Rated low	2

Notes: Study inclusion criteria as described in Chapter I of the main report are as follows: a study was eligible for review if the model was (1) a direct multicomponent model that provided a defined set of replicable program components including direct early learning services to infants and toddlers in out-of-home ECE settings; (2) a direct enhancement model with at least one replicable program component and provided direct early learning services to infants and toddlers in out-of-home ECE settings; or (3) an indirect enhancement model consisting of professional development programs with replicable program components focused on helping adult out-of-home caregivers to support infant and toddler early learning. ECE services had to be targeted to support learning for infants and toddlers broadly. The program could serve children who were enrolled after age 36 months, as long as it primarily targeted infants and toddlers. The study had to include at least one outcome in any of the primary child outcome domains identified for this review (language, cognition, social-emotional/behavioral). The study also had to be in a policy-relevant nondeveloping-world context. The study design had to be an RCT, SCD, MCGD, or RDD. The study had to be a primary study published in English by 1960 or later. RCTs with low attrition and no reassignment receive a high rating.

RCTs with high attrition or reassignment and MCGDs receive a moderate rating if they established baseline equivalence on required measures and used statistical controls. Studies that did not meet the requirements for a high or moderate rating received a low rating. Studies that had at least one high-rated outcome received a high rating. We further define study quality ratings in Chapter II of the main report.

Information from this table is summarized in the main report in Table III.1.

YALE CHILD WELFARE RESEARCH PROGRAM

Description

The **Yale Child Welfare Research Program** was a comprehensive intervention that served low-income children and families from pregnancy until the children were 30 months old. Two goals of the program were to help parents raise healthy children and to improve children's cognitive and social-emotional development. The intervention had four components: (1) home visits, (2) pediatric care, (3) developmental evaluation, and (4) child care or toddler school.

Home visits occurred at least monthly, and families received additional services as needed. Routine pediatric visits occurred monthly from birth to age 12 months and then at ages 15, 16, 21, 24, 27, and 30 months. Pediatricians were also available in case of illness. Developmental examinations conducted by pediatricians and psychologists were administered at ages 2, 3, 6, 9, 12, 18, 24, and 30 months. All results were shared with the families to help them understand their children's development. Families were given the option of full-day child care or traditional toddler nursery school for the children. Services were flexible and adjusted according to the families' needs. Children could start and end enrollment in the child care program at any age. Toddler school began when the children were between 15 and 18 months old, and children met twice a week for an hour and a half with their mothers present.

We identified one Yale Child Welfare Research Program study as eligible for review (Trickett et al., 1982). The study was a MCGD follow-up five years after the children were in the program. The follow-up intervention group sample consisted of 17 children from 16 families. Eighteen children from 17 families were originally enrolled between 1968 and 1970. Two comparison groups were formed retrospectively by sampling children from two neighborhoods similar to those of the intervention group families.

This eligible study received a low rating. It was an MCGD in which the comparison group was formed retrospectively. Thus, this review could not assess baseline equivalence on socioeconomic status, one of the measures required for MCGDs to receive a moderate rating.

Table A.24. Yale Child Welfare Research Program, extent of research

	Number of studies
Eligible for review	1
Publication date of eligible studies	
1960 to 1979	0
1980 to 1999	1
2000 or later	0
Rating of eligible studies	
Rated high	0
Rated moderate	0
Rated low	1

Notes: Study inclusion criteria as described in Chapter I of the main report are as follows: a study was eligible for review if the model was (1) a direct multicomponent model that provided a defined set of replicable program components including direct early learning services to infants and toddlers in out-of-home ECE settings; (2) a direct enhancement model with at least one replicable program component and provided direct early learning services to infants and toddlers in out-of-home ECE settings; or (3) an indirect enhancement model consisting of professional development programs with replicable program components focused on helping adult out-of-home caregivers to support infant and toddler early learning. ECE services had to be targeted to support learning for infants and toddlers broadly. The program could serve children who were enrolled after age 36 months, as long as it primarily targeted infants and toddlers. The study had to include at least one outcome in any of the primary child outcome domains identified for this review (language, cognition, social-emotional/behavioral). The study also had to be in a policy-relevant nondeveloping-world context. The study design had to be an RCT, SCD, MCGD, or RDD. The study had to be a primary study published in English by 1960 or later. RCTs with low attrition and no reassignment receive a high rating.

RCTs with high attrition or reassignment and MCGDs receive a moderate rating if they established baseline equivalence on required measures and used statistical controls. Studies that did not meet the requirements for a high or moderate rating received a low rating. Studies that had at least one high-rated outcome received a high rating. We further define study quality ratings in Chapter II of the main report.

Information from this table is summarized in the main report in Table III.1.

APPENDIX B

DIRECT AND INDIRECT ENHANCEMENT MODEL SUMMARIES

This page has been left blank for double-sided copying.

PROGRAM FOR INFANT/TODDLER CARE (PITC)

Description

The **Program for Infant/Toddler Care (PITC)** combines direct caregiver training and on-site coaching or tailored assistance. The PITC program aims to indirectly improve children's language, cognitive, and social-emotional development through professional development for teachers. It has six essential policies: (1) primary care (assignment of a primary caregiver to each child); (2) small groups (0 to 8 months: 6 children per group, 8 to 18 months: 9 children per group, 18 months and older: 12 children per group, or mixed-ages: 8 children per group); (3) continuity of care; (4) individualized schedules and routines; (5) inclusion of children with special needs; and (6) cultural sensitivity. PITC-certified trainers deliver 64 hours of training and 40 hours of coaching or other support to each participating child care center or group. Center-based programs receive the intervention on site; family child care providers participate in groups of 5 to 10 programs, and the intervention is delivered in a provider's home, a community center, or a school.

We identified one PITC study as eligible for review (Weinstock et al., 2012). The study used a cluster-based random assignment sample of 251 childcare programs that were randomly assigned to the intervention (N = 124 programs [46 centers and 78 family child care homes]) and comparison (N = 127 programs [46 centers and 81 family child care homes]) groups. Enrollment took place on a rolling basis from October 2007 to July 2008 in six Southern California counties and four Arizona counties. The study focused on children who received child care from a PITC provider for at least 20 hours per week and were younger than 27 months when enrolled. There were 936 children (480 in the intervention group and 456 in the comparison group) in the study. Children and programs were followed until 22 months after random assignment.

PITC received an unfavorable effect rating at the end of the intervention in the social-emotional/behavioral development domain based on one significant unfavorable effect. PITC also had one substantial favorable effect in an interim outcome domain: parent- or caregiver-child interaction.

Domain-specific evidence of effectiveness ratings

Table B.1. PITC, full sample

Outcome domain	End of intervention effects	Sustained or delayed effects
Cognitive development	No discernible effects	Not applicable
Language development	No discernible effects	Not applicable
Social-emotional/ behavioral development	Unfavorable effect	Not applicable
Child health	Not applicable	Not applicable

Notes: Possible ratings are as follows: **favorable effects** (at least one high- or moderate-rated study shows at least one significant or substantial favorable effect, no high-rated study shows any significant or substantial favorable effects, and no high- or moderate-rated study shows any significant or substantial unfavorable effects); **mixed effects** (at least one high-or moderate rated study shows at least one significant or substantial favorable effect and at least one high- or moderate-rated study shows at least one significant or substantial unfavorable effect); **no discernible effects** (no high- or moderate-rated study shows any significant or substantial effects, either favorable or unfavorable); and **unfavorable effects** (at least one high-or moderate-rated study shows at least one significant or substantial unfavorable effect, no high-rated study shows any significant or substantial unfavorable effects, and no high- or moderate-rated study shows any significant or substantial favorable effects). **End of intervention effects** are measured at 36 months and/or at the end of the intervention, if the intervention extends beyond 36 months. **Sustained or delayed effects** are measured one year or more after the end of the intervention. **Not applicable** indicates that no outcomes were measured that rated at least moderate in a study that met eligibility criteria.

Summary of findings

Table B.2. PITC, full sample, primary domains and child health

Outcome domain	Significant or substantial favorable effects			Significant or substantial unfavorable effects			Null effects[a]
	During intervention	End of intervention	Sustained or delayed	During intervention	End of intervention	Sustained or delayed	
Cognitive development	0	0	Not applicable	0	0	Not applicable	2
Language development	0	0	Not applicable	0	0	Not applicable	1
Social-emotional/ behavioral development	0	0	Not applicable	0	1	Not applicable	3
Child health	Not applicable	Not applicable	Not applicable	Not applicable	Not applicable	Not applicable	Not applicable

Notes: A **significant or substantial effect** is one that is statistically significant ($p \leq 0.05$) or has an effect size greater than or equal to 0.2 standard deviations in absolute value. A **sustained or delayed effect** is one that is statistically significant ($p \leq 0.05$) or has an effect size greater than or equal to 0.2 standard deviations and is measured one year or more after the end of the intervention. A **null effect** is an effect that is neither a significant or substantial favorable effect nor a significant or substantial unfavorable effect. **Not applicable** indicates that no outcomes were measured that rated at least moderate in a study that met eligibility criteria.

[a]Includes effects measured during and immediately after the intervention, as well as in longer-term follow-ups.

Table B.3. PITC, full sample, interim domains

Outcome domain	Significant or substantial favorable effects	Significant or substantial unfavorable effects	Null effects
Global child care quality	0	0	2
Parent- or caregiver-child interaction	1	0	1

Notes: A **significant or substantial effect** is one that is statistically significant ($p \leq 0.05$) or has an effect size greater than or equal to 0.2 standard deviations in absolute value. A **null effect** is an effect that is neither a significant or substantial favorable effect nor a significant or substantial unfavorable effect. Subscale results do not count toward evidence of effectiveness. This table includes effects measured during and immediately after the intervention, as well as in longer-term follow-ups.

Table B.4. PITC, extent of research

	Number of studies
Eligible for review	1
Publication date of eligible studies	
1960 to 1979	0
1980 to 1999	0
2000 or later	1
Rating of eligible studies	
Rated high	1[a]
Rated moderate	0
Rated low	0

Notes: Possible ratings are as follows: **favorable effects** (at least one high- or moderate-rated study shows at least one significant or substantial favorable effect, no high-rated study shows any significant or substantial favorable effects, and no high- or moderate-rated study shows any significant or substantial unfavorable effects); **mixed effects** (at least one high-or moderate rated study shows at least one significant or substantial favorable effect and at least one high- or moderate-rated study shows at least one significant or substantial unfavorable effect); **no discernible effects** (no high- or moderate-rated study shows any significant or substantial effects, either favorable or unfavorable); and **unfavorable effects** (at least one high-or moderate-rated study shows at least one significant or substantial unfavorable effect, no high-rated study shows any significant or substantial unfavorable effects, and no high- or moderate-rated study shows any significant or substantial favorable effects). **End of intervention effects** are measured at 36 months and/or at the end of the intervention, if the intervention extends beyond 36 months. **Sustained or delayed effects** are measured one year or more after the end of the intervention. **Not applicable** indicates that no outcomes were measured that rated at least moderate in a study that met eligibility criteria.

Information from this table is summarized in the main report in Table III.6.

[a]This study contains both high and moderate rated outcomes.

EAGER AND ABLE TO LEARN

Eager and Able to Learn

Description

Eager and Able to Learn is a professional development program for caregivers of children ages 2 to 3 in Northern Ireland aimed at improving young children's eagerness and ability to learn through physical, social, emotional, and language development. The program emphasizes physical movement, the physical design of the program settings, and relationships between children, parents, and practitioners. Eager and Able to Learn has six core components: (1) initial practitioner training, (2) ongoing training for practitioners, (3) Early Years Specialist support (SEYS), (4) service design manual, (5) resources for group environment, and (6) home learning package for parents. The initial preparation is a 42-hour training across six days that helps practitioners understand the theory and rationale of the program and supports teachers' skills and knowledge to begin delivering the program. The ongoing "cluster" training sessions take place bi-monthly and include four three-hour sessions that expand on the initial training and address any difficulties during implementation. SEYS works with settings for at least five hours per month and provides mentoring, modeling, and peer support training. The service design manual provides theoretical background, details various elements of the program, and describes the roles of those involved. The resources provided include toys and apparatus to promote physical development, such as slides, balls, and clay. Finally, the home resources for parents include a manual for directing play activities and resources to carry out the activities. Caregivers also provide three home visits throughout the year to emphasize the program.

We identified one Eager and Able to Learn study as eligible for review (McGuinness et al., 2012). The study, which took place from September 2008 to June 2010, was an MCGD with a partial crossover design. Children from 18 child care settings and 18 Sure Start settings participated in the study.[57] The comparison group included 197 children between ages 24 months and 33 months who attended the centers in the 2008–2009 school year before the intervention was implemented. The intervention group comprised of 257 children who attended the centers in the 2009–2010 school year when Eager and Able to Learn was implemented. Both groups were assessed at the end of their school year.

This eligible study received a low rating because of a time confounding factor. The comparison and intervention groups comprised children from adjacent cohorts. In line with WWC standards, this factor is considered confounding, as history is a threat to internal validity.

[57] Child care programs include a mix of private businesses from varying socioeconomic communities, while Sure Start programs target small socially disadvantaged communities. Both child care centers and Sure Start programs were in the intervention and comparison groups.

Table B.5. Eager and Able to Learn, extent of research

	Number of studies
Eligible for review	1
Publication date of eligible studies	
1960 to 1979	0
1980 to 1999	0
2000 or later	1
Rating of eligible studies	
Rated high	0
Rated moderate	0
Rated low	1

Notes: Possible ratings are as follows: **favorable effects** (at least one high- or moderate-rated study shows at least one significant or substantial favorable effect, no high-rated study shows any significant or substantial favorable effects, and no high- or moderate-rated study shows any significant or substantial unfavorable effects); **mixed effects** (at least one high-or moderate rated study shows at least one significant or substantial favorable effect and at least one high- or moderate-rated study shows at least one significant or substantial unfavorable effect); **no discernible effects** (no high- or moderate-rated study shows any significant or substantial effects, either favorable or unfavorable); and **unfavorable effects** (at least one high-or moderate-rated study shows at least one significant or substantial unfavorable effect, no high-rated study shows any significant or substantial unfavorable effects, and no high- or moderate-rated study shows any significant or substantial favorable effects). **End of intervention effects** are measured at 36 months and/or at the end of the intervention, if the intervention extends beyond 36 months. **Sustained or delayed effects** are measured one year or more after the end of the intervention. **Not applicable** indicates that no outcomes were measured that rated at least moderate in a study that met eligibility criteria.

Information from this table is summarized in the main report in Table III.6. Extent of research on LITES Direct and indirect enhancement models.

LEARNING LANGUAGE AND LOVING IT

Description

Learning Language and Loving It is an in-service group-training program for caregivers designed to improve children's language development. The program is intended to be delivered by a trained facilitator to a group of caregivers over 15 to 20 hours. Each educator also receives four to six individual video recording and feedback sessions. Finally, the program provides additional resources to learn about language strategies. The program includes the Learning Language and Loving It guidebook, which contains language development strategies to use in the classroom. The strategies are targeted to all developmental levels from toddler through preschool. The study eligible for review evaluated professional development strategies derived from the full Learning Language and Loving It program.

We identified one Learning Language and Loving It study as eligible for review (Ahrens, 2009). Six centers that served toddlers in six regions of a Midwestern metropolitan city were selected to participate in the study. Each center was randomly assigned to either the coaching and direct modeling group (N = 3) or the in-service training group (N = 3). Direct modeling involved the researcher interacting with children to demonstrate how to enhance language abilities during three 30-minute sessions. The behaviors modeled were based on strategies from Learning Language and Loving It. The researcher also coached the caregiver reflecting on experiences interacting with the children and answering the lead teacher's questions. The in-service training was a 90-minute session on enhancing children's language abilities in the classroom, based on Learning Language and Loving It. The child care director or lead teacher from each center selected four 18- to 36-month-old toddlers to be assessed for the study. One child withdrew from the experimental center, following pretest data collection, resulting in the total analysis sample of 23 children. Follow-up data collection occurred two weeks after implementation.

This eligible study received a low rating. It was a MCGD study that did not establish baseline equivalence on race/ethnicity, one of the measures required for MCGDs to receive a moderate rating.

Table B.6. Learning Language and Loving It, extent of research

	Number of studies
Eligible for review	1
Publication date of eligible studies	
1960 to 1979	0
1980 to 1999	0
2000 or later	1
Rating of eligible studies	
Rated high	0
Rated moderate	0
Rated low	1

Notes: Possible ratings are as follows: **favorable effects** (at least one high- or moderate-rated study shows at least one significant or substantial favorable effect, no high-rated study shows any significant or substantial favorable effects, and no high- or moderate-rated study shows any significant or substantial unfavorable effects); **mixed effects** (at least one high-or moderate rated study shows at least one significant or substantial favorable effect and at least one high- or moderate-rated study shows at least one significant or substantial unfavorable effect); **no discernible effects** (no high- or moderate-rated study shows any significant or substantial effects, either favorable or unfavorable); and **unfavorable effects** (at least one high-or moderate-rated study shows at least one significant or substantial unfavorable effect, no high-rated study shows any significant or substantial unfavorable effects, and no high- or moderate-rated study shows any significant or substantial favorable effects). **End of intervention effects** are measured at 36 months and/or at the end of the intervention, if the intervention extends beyond 36 months. **Sustained or delayed effects** are measured one year or more after the end of the intervention. **Not applicable** indicates that no outcomes were measured that rated at least moderate in a study that met eligibility criteria.

Information from this table is summarized in the main report in Table III.6. Extent of research on LITES direct and indirect enhancement models.

LEARNINGGAMES

Description

LearningGames is an intervention designed to promote children's cognitive, language, motor, and social development. It is composed of five components: (1) LearningGames activities, (2) support for enriched caregiving, (3) conversational books, (4) strategies to support language development, and (5) organizational plans and records. The LearningGames activities consist of 200 games divided into five volumes by age range for children from birth to age 5. The games are designed to support specific development topics such as social-emotional, early literacy, oral language, cognitive, and space and action. Child care providers are trained in implementing appropriate games. Providers also incorporate "enriched caregiving" during routine care and transition activities. Examples include naming food items during meals, singing songs, or going for a walk. Providers receive read-aloud books and are required to read at least one book to each child every day. Caregivers incorporate two language development strategies into the day. These strategies build on children's language abilities. Finally, caregivers receive planning and record-keeping materials to make weekly plans for each child and track progress.

We identified one LearningGames study as eligible for review (Collins et al., 2010). The study was a clustered RCT with family child care providers. Three-hundred-and-fifty-three family child care homes were randomly assigned to either the intervention group (N = 173 homes) or the comparison group (N = 180 homes) within 22 family child care agencies. LearningGames training was provided between 2006 and 2008. Child outcomes were assessed after two years of implementation. To be eligible for assessments, children had to have been with the family child care provider for at least six months. One-hundred-twenty-one children ages 6 months to 72 months participated in the follow-up assessment, 59 from the intervention group, and 62 from the comparison group.

This eligible study received a low rating. It was an RCT with high attrition that did not establish baseline equivalence on any of the required measures. Authors reported that child-level baseline data were not available.

Table B.7. LearningGames, extent of research

	Number of studies
Eligible for review	1
Publication date of eligible studies	
1960 to 1979	0
1980 to 1999	0
2000 or later	1
Rating of eligible studies	
Rated high	0
Rated moderate	0
Rated low	1

Notes: Possible ratings are as follows: **favorable effects** (at least one high- or moderate-rated study shows at least one significant or substantial favorable effect, no high-rated study shows any significant or substantial favorable effects, and no high- or moderate-rated study shows any significant or substantial unfavorable effects); **mixed effects** (at least one high-or moderate rated study shows at least one significant or substantial favorable effect and at least one high- or moderate-rated study shows at least one significant or substantial unfavorable effect); **no discernible effects** (no high- or moderate-rated study shows any significant or substantial effects, either favorable or unfavorable); and **unfavorable effects** (at least one high-or moderate-rated study shows at least one significant or substantial unfavorable effect, no high-rated study shows any significant or substantial unfavorable effects, and no high- or moderate-rated study shows any significant or substantial favorable effects). **End of intervention effects** are measured at 36 months and/or at the end of the intervention, if the intervention extends beyond 36 months. **Sustained or delayed effects** are measured one year or more after the end of the intervention. **Not applicable** indicates that no outcomes were measured that rated at least moderate in a study that met eligibility criteria.

Information from this table is summarized in the main report in Table III.6. Extent of research on LITES direct and indirect enhancement interventions.

MUSIC EDUCATION

Description

Music Education was a 15-month intervention in which children were exposed to a musically enriched setting. Once a week, children received a 30- to 40-minute group music lesson. The music lesson was informal and led by a teacher and assistant. Children were exposed to songs, tunes, chants, rhymes, and tonal and rhythm patterns. The program aimed to establish musical aptitudes for singing, moving, and listening, and to offer models for bodily imitation.

We identified one Music Education study as eligible for review (Gruhn, 2002). The study, which began in 1998, was a MCGD with interim assessments and an immediate post-test. The study measured children's attention, imitation, coordination, and vocal patterns among other outcomes. Twelve children ages 1 to 2 years were recruited from an urban, upper middle class area in Freiburg, Germany, to serve as the intervention group. Three children dropped out of the program; therefore, there were nine children in the analytic sample. The comparison group was composed of nine children recruited from a local nursery school. Children in the comparison group did not receive any specialized music education.

This eligible study received a low rating. It was an MCGD that did not establish baseline equivalence on any of the required measures.

Table B.8. Music Education, extent of research

	Number of studies
Eligible for review	1
Publication date of eligible studies	
1960 to 1979	0
1980 to 1999	0
2000 or later	1
Rating of eligible studies	
Rated high	0
Rated moderate	0
Rated low	1

Notes: Possible ratings are as follows: **favorable effects** (at least one high- or moderate-rated study shows at least one significant or substantial favorable effect, no high-rated study shows any significant or substantial favorable effects, and no high- or moderate-rated study shows any significant or substantial unfavorable effects); **mixed effects** (at least one high-or moderate rated study shows at least one significant or substantial favorable effect and at least one high- or moderate-rated study shows at least one significant or substantial unfavorable effect); **no discernible effects** (no high- or moderate-rated study shows any significant or substantial effects, either favorable or unfavorable); and **unfavorable effects** (at least one high-or moderate-rated study shows at least one significant or substantial unfavorable effect, no high-rated study shows any significant or substantial unfavorable effects, and no high- or moderate-rated study shows any significant or substantial favorable effects). **End of intervention effects** are measured at 36 months and/or at the end of the intervention, if the intervention extends beyond 36 months. **Sustained or delayed effects** are measured one year or more after the end of the intervention. **Not applicable** indicates that no outcomes were measured that rated at least moderate in a study that met eligibility criteria.

Information from this table is summarized in the main report in Table III.6. Extent of research on LITES direct and indirect enhancement models.

PROJECT SECURE CHILD IN CHILD CARE

Description

Project Secure Child in Child Care was an emotional availability (EA) intervention for early childhood caregivers; which targeted children's social-emotional development. The intervention consisted of an informational session and a practice component. Caregivers received two one-hour informational sessions in a group format. During the informational sessions, caregivers learned about EA, its links to attachment, and children's different attachment styles. For the practice component, EA coaches visited the centers three to four times over two to three months to hone the skills developed in the informational sessions. The EA coach used an EA checklist and narrative to note the strengths and weaknesses in the caregivers' EA skills. Also, during an EA coach visit the caregiver watched a video of interactions and the coach and caregiver discussed ways to improve the interactions.

We identified one Project Secure Child in Child Care study as eligible for review (Biringen et al., 2012). The study was a matched group comparison design (MCGD) with immediate posttest assessments. Twenty-one sites were recruited to participate in the study—10 received Project Secure Child in Child Care and 11 were in the comparison group. The comparison group sites did not receive any intervention. Within each site, dyadic child-caregiver pairs were observed, totaling 33 intervention group pairs and 24 comparison group pairs. Children were ages 11 to 32 months at the start of the study.

This eligible study received a low rating. It did not establish baseline equivalence on child age, one of the measures required for MCGDs to receive a moderate rating. Authors reported a statistically significant difference between the intervention and comparison groups on child age at baseline.

Table B.9. Project Secure Child in Child Care, extent of research

	Number of studies
Eligible for review	1
Publication date of eligible studies	
1960 to 1979	0
1980 to 1999	0
2000 or later	1
Rating of eligible studies	
Rated high	0
Rated moderate	0
Rated low	1

Notes: Possible ratings are as follows: **favorable effects** (at least one high- or moderate-rated study shows at least one significant or substantial favorable effect, no high-rated study shows any significant or substantial favorable effects, and no high- or moderate-rated study shows any significant or substantial unfavorable effects); **mixed effects** (at least one high-or moderate rated study shows at least one significant or substantial favorable effect and at least one high- or moderate-rated study shows at least one significant or substantial unfavorable effect); **no discernible effects** (no high- or moderate-rated study shows any significant or substantial effects, either favorable or unfavorable); and **unfavorable effects** (at least one high-or moderate-rated study shows at least one significant or substantial unfavorable effect, no high-rated study shows any significant or substantial unfavorable effects, and no high- or moderate-rated study shows any significant or substantial favorable effects). **End of intervention effects** are measured at 36 months and/or at the end of the intervention, if the intervention extends beyond 36 months. **Sustained or delayed effects** are measured one year or more after the end of the intervention. **Not applicable** indicates that no outcomes were measured that rated at least moderate in a study that met eligibility criteria.

Information from this table is summarized in the main report in Table III.6. Extent of research on LITES direct and indirect enhancement models.

APPENDIX C

SCREENING DISPOSITION TABLES

This page has been left blank for double-sided copying.

APPENDIX C. SCREENING DISPOSITIONS FOR STUDIES

Dispositions for studies for models are included in this appendix. Citations for studies are included in Appendix D.

Table C.1. Screening dispositions for direct multicomponent models

Screening disposition	Number of studies								
	Abecedarian	Brookline Early Education Project	Early Head Start	Infant Health and Development Program	Milwaukee Infant Stimulation Project	Parent Child Development Centers	Project CARE	Sure Start Local Programmes	Yale Child Welfare Research Program
Total studies identified	32	7	58	44	6	8	6	8	3
Screened in	10	4	5	11	3	6	2	2	1
Ineligible for review	4	1	4	1	0	0	1	0	0
Handbook or conference proceedings	0	0	0	0	0	0	0	0	0
Could not obtain full text	1	1	1	0	0	0	1	0	0
Supplemental materials	2	0	3	1	0	0	0	0	0
Study not most recent/ complete version available	1	0	0	0	0	0	0	0	0
Screened out	18	2	49	35	3	2	3	6	2
Additional source	4	0	2	6	2	1	0	0	2
Home visiting is primary program element	0	0	3	0	0	0	0	0	0
Ineligible study design	5	2	19	11	0	0	0	5	0
International and not policy relevant	0	0	0	0	0	0	0	0	0
Not possible to attribute effects solely to the model of interest	3	0	6	0	0	0	3	0	0
No eligible outcomes	1	0	17	4	0	1	0	3	0
Non-English	0	0	0	0	0	0	0	0	0
Not a primary study (e.g., a literature review)	6	0	15	7	1	0	0	1	0
Study does not examine a replicable program	0	0	0	0	0	0	0	0	0
Publication date out of range	0	0	0	0	0	0	0	0	0
Study on services not relevant to review	0	0	5	4	0	0	0	0	0
Subgroups out of scope	1	0	0	2	0	0	0	0	0
Target population out of range	0	0	0	0	0	0	0	0	0

Note: Some studies screened out for multiple reasons.

Table C.2. Screening dispositions for direct and indirect enhancement models

Screening disposition	Number of studies					
	Direct enhancement models		Indirect enhancement models			
	LearningGames	Music Education	Eager and Able to Learn	Learning Language and Loving It	Project Secure Child in Child Care	Program for Infant/Toddler Care
Total studies identified	1	5	1	6	1	1
Screened in	1	1	1	1	1	1
Ineligible for review	0	0	0	0	0	0
Handbook or conference proceedings	0	0	0	0	0	0
Could not obtain full text	0	0	0	0	0	0
Supplemental materials	0	0	0	0	0	0
Study not most recent/ complete version available	0	0	0	0	0	0
Screened out	0	4	0	5	0	0
Additional source	0	1	0	0	0	0
Home visiting is primary program element	0	0	0	0	0	0
Ineligible study design	0	0	0	1	0	0
International and not policy relevant	0	0	0	0	0	0
Not possible to attribute effects solely to the model of interest	0	0	0	0	0	0
No eligible outcomes	0	0	0	0	0	0
Non-English	0	0	0	0	0	0
Not a primary study (e.g., a literature review)	0	0	0	1	0	0
Study does not examine a replicable program	0	0	0	0	0	0
Publication date out of range	0	0	0	0	0	0
Study on services not relevant to review	0	1	0	0	0	0
Subgroups out of scope	0	0	0	0	0	0
Target population out of range	0	2	0	3	0	0

Note: Some studies screened out for multiple reasons.

Table C.3. Screening dispositions for direct multicomponent models that were considered but later deemed ineligible

Screening disposition	Number of Studies Even Start
Total studies identified	13
Screened in	0
Ineligible for review	3
Handbook or conference proceedings	0
Could not obtain full text	1
Supplemental materials	2
Study is not the most recent and complete version available	0
Screened out	10
Additional source	0
Home visiting is the primary program element	3
Ineligible study design	2
International and not policy relevant	0
Intervention combined with another intervention	0
No eligible outcomes	1
Non-English	0
Not a primary study (for example, a literature review)	0
Study does not examine a replicable program	0
Publication date out of range	0
Study on services not relevant to review	0
Subgroups out of scope	0
Target population out of range	5

Note: Some studies screened out for multiple reasons.

Table C.4. Screening dispositions for direct and indirect enhancement models that were considered but later deemed ineligible

Screening disposition	Number of studies	
	Child Care Expulsion Prevention	Responsive Early Childhood Curriculum
Total studies identified	1	1
Screened in	0	0
Ineligible for review	0	0
Handbook or conference proceedings	0	0
Could not obtain full text	0	0
Supplemental materials	0	0
Study is not the most recent and complete version available	0	0
Screened out	1	1
Additional source	0	0
Home visiting is the primary program element	0	0
Ineligible study design	0	0
International and not policy relevant	0	0
Intervention combined with another intervention	0	0
No eligible outcomes	0	0
Non-English	0	0
Not a primary study (for example, a literature review)	0	0
Study does not examine a replicable program	0	0
Publication date out of range	0	0
Study on services not relevant to review	0	0
Subgroups out of scope	0	0
Target population out of range	1	1

Note: Some studies screened out for multiple reasons.

APPENDIX D

CITATION OF STUDIES

This page has been left blank for double-sided copying.

APPENDIX D. CITATION OF STUDIES

Citations in this appendix correspond to the screening dispositions in Appendix C.

DIRECT MULTICOMPONENT MODELS

Abecedarian

Screened in

High

Ramey, C. T., & Haskins, R. (1981). The modification of intelligence through early experience. *Intelligence, 5*(1), 5–19.

- *Study Rating Disposition*: RCT with low attrition and no reassignment

Moderate

Campbell, F. A., Helms, R., Sparling, J. J., & Ramey, C. T. (1998). Early-childhood programs and success in school: The Abecedarian study. In: Barness, W. S., Boocock, S. S., (Eds.), *Early Care and Education for Children in Poverty: Promises, Programs, and Long Term Results. SUNY Series, Youth Social Services, Schooling, and Public Policy/SUNY Series, Early Childhood Education: Inquiries and Insights.* Albany, NY: State University of New York Press, pp. 145–166.

- *Study Rating Disposition:* RCT with reassignment that establishes baseline equivalence and uses statistical controls

Campbell, F. A., Pungello, E. P., Burchinal, M., Kainz, K., Pan, Y., Wasik, B. H., Barbarin, O. A., Sparling, J. J., & Ramey, C. T. (2012, July). Adult outcomes as a function of an early childhood educational program: An Abecedarian Project follow-up. *Developmental Psychology, 48*(4), 1033.

- *Study Rating Disposition:* RCT with reassignment that establishes baseline equivalence and uses statistical controls

Campbell, F. A., & Ramey, C. T. (1995, winter). Cognitive and school outcomes for high-risk African-American students at middle adolescence: Positive effects of early intervention. *American Educational Research Journal, 32*(4), 743–772.

- *Study Rating Disposition:* RCT with reassignment that establishes baseline equivalence and uses statistical controls

Campbell, F. A., Ramey, C. T., Pungello, E. P., Sparling, J., & Miller-Johnson, S. (2002). Early childhood education: Young adult outcomes from the Abecedarian project. *Applied Developmental Science, 6*(1), 42–57.

- *Study Rating Disposition:* RCT with reassignment that establishes baseline equivalence and uses statistical controls

Muennig, P., Robertson, D., Johnson, G., Campbell, F., Pungello, E. P., & Neidell, M. (2011, March). The effect of an early education program on adult health: The Carolina Abecedarian project randomized controlled trial. *American Journal of Public Health*, *101*(3), 512–516.

- *Study Rating Disposition:* RCT with reassignment that establishes baseline equivalence and uses statistical controls

Low

Campbell, F. A., & Ramey, C. T. (1994). Effects of early intervention on intellectual and academic achievement: A follow-up study of children from low-income families. *Child Development*, *65*(2), 684–698.

- *Study Rating Disposition:* RCT with reassignment that does not establish baseline equivalence

Campbell, F. A., Pungello, E. P., Miller-Johnson, S., Burchinal, M., & Ramey, C. T. (2001, March). The development of cognitive and academic abilities: Growth curves from an early childhood educational experiment. *Developmental Psychology*, *37*(2), 231–242.

- *Study Rating Disposition:* RCT with reassignment that does not use statistical controls

Horacek, H. J., Ramey, C. T., Campbell, F. A., Hoffman, K. P., & Fletcher, R. H. (1987). Predicting school failure and assessing early interventions with high-risk children. *Journal of the American Academy of Child & Adolescent Psychiatry*, *26*(5), 758.

- *Study Rating Disposition:* RCT with reassignment that does not use statistical controls

Ramey, C. T., & Campbell, F. A. (1984). Preventive education for high-risk children: Cognitive consequences of the Carolina Abecedarian project. *American Journal of Mental Deficiency*, *88*(5), 515–523.

- *Study Rating Disposition:* RCT with reassignment that does not use statistical controls

Ineligible for review

Campbell, F. A., Conti, G., Heckman, J. J., Moon, S. H., Pinto, R., Pungello, E. P., & Pan, Y. (2014). Early childhood investments substantially boost adult health. *Science, 343*(6178), 1478–1485. doi:10.1126/science.1248429

Campbell, F. A. (2000). Early learning, later success: The Abecedarian study. Early childhood educational intervention for poor children. Executive summary. Chapel Hill, NC: Frank Porter Graham Child Development Center.

Campbell, F. A., & Ramey, C. T. (2010). Carolina Abecedarian project. In A. Reynolds, A. Rolnick, & J. Temple (Eds.), *Childhood programs and practices in the first decade of life: A human capital integration* (pp. 76-98). New York: Cambridge University Press.

Campbell, F. A., Gunn, E., & Pungello, E. P. (2014). Carolina Abecedarian project and the Carolina approach to responsive education (CARE), age 21 follow up study. *Inter-university Consortium for Political and Social Research.*

Ramey, C. T., Gallagher, J. J., Campbell, F. A., Wasik, B. H., & Sparling, J. (2004). Carolina Abecedarian project and the Carolina approach to responsive education (CARE), 1972-1992. *Inter-university Consortium for Political and Social Research.*

Screened out

Burchinal, M., Lee, M., & Ramey, C. (1989). Type of day-care and preschool intellectual development in disadvantaged children. *Child Development, 60*(1), 128–137.

Burchinal, M. R., Campbell, F. A., Bryant, D. M., Wasik, B. H., & Ramey, C. T. (1997, October). Early intervention and mediating processes in cognitive performance of children of low-income African American families. *Child Development, 68*(5), 935–954.

Campbell, F. A., & Pungello, E. (2000). *High quality child care has long-term educational benefits for poor children.* Paper presented at the Head Start National Research Conference, Washington DC.

Campbell, F. A., Wasik, B. H., Pungello, E. P., Burchinal, M., Barbarin, O., Kainz, K., and Ramey, C. T. (2008). Young adult outcomes of the Abecedarian and CARE early childhood educational interventions. *Early Childhood Research Quarterly, 23*(4), 452–466.

Clarke, S. H., & Campbell, F. A. (1998). Can intervention early prevent crime later?: The Abecedarian project compared with other programs. *Early Childhood Research Quarterly, 13*(2), 319–343.

Finkelstein, N. W. (1982). Aggression: Is it stimulated by day care? *Young Children, 37*(6), 3–9.

Park, B. (2008). The earlier, the better: Early intervention programs for infants and toddlers at risk. *Dimensions of Early Childhood, 36*(1), 3–7.

Pungello, E. P., Campbell, F. A., & Barnett, W. S. (2006). *Poverty and early childhood educational intervention.* (Policy brief) Center on Poverty, Work and Opportunity Policy Brief Series. Retrieved from: http://www.law.unc.edu/documents/poverty/publications/pungelloandcampbellpolicybrief.pdf

Ramey, C. T. (1977). Social and intellectual consequences of daycare for high-risk infants. In R. Webb (Ed.), Social development in childhood: Day-care programs and research from the Hyman Blumberg Symposium on Research in Early Childhood Education (pp. 79-110). Baltimore: Johns Hopkins University Press.

Ramey, C. T., & Campbell, F. A. (1982). Compensatory education for disadvantaged children. In J. Belsky (Ed.), *In the beginning: Readings in infancy* (pp. 259-269). New York: Columbia University Press.

Ramey, C. T., & Campbell, F. A. (1987). The Carolina Abecedarian project: An educational experiment concerning human malleability. In J. J. Gallagher (Ed.), *The malleability of children* (pp. 127-139). Baltimore: Paul H. Brookes Publishing Co.

Ramey, C. T., & Campbell, F. A. (1992). Poverty, early childhood education and academic competence: The Abecedarian experiment. In A. C. Huston (Ed.), *Children in Poverty: Child development and public policy* (pp. 190-221). New York: Cambridge University Press.

Ramey, C. T., Campbell, F. A., & Blair, C. (1998). Enhancing the life course for high-risk children: Results from the Abecedarian project. In *Social programs that really work* (pp. 163-183). New York: Russell Sage Foundation.

Ramey, C. T., Campbell, F. A., Bryant, D. M., Burchinal, M., Sparling, J., & Wasik, B. H. (1992). Early intervention and long-term predictors of school status. In *New directions in child and family research: Shaping Head Start in the 90s: The First Head Start National Research Conference, June 24-26, 1991: Summary of conference proceedings* (pp. 170-174). Washington, DC: U.S. Administration on Children, Youth, and Families.

Ramey, C. T., Campbell, F. A., Burchinal, M., Skinner, M. L., Gardner, D. M., & Ramey, S. L. (2000). Persistent effects of early childhood education on high-risk children and their mothers. *Applied Developmental Science, 4*(1), 2–14.

Ramey, C. T., McGinness, G. D., Cross, L., Collier, A. M., & Barrie-Blackley, S. (1982). The Abecedarians approach to social competence: Cognitive and linguistic intervention for disadvantaged preschoolers. In K. Borman (Ed.), *The social life of children in a changing society* (pp. 145-174). Hillsdale, NJ: Lawrence Erlbaum Associates.

Ramey, C. T., & Ramey, S. L. (2004). Early learning and school readiness: Can early intervention make a difference? *Merrill-Palmer Quarterly, 50*(4), 471–491. doi:10.1353/mpq.2004.0034.

Brookline Early Education Project (BEEP)

Screened in

Low

Bronson, M. B., Pierson, D. E., & Tivnan, T. (1984, October). The effects of early education on children's competence in elementary school. *Evaluation Review*, *8*(5), 615–629.

- *Study Rating Disposition:* MCGD that does not establish baseline equivalence

Palfrey, J. S., Hauser-Cram, P., Bronson, M. B., Warfield, M. E., Sirin, S., & Chan, E. (2005). The Brookline Early Education Project: A 25-year follow-up study of a family-centered early health and development intervention. *Pediatrics*, *116*(1), 144–152.

- *Study Rating Disposition:* MCGD that does not establish baseline equivalence

Palfrey, J. S., Walker, D. K., Sullivan, M., & Levine, M. D. (1987). Targeted early childhood programming. The promise half fulfilled. *American Journal of Diseases of Children*, *141*(1), 55–59.

- *Study Rating Disposition:* MCGD that does not establish baseline equivalence

Tivnan, T., & Pierson, D. E. (1982). *Evaluation of a school-based early education program: Results from the Brookline Early Education Project.* Paper presented at the Annual Meeting of the American Educational Research Association. New York, NY.

- *Study Rating Disposition:* MCGD that does not establish baseline equivalence

Ineligible for review

Weiss, H. M. B. (1979). *Parent support and education: An analysis of the Brookline Early Education Project.* (Doctoral dissertation). ProQuest Dissertations and Theses. (302915331)

Screened out

Levine, M. D., Palfrey, J. S., Lamb, G. A., Weisberg, H. I., & Bryk, A. S. (1977). Infants in a public school system: The indicators of early health and educational need. *Pediatrics*, *60*(4), 579.

Levine, M. D., Wolman, R., Oberklaid, F., & Pierson, D. E. (1982). The longitudinal study of findings in childhood. Analysis of an interdisciplinary process. *American Journal of Diseases of Children*, *136*(4), 303–309.

Early Head Start (EHS)

Screened in

High

Chazan-Cohen, R., Raikes, H.H., & Vogel, C. (2013). Program subgroups: Patterns of impacts for home-based, center-based, and mixed-approach programs. In J. M. Love, R. Chazan-Cohen, H. Raikes, & J. Brooks-Gunn (Eds.), *What Makes a Difference: Early Head Start Evaluation Findings in a Developmental Context*. Boston: Wiley.

- *Study Rating Disposition:* high rated outcomes had the disposition RCT with low attrition and no reassignment; moderate rated outcomes had the disposition RCT with high attrition that establishes baseline equivalence and uses statistical controls

U.S. Administration for Children and Families. (2001). *Building their futures: How Early Head Start programs are enhancing the lives of infants and toddlers in low-income families.* Washington, DC: U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research and Evaluation.

- *Study Rating Disposition:* Center-based programs: moderate rated outcomes had the disposition RCT with high attrition that establishes baseline equivalence and uses statistical controls; Mixed-approach programs: high rated outcomes had the disposition RCT with low attrition and no reassignment

U.S. Administration for Children and Families. (2002). Making a difference in the lives of infants and toddlers and their families: The impacts of Early Head Start: Vol. I. Final technical report. Washington, DC: U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research and Evaluation.

- *Study Rating Disposition:* Center-based programs: moderate rated outcomes had the disposition RCT with high attrition that establishes baseline equivalence and uses statistical controls; Mixed-approach programs: high rated outcomes had the disposition RCT with low attrition and no reassignment

Moderate

C. A. Vogel, Y. Xue, E. M. Moiduddin, B. L. Carlson, E. E. Kisker (2010). Early Head Start children in grade 5: Long-term follow-up of the Early Head Start Research and Evaluation Study sample. OPRE Report #2011-8. Washington, DC: U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research and Evaluation.

- *Study Rating Disposition:* RCT with high attrition that establishes baseline equivalence and uses statistical controls

Low

Whiteside-Mansell, L., Bradley, R., McKelvey, L., & Lopez, M. (2009, November). Center-based Early Head Start and children exposed to family conflict. *Early Education and Development*, *20*(6), 942–957.

- *Study Rating Disposition*: RCT with high attrition that does not use statistical controls

Ineligible for review

Abstract. (2013). *Monographs of the Society for Research in Child Development, 78*(1), vii–viii. doi:10.1111/j.1540-5834.2012.00699.x.

Love, J. M., Kisker, E. E., Ross, C. M., Schochet, P. Z., Paulsell, D., Boller, K., et al. (2002). *Making a difference in the lives of infants and toddlers and their families: The impacts of Early Head Start: Executive summary*. Washington, DC: U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research and Evaluation.

Love, J. M., Kisker, E. E., Ross, C. M., Schochet, P. Z., Paulsell, D., Boller, K., et al. (2002). *Making a difference in the lives of infants and toddlers and their families: The impacts of Early Head Start: Vol. II. Final technical report append*ixes. Washington, DC: U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research and Evaluation.

U.S. Administration for Children and Families. *Early Head Start Research and Evaluation Project.* United States. Department of Health and Human Services.

Screened out

Ayoub, C., Rappolt-Schlichtmann, G., Vallotton, C., Raikes, H., Chazan-Cohen, R., & O'Connor, E. (2009). Cognitive skill performance among young children living in poverty: Risk, change, and the promotive effects of Early Head Start. *Early Childhood Research Quarterly*, *24*(3), 289-305.

Ayoub, C., Vallotton, C. D., & Mastergeorge, A. M. (2011). Developmental pathways to integrated social skills: The roles of parenting and early intervention. *Child Development, 82*(2), 583–600.

Beerer, L. S., Chazan-Cohen, R., Squires, J., Harden, B. J., Boris, N. W., Heller, S. S., & Malik, N. M. (2007). The early promotion and intervention research consortium (E-PIRC): Five approaches to improving infant/toddler mental health in Early Head Start. *Infant Mental Health Journal, 28*(2), 130–150.

Bohlander, A. (2011). The influence of poverty correlates on cognitive ability of toddlers in the Early Head Start program. *Dissertation Abstracts International: Section A. Humanities and Social Sciences, 71*(12-A), 4280.

Brophy-Herb, H., Schiffman, R., McKelvey, L., Cunningham-DeLuca, M., & Hawver, M. (2001). Innovations in practice. Quality improvement: Lessons learned from an infant mental health-based Early Head Start program. *Infants & Young Children: An Interdisciplinary Journal of Special Care Practices*, *14*(2), 77–85.

Brophy-Herb, H., Zajicek-Farber, M., Bocknek, E. L., McKelvey, L. M., & Stansbury, K. (2013). Longitudinal connections of maternal supportiveness and early emotion regulation to children's school readiness in low-income families. *Journal of the Society for Social Work & Research*, *4*(1), 2–19.

Buell, M. J., Pfister, I., & Gamel-McCormick, M. (2002). Caring for the caregiver: Early head Start/family child care partnerships. *Infant Mental Health Journal, 23*(1–2), 213–230.

Buhrmann, J., Eiserman, W., & Shisler, L. (2004). *Assessing the effectiveness of hearing screening in Early Head Start programs.* American Sociological Association. Paper presented at The Annual Meeting of the American Sociological Association, San Francisco. doi: asa_proceeding_35571.PDF.

Center for Law and Social Policy, & Zero to Three. (2012). *Missouri: Early Head Start initiative.* Washington, DC: Center for Law and Social Policy. Retrieved from http://www.clasp.org/admin/site/publications/files/MO-EHS.pdf

Chapin, L. A., & Altenhofen, S. (2010). Neurocognitive perspectives in language outcomes of Early Head Start: Language and cognitive stimulation and maternal depression. *Infant Mental Health Journal*, *31*(5), 486–498.

Chazan-Cohen, R., Ayoub, C., Pan, B. A., Roggman, L., Raikes, H., McKelvey, L., Whiteside-Mansell, L., Hart, A. (2007). It takes time: Impacts of Early Head Start that lead to reductions in maternal depression two years later. *Infant Mental Health Journal*, *28*(2), 151–170.

Chazan-Cohen, R., Stark, D. R., Mann, T., & Fitzgerald, H. (2007). Early Head Start and infant mental health. *Infant Mental Health Journal, 28*(2), 99–105. doi:10.1002/imhj.20124

Child Trends. (2010). Early Head Start: Research findings. *Early Childhood Highlights, 1*(2).

Cline, K. D. (2010). *The instructional and emotional quality of parent-child book reading and Early Head Start children's learning outcomes.* (Doctoral dissertation). ProQuest Dissertations and Theses. (518634230)

Early Head Start National Resource Center (2003). *Responding to the mental health needs of infants, toddlers and families.* Washington, DC: Early Head Start National Resource Center.

Early Head Start National Resource Center (2004). *Transition strategies: Continuity and change in the lives of infants and toddlers.* Washington, DC: Early Head Start National Resource Center.

Ezell, S. (2013). *Early Head Start, curriculum intervention, and outcome measures.* (Doctoral dissertation). ProQuest Dissertations and Theses. (1372810039)

Famakinwa, O. (2012). *A program evaluation of Early Head Start health services in family child care homes.* (Doctoral dissertation). ProQuest Dissertations and Theses. (1102745357)

Greenwood, C. R., Walker, D., & Buzhardt, J. (2010). The early communication indicator for infants and toddlers: Early Head Start growth norms from two states. *Journal of Early Intervention, 32*(5), 310-334.

Illmer-Craciun, D. (2009). *The relationships between home support for language and emergent literacy in low-income families, mother's education and immigrant status, and children's language and emergent literacy development at kindergarten entry.* (Unpublished doctoral dissertation), Washington, DC: Catholic University.

Jameson, J. S. (2009). Attainment of parenting goals in an Early Head Start program. *Dissertation Abstracts International: Section A. Humanities and Social Sciences, 70*(2-A), 466.

Kisker, E. E., Paulsell, D., Love, J. M., & Raikes, H. (2002). *Early Head Start research: Pathways to quality and full implementation in Early Head Start programs*. Administration for Children, Youth, and Families (DHHS), Washington, DC. Head Start Bureau.

Kisker, E. E., Love, J. M., Raikes, H. H., Boller, K., Paulsell, D., Rosenberg, L. et al. (1999). *Leading the way: Characteristics and early experiences of selected Early Head Start programs: Volume I: Cross-site perspectives.* Report submitted to the Administration on Children, Youth and Families. Princeton, NJ: Mathematica Policy Research.

Kisker, E. E., Love, J. M., Raikes, H. H., Boller, K., Paulsell, D., Rosenberg, L. et al. (1999). *Leading the way: Characteristics and early experiences of selected Early Head Start programs: Volume II: Program profiles.* Report submitted to the Administration on Children, Youth and Families. Princeton, NJ: Mathematica Policy Research.

Lee, R., Zhai, F., Brooks-Gunn, J., Han, W., & Waldfogel, J. (2014). Head start participation and school readiness: Evidence from the early childhood longitudinal study-birth cohort. *Developmental Psychology, 50*(1), 202–215.

Lombardi, J., & Bogle, M. M. (2005). *Beacon of hope: The promise of Early Head Start for America's youngest children.* Washington D.C.: *ZERO TO THREE*.

Love, J. M., (2010). *Effects of Early Head Start prior to kindergarten entry: The importance of early experience.* Paper presented at the early childhood invited symposium, Society for Research on Educational Effectiveness.

Love, J. M., Kisker, E. E., Ross, C., Raikes, H., Constantine, J., Boller, K., et al. (2005, November). The effectiveness of Early Head Start for 3-year-old children and their parents: Lessons for policy and programs. *Developmental Psychology, 41*(6), 885–901.

Mann, T. L. (2002). The role of training and technical assistance in supporting the delivery of high quality services in Early Head Start. *Infant Mental Health Journal*, *23*(2), 36–47.

Mayer, L. M., & Blome, W. W. (2013). The importance of early, targeted intervention: The effect of family, maternal, and child characteristics on the use of physical discipline. *Journal of Human Behavior in the Social Environment*, *23*(2), 144–158.

McAllister, C. L., Green, B. L., Terry, M. A., Herman, V., & Mulvey, L. (2003). Parents, practitioners, and researchers: Community-based participatory research with Early Head Start. *American Journal of Public Health*, *93*(10), 1672–1679.

McAllister, C. L., & Thomas, T. (2007, March–April). Infant mental health and family support: Contributions of Early Head Start to an integrated model for community-based early childhood programs. *Infant Mental Health Journal*, *28*(2), 192–215.

Ontai, L. L., Hinrichs, S., Beard, M., & Wilcox, B. L. (2002). Improving child care quality in Early Head Start programs: A partnership model. *Infant Mental Health Journal*, *23*(1), 48–61.

Paulsell D., Cohen J., Stieglitz A., Lurie-Hurvitz E., Fenichel E., Kisker E. (2002). *Partnerships for quality: Improving infant-toddler child care for low-income families.* Princeton, NJ: Mathematica Policy Research.

Paulsell, D., Kisker, E. E., Love. J. M., Raikes, H., Boller, K., Rosenberg, L., et al. (2000). *Leading the way: Characteristics and early experiences of selected Early Head Start programs: Executive summary: Volumes I, II, III*. Report submitted to the Administration on Children, Youth and Families. Princeton, NJ: Mathematica Policy Research.

Paulsell, D., Kisker, E. E., Love, J. M., Raikes, H., Boller, K., Rosenberg, L., et al. (2000). *Leading the way: Characteristics and early experiences of selected Early Head Start programs. Volume III: Program implementation. Early Head Start research.* Report submitted to the Administration on Children, Youth and Families. Princeton, NJ: Mathematica Policy Research.

Poulakos, A. (2013). *Language development in context: Influences of perinatal risk and environmental characteristics on outcomes at 14 to 36 months.* (Doctoral dissertation). ProQuest Dissertations and Theses. (1467526239)

Raikes, H., & Love, J. M. (2002). Early Head Start: A dynamic new program for infants and toddlers and their families. *Infant Mental Health Journal*, *23*(1–2), 1–13. doi:10.1002/imhj.10000

Robinson, J., & Emde, R. N. (2004, January–March). Mental health moderators of Early Head Start on parenting and child development: Maternal depression and relationship attitudes. *Parenting: Science and Practice*, *4*(1), 73–97.

Robokos, D. (2007). Cognitive, language, and social-emotional development among infants and toddlers in Early Head Start: An examination of the impact of cumulative risk. Dissertation Abstracts International: Section B. Sciences and Engineering, 68(1-b), 652.

Schmit, S., Ewen, D., & Center for Law and Social Policy. (2012). Supporting our youngest children: Early Head Start programs in 2010. (Brief no. 11). Washington, DC: CLASP.

Schumacher, R., & DiLauro, E. (2008). Building on the promise: State initiatives to expand access to Early Head Start for young children and their families. Washington, DC: Center for Law and Social Policy.

Spieker, S., Nelson, E. M., & Condon, M. (2011). Validity of the TAS-45 as a measure of toddler-parent attachment: Preliminary evidence from Early Head Start families. Attachment and Human Development, 13(1), 69–90.

Tao, F., Swartz, J., St. Pierre, R., & Tarr, H. (1997). National evaluation of the Even Start family literacy program. 1995 interim report. Washington, DC: US Department of Education, Planning, and Evaluation Service.

U.S. Administration for Children and Families. (2004). The role of Early Head Start programs in addressing the child care needs of low-income families with infants and toddlers: Influences on child care use and quality. Washington, DC: U.S. Administration for Children and Families.

U.S. Administration for Children and Families, Office of Planning, Research, and Evaluation. (2011). Learning as we go: A first snapshot of early head start programs, staff, families, and children: Vol. I. First report. Washington, DC: U.S. Administration for Children and Families, Office of Planning, Research and Evaluation.

U.S. Office of Head Start, & U.S. Office of Child Care. (2011). Building partnerships between Early Head Start grantees and family child care providers: Lessons from the Early Head Start for Family Child Care project: Final report. Washington, DC: U.S. Office of Head Start.

Vallotton, C. D., Harewood, T., Ayoub, C. A., Pan, B., Mastergeorge, A. M., & Brophy-Herb, H. (2012). Buffering boys and boosting girls: The protective and promotive effects of Early Head Start for children's expressive language in the context of parenting stress. Early Childhood Research Quarterly, 27(4), 695–707.

Wacharasin, C. (2001). *Predicting child cognitive development in low-income families.* (Doctoral dissertation). Available from ProQuest Dissertations & Theses A&I. (304732375)

IHDP

Screened in

High

Brooks-Gunn, J., Klebanov, P. K., Liaw, F., & Spiker, D. (1993). Enhancing the development of low-birthweight, premature infants: Changes in cognition and behavior over the first three years. *Child Development, 64*(3), 736–753.

- *Study Rating Disposition:* high rated outcomes had the disposition RCT with low attrition and no reassignment; moderated rated outcomes had the disposition RCT with high attrition that establishes baseline equivalence and uses statistical controls

Casey, P. H., Bradley, R. H., Whiteside-Mansell, L., Barrett, K., Gossett, J. M., & Simpson, P. M. (2009). Effect of early intervention on 8-year growth status of low-birth-weight preterm infants. *Archives of Pediatrics and Adolescent Medicine, 163*(11), 1046–1053.

- *Study Rating Disposition:* high rated outcomes had the disposition RCT with low attrition and no reassignment (full sample); moderate rated outcomes had the disposition RCT wit®h high attrition that establishes baseline equivalence and uses statistical controls (subgroups)

Moderate

Brooks-Gunn, J., Gross, R. T., Kraemer, H. C., Spiker, D., & Shapiro, S. (1992). Enhancing the cognitive outcomes of low birth weight, premature infants: For whom is the intervention most effective? *Pediatrics, 89*(6), 1209–1215.

- *Study Rating Disposition:* RCT with high attrition that establishes baseline equivalence and uses statistical controls

Brooks-Gunn, J., Liaw, F. R., & Klebanov, P. K. (1992). Effects of early intervention on cognitive function of low birth weight preterm infants. *Journal of Pediatrics, 120*(3), 350–359.

- *Study Rating Disposition:* RCT with high attrition that establishes baseline equivalence and uses statistical controls

Brooks-Gunn, J., McCarton, C. M., Casey, P. H., McCormick, M. C., Bauer, C. R., Bernbaum, J. C., et al. (1994). Early intervention in low-birth-weight premature infants. Results through Age 5 years from the Infant Health and Development Program. *JAMA: The Journal of the American Medical Association, 272*(16), 1257–1262.

- Study *Rating Disposition:* RCT with high attrition that establishes baseline equivalence and uses statistical controls

Gross, R. T., Spiker, D., Constantine, N. A., Kreitman, W. L., Haynes, C. W., .Ramey, C. T. et al. (1990). Enhancing the outcomes of low-birth-weight, premature infants. A multisite, randomized trial. *JAMA: The Journal of the American Medical Association, 263*(22), 3035–3042.

- *Study Rating Disposition:* RCT with high attrition that establishes baseline equivalence and uses statistical controls

Liaw, F., & Brooks-Gunn, J. (1994, December). Cumulative familial risks and low-birthweight children's cognitive and behavioral development. *Journal of Clinical Child Psychology, 23*(4), 360–372.

- *Study Rating Disposition:* RCT with high attrition that establishes baseline equivalence and uses statistical controls

McCarton, C. M., Brooks-Gunn, J., Wallace, I. F., & Bauer, C. R. (1997, January). Results at age 8 years of early intervention for low-birth-weight premature infants: The Infant Health and Development Program. *JAMA: The Journal of the American Medical Association, 277*(2), 126–132.

- *Study Rating Disposition:* RCT with high attrition that establishes baseline equivalence and uses statistical controls

Spiker, D., Ferguson, J., & Brooks-Gunn, J. (1993). Enhancing maternal interactive behavior and child social competence in low birth weight, premature infants. *Child Development, 64*(3), 754–768.

- *Study Rating Disposition:* RCT with high attrition that establishes baseline equivalence and uses statistical controls

Low

Blair, C. (2002). Early intervention for low birth weight, preterm infants: The role of negative emotionality in the specification of effects. *Development and Psychopathology, 14*(2), 311–332.

- *Study Rating Disposition:* RCT with high attrition that does not use statistical controls

Hollomon, H. A., & Scott, K. G. (1998). Influence of birth weight on educational outcomes at age 9: The Miami site of the Infant Health and Development Program. *Journal of Developmental and Behavioral Pediatrics, 19*(6), 404–410.

- *Study Rating Disposition:* RCT with high attrition that does not establish baseline equivalence

Ineligible for review

Brooks-Gunn, J., McCarton, C. M., Casey, P. H., McCormick, M. C., Bauer, C. R., Bernbaum, J. C., et al. (1994). Early intervention in low-birth-weight premature infants. Results through age 5 years from the Infant Health and Development Program. *JAMA: The Journal of the American Medical Association, 272*(16), 1257–1262.

Spiker, D., Kraemer, H. C., Scott, D. T., & Gross, R. T. (1991). Design issues in a randomized clinical trial of a behavioral intervention: Insights from the Infant Health and Development Program. *Journal of Developmental and Behavioral Pediatrics, 12*(6), 386–393.

Screened out

Baumeister, A. A., & Bacharach, V. R. (2000). Early generic educational intervention has no enduring effects on intelligence and does not prevent mental retardation: The Infant Health and Development Program. *Intelligence, 28*(3), 161–192.

Berlin, L. J., Brooks-Gunn, J., McCarton, C., & McCormick, M. C. (1998). The effectiveness of early intervention: Examining risk factors and pathways to enhanced development. *Preventive Medicine, 27*(2), 238–245.

Berlin, L. J., O'Neal, C., & Brooks-Gunn, J. (1998). What makes early intervention programs work? The program, its participants, and their interaction. *Zero to Three, 18*(4), 4–15.

Blair, C., Ramey, C. T., & Hardin, J. M. (1995). Early intervention for low birthweight, premature infants: Participation and intellectual development. *American Journal of Mental Retardation, 99*(5), 542–554.

Bradley, R. H. (2000, October). Intervention & home environment. *NCEDL Spotlights*, no. 27. Chapel Hill, NC: National Center for Early Development and Learning.

Bradley, R. H., Burchinal, M. R., & Casey, P. H. (2001). Early intervention: The moderating role of the home environment. *Applied Developmental Science, 5*(1), 2–8.

Bradley, R. H., Mundfrom, D. J., Whiteside, L., Caldwell, B. M., Casey, P. H., Kirby, R. S., et al. (1994). A reexamination of the association between HOME scores and income. *Nursing Research, 43*(5), 260–266.

Bradley, R. H., Whiteside-Mansell, L., Casey, P. H., & Barrett, K. (2010). Impact of a two-generation early education program on parenting processes at age 18. *Journal of Family Psychology, 24*(4), 478–484.

Brooks-Gunn, J., McCormick, M. C., Shapiro, S., Benasich, A., & Black, G. W. (1994). The effects of early education intervention on maternal employment, public assistance, and health insurance: The Infant Health and Development Program. *American Journal of Public Health, 84*(6), 924–931.

Duncan, G. J., Brooks-Gunn, J., & Klebonov, P. K. (1993, January). *Economic deprivation and early-childhood development.* Paper presented at the Biennial Meeting of the Society for Research in Child Development, New Orleans.

Duncan, G. J., & Sojourner, A. J. (2013). Can intensive early childhood intervention programs eliminate income-based cognitive and achievement gaps? *Journal of Human Resources, 48*(4), 945–968.

Gross, R. T., & Hayes, C. (1991). Implementing a multi-site, multidisciplinary clinical trial: The Infant Health and Development Program. *Zero to Three*, *11*(4), 1–7.

Hill, J. L., Brooks-Gunn, J., & Waldfogel, J. (2003). Sustained effects of high participation in an early intervention for low-birth-weight premature infants. *Developmental Psychology*, *39*(4), 730–744.

Klebanov, P. K., & Brooks-Gunn, J. (2008). Differential exposure to early childhood education services and mother-toddler interaction. *Early Childhood Research Quarterly*, *23*(2), 213–232. doi:10.1016/j.ecresq.2007.12.001

Lee, K. (2005a). Effects of experimental center-based child care on developmental outcomes of young children living in poverty. *Social Service Review*, *79*(1), 158–180.

Lee, K. (2005b). Intervention effects on maternal concepts of development for children's cognitive outcomes. *Journal of Human Behavior in the Social Environment*, *11*(2), 77–95.

Leventhal, T. (1999, January). Poverty and turbulence: Familial and neighborhood influences on children's achievement. (cognition). *Dissertation Abstracts International: Section B. Sciences and Engineering, 60*(6-B), 2987.

Liaw, F., Meisels, S. J., & Brooks-Gunn, J. (1995). The effects of experience of early intervention on low birth weight, premature children: The Infant Health and Development Program. *Early Childhood Research Quarterly*, *10*(4), 405–431.

Martin, A., Brooks-Gunn, J., Klebanov, P., Buka, S. L., & McCormick, M. C. (2008). Long-term maternal effects of early childhood intervention: Findings from the Infant Health and Development Program (IHDP). *Journal of Applied Developmental Psychology*, *29*(2), 101–117.

McCarton, C. (1998). Behavioral outcomes in low birth weight infants. *Pediatrics*, *102*(5 suppl. E), 1293–1297.

McCarton, C. M., Wallace, I. F., & Bennett, F. C. (1995). Preventive interventions with low birth weight premature infants: An evaluation of their success. *Seminars in Perinatology*, *19*(4), 330–340.

McCarton, C. M., Wallace, I. F., & Bennett, F. C. (1996). Early intervention for low-birth-weight premature infants: What can we achieve? *Annals of Medicine*, *28*(3), 221–225.

McCormick, M. C., Brooks-Gunn, J., Buka, S. L., Goldman, J., Yu, J., Salganik, M., et al. (2006). Early intervention in low birth weight premature infants: Results at 18 years of age for the Infant Health and Development Program. *Pediatrics, 117*(3), 771–780.

McCormick, M. C., Brooks-Gunn, J., Shapiro, S., Benasich, A. A., Black, G., & Gross, R. T. (1991). Health care use among young children in day care. Results in a randomized trial of early intervention. *JAMA*, 265(17), 2212–2217.

McCormick, M. C., McCarton, C., Brooks-Gunn, J., Belt, P., & Gross, R. T. (1998, October). The Infant Health and Development Program: Interim summary. *Journal of Developmental and Behavioral Pediatrics, 19*(5), 359–370.

McCormick, M. C., McCarton, C., Tonascia, J., & Brooks-Gunn, J. (1993). Early educational intervention for very low birth weight infants: Results from the Infant Health and Development Program. *Journal of Pediatrics, 123*(4), 527–533.

National Forum on Early Childhood Program Evaluation. (2008). *Do early childhood programs have lasting effects on children?* (Evaluation Science Brief). Cambridge, MA. National Forum on Early Childhood Program Evaluation.

Ramey, C. T., Bryant, D. M., Wasik, B. H., Sparling, J. J., Fendt, K. H., & LaVange, L. M. Infant health and development program for low birth weight, premature infants: Program elements, family participation, and child intelligence. *Pediatrics, 89*(3), 454.

Tafti, J. H. (1999, July). Correlates of classroom behavior and competence in low birth weight children at age 6 1/2. *Dissertation Abstracts International: Section B. Sciences and Engineering, 60*(1-B), 0392.

Whiteside-Mansell, L., Bradley, R. H., Casey, P. H., Fussell, J. J., & Conners-Burrow, N. A. (2009). Triple risk: Do difficult temperament and family conflict increase the likelihood of behavioral maladjustment in children born low birth weight and preterm? *Journal of Pediatric Psychology 34*(4), 396–405.

Milwaukee Infant Stimulation Project

Screened in

Low

Garber, H. L. (1988). *The Milwaukee project: Preventing mental retardation in children at risk.* Washington, DC: American Association on Mental Retardation, 1988.

- *Study Rating Disposition:* RCT with high attrition that does not use statistical controls

Heber, R., et al. (1972). *Rehabilitation of families at risk for mental retardation: Progress report.* Madison, WI: University of Wisconsin, Rehabilitation Research and Training Center in Mental Retardation.

- *Study Rating Disposition:* RCT with reassignment that does not use statistical controls

McBride, J. A. (1989). Public school performance of disadvantaged children with extensive preschool experience: A study of the Milwaukee project children through high school. (Unpublished doctoral dissertation). Milwaukee, WI: University of Wisconsin.

- *Study Rating Disposition:* RCT with high attrition that does not use statistical controls

Screened out

Garber, H. L. (1979). Bridging the gap from preschool to school for the disadvantaged child. School Psychology Digest, 8(3), 303–310.

Garber, H., & Heber, R. (1973). The Milwaukee project: Early intervention as a technique to prevent mental retardation. Storrs CT: The University of Connecticut.

Heber, R., & Garber, H. (1975). The Milwaukee project: A study of the use of family intervention to prevent cultural-familial mental retardation. Exceptional Infant, 3, 399–433.

Parent Child Development Centers (PCDC)

Screened in

Moderate

Bridgeman, B. (1981). *Parent Child Development Center: Final evaluation report.* Washington D.C.: Department of Health and Human Services, Office of Human Development Services.

- *Study Rating Disposition:* RCT with high attrition that establishes baseline equivalence and uses statistical controls

Low

Andrews, S. R., Blumenthal, J. B., Johnson, D. L., Kahn, A. J., Ferguson, C. J., Lasater, T. M., et al. (1982). The skills of mothering: A study of parent child development centers. *Monographs of the Society for Research in Child Development*, 1–83.

- *Study Rating Disposition*: RCT with high attrition that does not establish baseline equivalence

Johnson, D. L., Kahn, A. J., & Leler, H. (1976). *Houston Parent-Child Development Center. Final report.* Washington D.C.: Office of Child Development.

- *Study Rating Disposition:* RCT with high attrition that does not establish baseline equivalence and does not use statistical controls

Johnson, D. L. (2006). Parent- Child Development Center follow-up project: Child behavior problem results. *Journal of Primary Prevention, 27*(4), 391–407.

- *Study Rating Disposition:* RCT with high attrition that does not establish baseline equivalence and does not use statistical controls

Johnson, D. L., & Breckenridge, J. N. (1982, June). The Houston Parent-Child Development Center and the primary prevention of behavior problems in young children. *American Journal of Community Psychology, 10*(3), 305–316.

- *Study Rating Disposition:* RCT with high attrition that does not establish baseline equivalence and does not use statistical controls

Johnson, D. L., & Walker, T. (1991). A follow-up evaluation of the Houston Parent-Child Development Center: School performance. *Journal of Early Intervention, 15*(3), 226–236.

- *Study Rating Disposition:* MCGD that does not establish baseline equivalence

Screened out

Johnson, D. L. (1976). *Parent education and the educationally disadvantaged child.* Paper presented at the Annual Meeting of the American Psychological Association. Washington, DC.

Johnson, D. L. (1976). *Measuring the learning environment of Mexican-American families in a parent education program.* Paper presented at the Annual Meeting of the American Educational Research Association. San Francisco, CA.

Project CARE

Screened in

Low

Ramey, C. T., Bryant, D. M., Sparling, J., & Wasik, B. H. (1985). Project CARE: A comparison of two early intervention strategies. *Topics in Early Childhood Special Education*, *5*(2), 12–25.

- *Study Rating Disposition:* RCT with high attrition that does not use statistical controls

Wasik, B. H., Ramey, C. T., Bryant, D. M., & Sparling, J. J. (1990). A longitudinal study of two early intervention strategies: Project CARE. *Child Development, 61*(6), 1682–1696.

- *Study Rating Disposition:* RCT with high attrition that does not use statistical controls

Ineligible for review

Campbell, F. A., Gunn, E., & Pungello, E. P. (2014). *Carolina Abecedarian project and the Carolina approach to responsive education (CARE), age 21 follow up study.* Inter-university Consortium for Political and Social Research. doi:10.3886/ICPSR32262.v1

Screened out

Burchinal, M., (1989). Type of day-care and preschool intellectual development in disadvantaged children. *Child Development, 60*(1), 128–137.

Burchinal, M. R., Campbell, F. A., Bryant, D. M., Wasik, B. H., & Ramey, C. T. (1997, October). Early intervention and mediating processes in cognitive performance of children of low-income African American families. *Child Development, 68*(5), 935–954.

Campbell, F. A., Wasik, B. H., Pungello, E. P., Burchinal, M., Barbarin, O., Kainz, K., Sparling, J. J., Ramey, C. T. (2008). Young adult outcomes of the Abecedarian and CARE early childhood educational interventions. *Early Childhood Research Quarterly, 23*(4), 452–466.

Sure Start local programmes (SSLPs)

Screened in

Low

Belsky, J., Melhuish, E., Barnes, J., Leyland, A. H., & Romaniuk, H. (2006). Effects of Sure Start Local Programmes on children and families: Early findings from a quasi-experimental, cross sectional study. *British Medical Journal*, *332*(7556), 1476.

- *Study Rating Disposition:* MCGD that does not establish baseline equivalence

Melhuish, E., Belsky, J., Leyland, A. H., Barnes, J., & National Evaluation of Sure Start Research Team. (2008). Effects of fully-established Sure Start Local Programmes on 3-year-old children and their families living in England: A quasi-experimental observational study. *Lancet*, *372*(9650), 1641–1647.

- *Study Rating Disposition:* MCGD that does not establish baseline equivalence

Screened out

Anning, A., Ball, M., Barnes, J., Belsky, J., Botting, B., Frost, M., et al. (2004). The national evaluation of Sure Start Local Programmes in England. *Child and Adolescent Mental Health*, *9*(1), 2–8.

Fuller, A. (2010). Speech and language therapy in Sure Start Local Programmes: A survey-based analysis of practice and innovation. *International Journal of Language & Communication Disorders*, *45*(2), 182–203.

Reading, R. (2006). The national evaluation of Sure Start research team effects of Sure Start Local Programmes on children and families: Early findings from a quasi-experimental, cross sectional study. *Child: Care, Health and Development*, *32*(6), 753–775.

Sure Start (Programme). (2003a). *Characteristics of* Sure Start Local Programme areas*: Rounds 1 to 4: Profiles of rounds 1 to 4 sure start local programme areas and a sample of round 5, to-be-sure start local programme areas [executive summary]*. (Report No. 03). Nottingham, United Kingdom: Great Britain, Department for Education and Skills.

Sure Start (Programme). (2003b). *National evaluation summary: Characteristics of Sure* Start Local Programme *areas: Rounds 1 to 4.* Nottingham, United Kingdom: Great Britain, Department for Education and Skills.

Sure Start (Programme). (2004). *National evaluation report: Characteristics of Sure Start Local Programme 2001/2: Interim report on characteristics of and changes in Sure Start local programme areas in rounds 1 to 4 and a sample from round 5 between 2000/2001 and 2001/2002.* (Report No. 05). Nottingham, United Kingdom: Great Britain, Department for Education and Skills.

Yale Child Welfare Research Program

Screened in

Low

Trickett, P. K., Apfel, N. H., Rosenbaum, L. K., and Zigler, E. F. (1982). A five-year follow-up of participants in the Yale Child Welfare Research Program. In E. F. Zigler and E. W. Gordon (Eds.), *Day Care: Scientific and Social Policy Issues* (200–222). Boston: Auburn House.

- *Study Rating Disposition:* MCGD that does not establish baseline equivalence

Screened out

Rescorla, L. A. (1979). *The Yale Child Welfare Research Program: Description and results.* Paper presented at the Biennial Meeting of the Society for Research in Child Development. San Francisco, CA.

Rescorla, L. A., & Zigler, E. (1981). The Yale Child Welfare Research Program: Implications for social policy. *Educational Evaluation & Policy Analysis*, *3*(6), 5–14.

DIRECT AND INDIRECT ENHANCEMENT MODELS

Eager and Able to Learn

Screened in

Low

McGuinness, C., Eakin, A. & Connolly, P. (2012). An evaluation of the effects of the eager and able to learn programme on outcomes for 2-3 years olds in early years settings. Belfast: Centre for Effective Education, Queen's University Belfast.

- *Study Rating Disposition:* Measures cannot be attributed solely to the program of interest due to a confound

LearningGames

Screened in

Low

Collins, A., Goodson, B., Luallen, J., Fountain, A. R., & Checkoway, A. (2010). *Evaluation of child care subsidy strategies: Massachusetts family child care study.* (OPRE 2011-1). Washington, DC: U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research and Evaluation.

- *Study Rating Disposition:* RCT with high attrition that does not establish baseline equivalence

Learning Language and Loving It

Screened in

Low

Ahrens, R. (2009). Training childcare providers to improve interactions with children: A comparison of in-service training versus direct modeling of desired behaviors. Unpublished master's thesis, University of Nebraska at Omaha.

- *Study Rating Disposition:* MCGD that does not establish baseline equivalence

Screened out

Flowers, H., Girolametto, L., Weitzman, E., & Greenberg, J. (2007). Promoting early literacy skills: Effects of in-service education for early childhood educators. *Canadian Journal of Speech-Language Pathology & Audiology*, *31*(1), 6–18.

Girolametto, L., & Weitzman, E. (2007). Promoting peer interaction skills: Professional development for early childhood educators and preschool teachers. *Topics in Language Disorders*, *27*(2), 93–110.

Girolametto, L., Weitzman, E., & Greenberg, J. (2003). Training day care staff to facilitate children's language. *American Journal of Speech-Language Pathology*, *12*(3), 299.

Girolametto, L., Weitzman, E., & Greenberg, J. (2004). The effects of verbal support strategies on small-group peer interactions. *Language, Speech & Hearing Services in Schools*, *35*(3), 254–268.

Girolametto, L., Weitzman, E., & Greenberg, J. (2006). Facilitating language skills: In-service education for early childhood educators and preschool teachers. *Infants and Young Children*, *19*(1), 36–49.

Music Education

Screened in

Low

Gruhn, W. (2002). Phases and stages in early music learning. A longitudinal study on the development of young children's musical potential. *Music Education Research*, *4*(1), 51–71.

- *Study Rating Disposition:* MCGD that does not establish baseline equivalence

Screened out

Bolduc, J. (2009). Effects of a music programme on kindergartners' phonological awareness skills. *International Journal of Music Education*, *27*(1), 37–47.

Vaiouli, P. (2014). *Music, engagement, and early literacy in inclusive early childhood settings.* (Doctoral dissertation). ProQuest Dissertations and Theses. (1545896786)

Walworth, D. D. (2007). *The effect of developmental music groups for parents and premature or typical infants under two years on parental responsiveness and infant social development.* (Doctoral dissertation). ProQuest Dissertations and Theses. (304873039)

Walworth, D. D. (2009). Effects of developmental music groups for parents and premature or typical infants under two years on parental responsiveness and infant social development. *Journal of Music Therapy*, *46*(1), 32–52.

Project Secure Child in Child Care

Screened in

Low

Biringen, Z., Altenhofen, S., Aberle, J., Baker, M., Brosal, A., Bennett, S., et al. (2012). Emotional availability, attachment, and intervention in center-based child care for infants and toddlers. *Development & Psychopathology*, *24*(1), 23–34.

- *Study Rating Disposition:* MCGD that does not establish baseline equivalence and does not use statistical controls

Program for Infant/Toddler Care (PITC)

Screened in

High

Weinstock, P., Bos, J., Tseng, F., Rosenthal, E., Ortiz, L., Dowsett, C., et al. (2012). *Evaluation of Program for Infant/Toddler Care (PITC): An On-site Training of Caregivers* (NCEE 2012-4003). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. Retrieved December 19, 2014 from http://ies.ed.gov/ncee/edlabs/projects/project.asp?ProjectID=90

- *Study Rating Disposition:* high rated outcomes had the disposition RCT with low attrition and no reassignment (full sample); moderate rated outcomes had the disposition RCT with high attrition or reassignment that establishes baseline equivalence and uses statistical controls (subgroups)

APPENDIX E

GLOSSARY

This page has been left blank for double-sided copying.

APPENDIX E. RESEARCH TERMS GLOSSARY

A

Absolute value. The value of a number, as a distance from zero, disregarding whether the number is positive or negative. For example, the absolute value for both +4 and -4 is 4.

Attrition. The loss of sample members from the study. Attrition typically occurs several ways. For example, some sample members refuse to participate; researchers may be unable to locate some sample members (for example, if they have moved); or researchers may exclude sample members from the study (for example, if a sample member was determined to be ineligible for the program or did not have data for all the required outcomes) although this may negatively affect the research design.

B

Baseline. The study's onset.

Baseline equivalence. Occurs when the intervention and comparison groups have similar characteristics (such as race and age) at the study's onset. For LITES, baseline equivalence was established when no statistically significant differences were detected on required measures at baseline.

C

Clustered randomized controlled trial (clustered RCT). Clusters (such as child care centers) are randomly assigned to the intervention.

Comparison group. A group with characteristics similar to those of intervention group members, except that they do not receive the services of interest. The comparison group is intended to represent what would have happened to members of the intervention group if they had not received the services from the model of interest. The more similar a comparison group is to the intervention group, the more likely it is that any difference in outcomes between the two groups can be attributed to the intervention.

Confounding factor. Occurs when an aspect of the study design, other than the model of interest, aligns with the intervention or comparison group, making it impossible to measure unbiased impact. For example, if one classroom caregiver administers all program ECE services, it is impossible to distinguish the effectiveness of that person from the effectiveness of the program. Confounding factors may also arise from systematic differences in the way data are collected from participants in the intervention group versus the comparison group. For example, participants may report information differently to someone they know than to someone they do not know. Familiarity with the data collector may change the way participants answer the questions. The presence of confounding factors can impede the ability of a study to capture an estimate of the actual effect of a program (that is, an unbiased impact).

Cronbach's coefficient alpha. An estimate of internal consistency reliability that indicates how well groups of items in an assessment "hang together" and contribute to measurement of the

same construct. The estimate captures the extent to which the separate items on the measure all seem to move in the same direction (that is, if a person is high on one item of a construct, they rate themselves high on all of the items related to that construct on a measure). The greater the similarity among items, the higher the reliability (and thus the higher the value of Cronbach's coefficient alpha). Values of the alpha can range from -1.0 to 1.0, with greater values indicating stronger internal consistency.

D

Differential attrition. Differential attrition rate is the absolute value of the difference between the attrition rates in the intervention and comparison groups.

E

Effect size. A measure of the magnitude of the difference between the intervention group and the comparison group. The effect size shows the magnitude of the impact (or the difference between the intervention and comparison group) relative to the standard deviation of the measure. A benefit of using the effect size is that it allows for comparisons of impacts across outcomes that may have been measured using different units. In the LITES review, a negative value indicated that the comparison group (which did not receive the services or program) had larger outcomes, on average, than the intervention group (which did receive services). A positive value indicated that the outcomes for the intervention group were greater than those for the comparison group. Values of 0 (referred to as a neutral effect) indicated there was no difference, on average, between the intervention and comparison groups.

F

Favorable effect. An estimated impact on an outcome measure in a direction that is beneficial for children and parents. This impact could be positive or negative, and is determined to be "favorable" based on the end result. For example, a favorable impact could be an increase in children's vocabulary or a reduction in harsh parenting practices.

Follow-up. A time point after the onset of the intervention for measuring participant outcomes.

I

Internal validity. A study's ability to isolate the effects of an intervention from other factors that may influence participants' outcomes.

Intervention group. The sample members who receive the early care and education services or program of interest.

M

Matched comparison group design (MCGD). A study design in which sample members (children, parents, or families) are selected for the intervention and comparison conditions in a nonrandom way.

Mean. A measure of the average value for a sample that equals the sum of all values divided by the number of sample members.

N

Null effect. An effect that is neither a significant or substantial favorable effect nor a significant or substantial unfavorable effect.

O

Outcome domain. A group of related outcomes that measure the same or similar constructs. The LITES review includes three primary child outcome domains: (1) cognitive development, (2) social-emotional/behavioral development, or (3) language development. Child health outcomes such as height, weight, gross and fine motor skills, and hospitalizations were reported if present in a study of a model that had at least one study with child outcomes in a cognitive, social-emotional/behavioral, or language domain. The LITES review also included long-term risk and economic well-being outcomes and several interim domains.

Overalignment. When outcome measures more closely align to one of the study groups than the other and could bias a study's results.

Overall attrition. The total number of sample members who are not participating at follow-up.

P

***p*-value.** The probability that the observed finding was obtained by chance when there is no true relationship in the population. For example, a sample may show a positive mean difference, suggesting that the intervention group has better outcomes than the comparison group, with a *p*-value of 0.05. The 0.05 *p*-value means that there is a 5 percent chance that the positive finding for the intervention group was obtained by chance and does not occur in the population.

R

Randomized controlled trial (RCT). A study design in which sample members (children, parents, or families) are assigned to the intervention and comparison groups by chance.

Reassignment. Compromising or violating random assignment—for example, children being switched from the comparison group to the intervention group after random assignment. If these children's outcome data were included as part of the intervention group's results, the study would suffer from reassignment and could not be reviewed as an RCT.

Regression discontinuity design (RDD). A design in which a continuous scoring variable is used to assign an intervention to study units. Units with scores below a pre-set cutoff value are assigned to the intervention group, and units with scores above the cutoff value are assigned to the comparison group, or vice versa. The effect of the intervention is estimated as the difference in mean outcomes between intervention and comparison group units, adjusting statistically for the relationship between the outcomes and the variable used to assign units to the intervention, typically referred to as the "forcing" variable.

Replicated effect. An effect that is statistically significant ($p \leq 0.05$) or has an effect size greater than or equal to 0.2 standard deviations and is measured in two or more non-overlapping analytic study samples.

S

Sample. Persons (children, caregivers, or families) included in the study. For the LITES review, sites that were analyzed separately were considered separate samples.

Significant effect. An impact estimate that is statistically significant with $p \leq 0.05$.

Single case design. These designs often involve repeated, systematic measurement of a dependent variable (outcome) before, during, and after the active manipulation of an independent variable (the intervention). These designs can provide a strong basis for establishing causal inference and are widely used in applied and clinical disciplines in psychology and education.

Standard deviation. A measure of the spread or variation of values in the sample. The standard deviation approximates the distribution around the mean with 68 percent of the sample having values that are between one standard deviation below the mean and one standard deviation above the mean. Smaller standard deviations indicate that the values for individual sample members are closer to the mean, whereas larger standard deviations indicate there is more variation in values.

Standardized (normed) instrument. An outcome measure that uses a uniform or standard set of procedures for administration and scoring. A norming sample, selected to be representative of the population of interest, was used to establish the standardized scoring system, or norms, for the measure.

Statistical controls. Methods of adjusting for characteristics that may differ between the intervention and comparison groups at baseline to make the groups more comparable.

Statistical significance. An indication of the probability that the observed finding was obtained by chance (when there is not a real relationship in the population). If the *p*-value is equal to or less than a predetermined cutoff (in the LITES review, 0.05), the finding is considered statistically significant because it has a low probability of having occurred by chance (5 percent or less).

Substantial effect. An impact estimate that has an effect size greater than or equal to 0.2 standard deviations in absolute value.

Sustained or delayed effect. An effect that is statistically significant ($p \leq 0.05$) or has an effect size greater than or equal to 0.2 standard deviations and is measured one year or more after the end of the intervention.

www.mathematica-mpr.com

Improving public well-being by conducting high quality, objective research and data collection

PRINCETON, NJ ■ ANN ARBOR, MI ■ CAMBRIDGE, MA ■ CHICAGO, IL ■ OAKLAND, CA ■ WASHINGTON, DC

MATHEMATICA
Policy Research