

UC Irvine

UC Irvine Previously Published Works

Title

Maximizing the potential of early childhood education to prevent externalizing behavior problems: A meta-analysis

Permalink

<https://escholarship.org/uc/item/0h24n1s3>

Journal

Journal of School Psychology, 53(3)

ISSN

0022-4405

Authors

Schindler, Holly S
Kholoptseva, Jenya
Oh, Soojin S
et al.

Publication Date

2015-06-01

DOI

10.1016/j.jsp.2015.04.001

Peer reviewed



Maximizing the potential of early childhood education to prevent externalizing behavior problems: A meta-analysis

Holly S. Schindler^{a,1}, Jenya Kholoptseva^b, Soojin S. Oh^b, Hirokazu Yoshikawa^c, Greg J. Duncan^d, Katherine A. Magnuson^e, Jack P. Shonkoff^b

^a University of Washington, College of Education, Miller Hall, Box 353600, Seattle, WA 98195, USA

^b Harvard University, USA

^c New York University, USA

^d University of California, Irvine, USA

^e University of Wisconsin–Madison, USA

ARTICLE INFO

Article history:

Received 20 January 2014

Received in revised form 8 April 2015

Accepted 8 April 2015

Keywords:

Early childhood education

Prevention

Externalizing behavior problems

Meta-analysis

Child social skills training

ABSTRACT

Early childhood education (ECE) programs offer a promising mechanism for preventing early externalizing behavior problems and later antisocial behavior; yet, questions remain about how to best maximize ECE's potential. Using a meta-analytic database of 31 studies, we examined the overall effect of ECE on externalizing behavior problems and the differential effects of 3 levels of practice, each with increasing specificity and intensity aimed at children's social and emotional development. In short, we found that each successive level of programs did a better job than the prior level at reducing externalizing behavior problems. Level 1 programs, or those without a clear focus on social and emotional development, had no significant effects on externalizing behavior problems relative to control groups ($ES = .13 SD, p < .10$). On the other hand, level 2 programs, or those with a clear but broad focus on social and emotional development, were significantly associated with modest decreases in externalizing behavior problems relative to control groups ($ES = -.10 SD, p < .05$). Hence, level 2 programs were significantly better at reducing externalizing behavior problems than level 1 programs ($ES = -.23 SD, p < .01$). Level 3 programs, or those that more intensively targeted children's social and emotional development, were associated with additional significant reductions in externalizing behavior problems relative to level 2 programs ($ES = -.26 SD, p < .05$). The most promising effects came from level 3 child social skills training programs, which reduced externalizing behavior problems half of a standard deviation more than level 2 programs ($ES = -.50 SD, p < .05$).

© 2015 Society for the Study of School Psychology. Published by Elsevier Ltd. All rights reserved.

1. Introduction

In recent years, there has been a growing interest in early childhood intervention strategies to prevent externalizing behavior problems (Farrington & Welsh, 2007). In childhood, externalizing behavior refers to a cluster of aggressive and disruptive behaviors including fighting, arguing, throwing tantrums, disturbing activities, or harming others (Achenbach & Edelbrock, 1978; Campbell, 1995; Duncan & Magnuson, 2011; Hinshaw, 2002; Zachrisson, Dearing, Lekhal, & Toppelberg, 2013). These behaviors emerge nearly universally in infancy, peak between ages 2 and 4, and then optimally decline (Keenan & Wakschlag, 2000; Shaw, Lacourse, & Nagin,

E-mail address: hschindl@uw.edu (H.S. Schindler).

Action Editor: Patricia Manz.

¹ Tel.: +1 206 616 0853.

2005; Tremblay, 2010; Tremblay et al., 1999). Yet, there is great variability in children's trajectories, with a concerning number leading to later antisocial and criminal activity (Campbell, Spieker, Burchinal, Poe, & NICHD Early Child Care Research Network, 2006; Moffitt, 1993; Tremblay, 2010; Tremblay et al., 2004). Approximately one third of children who display high levels of early externalizing behavior problems continue on a trajectory of problem behavior through adolescence (Loeber, Farrington, & Waschbusch, 1998; Losel & Bender, 2003, 2012). Young children who display high levels of early externalizing behavior problems are also at risk for a host of subsequent academic issues, including school retention, school dropout, and lower school engagement (Bub, McCartney, & Willett, 2007; Bulotsky & Fantuzzo, 2011; Duncan & Magnuson, 2011; Fantuzzo et al., 2007; Raver, 2002). Furthermore, externalizing behavioral concerns in early primary classrooms take up significant portions of teachers' time and efforts, thereby undermining the learning experience for all students (Houts, Caspi, Pianta, Arseneault, & Moffitt, 2010).

In response, a range of approaches for improving externalizing behavior prior to formal school entry have been developed and tested. These approaches have been heavily influenced by ecological models of development (Campbell, Shaw, & Gilliom, 2000). Specifically, Bronfenbrenner's (1977) ecological model suggests that variability in children's developmental trajectories is the result of different interactions between a child and a complex web of settings that span from immediate contexts (e.g., family and school contexts) to more distal ones (e.g., neighborhood, cultural, and political contexts). During the early childhood years, the characteristics of these nested contexts and the connections among them are thought to influence whether or not children's behavior improves as they learn to engage in more positive and less negative ways of interacting with others, or alternatively persist into middle childhood and even later (Campbell et al., 2000; Moffitt, 1993). Indeed, within children's immediate environments, many theorists and empiricists have documented a link between early harsh, inconsistent, and coercive caregiver–child relationships and higher levels of externalizing behavior problems (e.g., Hirschi, 1969; Patterson, DeBaryshe, & Ramsey, 1989; Smith et al., 2014; Tolan, Dodge, & Rutter, 2013). At more distal levels, poverty, high-crime neighborhoods, and persistent discrimination have also been linked to externalizing behavior problems (Campbell et al., 2000). In accordance with these findings, early prevention programs have operated in a number of different contexts, including families' homes, childcare and preschool centers, clinics, and communities (Losel & Bender, 2012).

One prevention approach that deserves greater scrutiny is early childhood education (ECE), defined as center-based education for children from birth to age 5. The number of children attending ECE programs has grown rapidly in recent years (Burgess, Chien, Morrissey, & Swenson, 2014), with programs reaching an unprecedentedly large proportion of low-income children, who are at an increased risk for displaying externalizing behavior problems (Huaqing Qi & Kaiser, 2003; Mattered, Lloyd, Fishman, & Bangser, 2013). This growth makes widespread early prevention within an ECE context more feasible but also raises important questions about best practices. As policymakers, educators, and researchers increasingly consider the merits of ECE programs, the time is right to take stock of what we know about their capacity to address externalizing behavior problems and to identify which program features are associated with the most beneficial effects.

In this study, we employ meta-analytic techniques to conduct a systematic analysis of what we know about how ECE programs affect externalizing behavior problems in young children, and to better understand how externalizing behavior problems might vary as a function of ECE practices.

2. Background

2.1. ECE and externalizing behavior problems: an overview

Despite several decades of evaluation research, it has been difficult to draw conclusions about ECE programs' effects on externalizing behavior problems. Theoretical and empirical research can be found to support both a positive and a negative relationship.

On the one hand, ECE programs theoretically offer a unique opportunity to target several early protective factors associated with lower externalizing behavior problems, including children's early cognitive abilities (particularly verbal abilities), self-regulation, and social skills; and caregiver warmth, responsiveness, and behavioral management techniques (Schindler & Yoshikawa, 2012; Yoshikawa, Schindler, & Caronongan, 2009). Some empirical studies have supported this view, finding that participation in ECE programs resulted in reductions in externalizing behavior problems. Perry Preschool is a well-known example of a high-quality ECE program for 4- and 5 year-olds implemented in the 1960s that successfully reduced externalizing behavior problems and later criminal and antisocial activity (Schweinhart et al., 2005). In addition, two reviews of evaluations of center-based preschool programs for 3 to 5 year olds found small, positive effects on social and emotional outcomes, though neither included evaluations of programs for younger children or examined externalizing behavior problems separately (Camilli, Vargas, Ryan, & Barnett, 2010; Nelson, Westhues, & MacLeod, 2003).

On the other hand, some theoretical perspectives argue for ECE increasing externalizing behavior. It is possible that ECE programs do not promote protective factors and instead may increase exposure to some risk factors, such as aggressive peers. The prolonged group interactions that are common in ECE settings may also increase children's physiological stress and thereby negatively impact their development, particularly for children under age 3 (Vermeer & van IJzendoorn, 2006). Several empirical investigations have supported a positive association between ECE enrollment and higher levels of externalizing behavior problems. Coley, Votruba-Drzal, Miller, and Koury (2013) examined a nationally representative sample of 6000 children from the Early Childhood Longitudinal Study- Birth Cohort. They found that center-based preschool for 4-year-olds predicted heightened externalizing behavior problems in kindergarten relative to children who had remained in parental care. ECE during infancy and toddlerhood showed no independent associations but did exacerbate the negative relationship between preschool and externalizing behavior problems. Two other studies came to similar conclusions about preschool using another nationally representative longitudinal dataset (Early Childhood

Longitudinal Study – Kindergarten Class of 1998–1999). One specifically examined participation in school-based and publicly funded prekindergarten programs and found that participation was associated with higher levels of behavior problems (Magnuson, Ruhm, & Waldfogel, 2007). The other study examined the effects of any center-based care and found that, overall, center-based care predicted heightened behavior problems. These negative associations were larger the younger the children enrolled in centers (Loeb, Bridges, Bassok, Fuller, & Rumberger, 2007).

Finally, some studies have reported no association between ECE and externalizing problem behavior (Howes et al., 2008; Loeb, Fuller, Kagan, & Carrol, 2004). Loeb, Fuller, Kagan, and Carrol compared outcomes for low-income children between 12 and 42 months who were in ECE programs to similar children cared for by kith or kin and found no effects on externalizing behavior problems. Howes and colleagues similarly found that enrollment in pre-kindergarten failed to predict any differences in children's externalizing behavior problems.

These studies, which have attempted to examine average impacts of a range of ECE programs, have come to quite different conclusions. Hence, our first research question is: What is the average treatment effect of ECE on children's externalizing behavior problems? Several of the aforementioned studies also examined whether various elements, such as timing and program duration, affected the relationship between ECE and externalizing behavior problems. However, to date, few studies looking across ECE programs have attempted to delineate program practices that may differentiate the programs that effectively reduce behavior problems and those that do not.

2.2. ECE and the prevention of externalizing behavior problems: three levels of practice

Our remaining research questions focus on understanding the relative effects of three levels of ECE practice for preventing externalizing behavior problems. ECE programs differ a lot in their approaches, including the extent to which the programs focus on improving or building children's social and emotional development. In our analyses, we examine three levels of emphasis on social and emotional development from no focus (level 1) to a clear and intensive focus (level 3). The framework guiding our analyses is presented in Fig. 1 and is described below. We use this framework to differentiate among common approaches to social and emotional practices in ECE programs.

2.2.1. Level 1: Programs without a clear focus on social and emotional development

Level 1 ECE programs operate without an explicit focus on social and emotional development. Even so, these programs may still be able to reduce externalizing behavior problems through the provision of an emotionally supportive environment that offers responsive and nurturing interactions (Gormley, Phillips, Newmark, Welti, & Adelstein, 2012). In addition, difficulties with early cognitive and achievement skills, particularly verbal abilities, put children at risk for externalizing behavior problems. Hence, programs that place an emphasis on academic preparedness, even without a content focus on social and emotional development, may still have important benefits for these behaviors. Such programs, particularly in preschool, have become increasingly popular in recent decades as both researchers and policymakers have emphasized the importance of literacy and math skills for school readiness (NAEYC & NCTM,

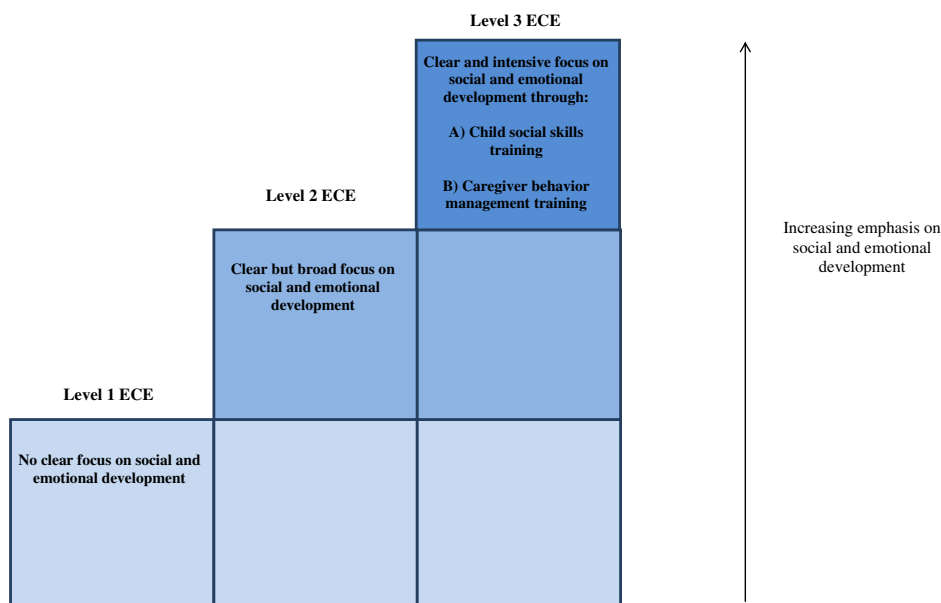


Fig. 1. ECE and the prevention of externalizing behavior problems: Three levels of practice.

2002; Snow, Burns, & Griffin, 1998). In practice, a level 1 program might focus on promoting child-directed group and individual activities, in which a teacher follows a child's cues and responds with appropriate, language-rich interaction, without a specific instructional focus. A level 1 program may also incorporate particular instructional activities focused on pre-academic skills, such as literacy or math, without incorporating specific activities focused on social and emotional development.

2.2.2. Level 2: programs with a clear but broad focus on social and emotional development

Level 2 programs build on level 1 programs by incorporating components focused on social and emotional development; however, these programs do so in the context of activities promoting skills in a range of developmental domains. In other words, level 2 programs include goals both for improving cognitive and achievement skills and for improving social and emotional development. The idea behind this approach is that social and emotional needs are intertwined with cognitive abilities, all of which are important for healthy child development, successful school performance, and later adult productivity (National Scientific Council on the Developing Child, 2007). Global curricula, which have a wide scope, fall into the level 2 category. For example, the High Scope model for infants, toddlers, and preschoolers pairs a focus on adult-child interaction with daily routines and suggested activities focused on building social, emotional, cognitive, and physical development (High Scope Educational Research Foundation, 2014).

Level 2 programs are widely implemented. Eighty percent of Head Start programs use a global curriculum (U.S. Department of Health and Human Services, 2005), and both Head Start and Early Head Start stand out as prominent examples of programs that have included social and emotional goals in their programming (Zigler & Styfco, 1995). Though many level 2 programs have been evaluated relative to control groups, research has not yet been conducted to empirically compare effects on externalizing behavior problems of programs that include a broad focus on social and emotional development versus programs without such a focus. Hence, our second research question is: Are level 2 ECE programs that report a clear but broad focus on social and emotional development associated with fewer externalizing behavior problems than level 1 programs that do not articulate that focus?

2.3. Level 3: programs with a clear and intensive focus on social and emotional development

In recent years, a third level of programs has emerged that places even more emphasis on social and emotional learning as well as positive behavior. These programs often supplement a global curriculum with program components that target social and emotional competencies more intensively than in typical ECE settings (Morris et al., 2014; Raver, 2002). These additional program components are considered "enhancements" to the more standard level 2 practices (Mattera et al., 2013). We classify level 3 enhancements into two categories: a) child social skills training, and b) caregiver behavior management training.

2.3.1. Level 3a: child social skills training

Deficits in social skills and competencies have been empirically linked to externalizing behavior problems and antisocial development (Farrington & Welsh, 2007; Losel & Bender, 2006). Hence, social skills training has been recommended not just for improving social development in general but also specifically for preventing externalizing behavior problems. Level 3 social skills training enhancements assert that some children may need "explicit instruction to ensure they develop competence in emotional literacy, anger and impulse control, interpersonal problem solving, and friendship skills" (Fox, Dunlap, Hemmeter, Joseph, & Strain, 2003, p. 2). This program enhancement is directly informed by the social information processing model of externalizing behavior problems, which assumes that individuals differ in how they interpret situations, read social cues, and initiate and evaluate reactions to situations, and that those differences are learned through experience (Crick & Dodge, 1994). Hence, social skills training programs aim to prevent problematic trajectories through teaching children appropriate behaviors (e.g., making friends, accepting others' rights, expressing needs and emotions) and/or targeting the basic cognitive skills needed for effective social problem solving (e.g., perspective taking, self-control, and anger management) (Losel & Bender, 2012).

Social skills training programs that act as program enhancements in an ECE context often use a series of "circle time" structured group lessons to promote social skills and social competence in children. For example, preschool PATHS ("Promoting Alternative Thinking Strategies") is a 30-lesson curricular enhancement focused on teaching children about emotions and how to respond in social situations (Domitrovich, Cortes, & Greenberg, 2007; Morris et al., 2014). Preschool PATHS was originally developed to supplement Head Start's typical programming. Thematic units include lessons on basic and advanced feelings, a self-control strategy, and problem solving. For example, in a lesson on "feelings," the teacher instructs children to look at a drawing of faces. From there, the teacher discusses how the physical features of faces can provide clues for how someone is feeling on the inside. Then, the teacher asks children to think about the parts of a face that show when someone is happy and to identify pictures of people who are happy or not happy. The lesson ends with an activity in which children use a blank face template and cutouts of mouths, eyes, ears, and noses, and the teacher asks them to make a happy face and a sad face with the materials. Later in the day, the teacher asks children to try and identify how their peers are feeling (Mattera et al., 2013).

2.3.2. Level 3b: caregiver behavior management training

The second type of level 3 enhancement takes a different approach to preventing externalizing behavior problems. This program enhancement focuses on promoting caregivers' positive behavioral management techniques, which are known protective factors against behavior problems (Patterson, Reid, & Dishion, 1992; Schindler & Yoshikawa, 2012). Caregiver behavior management trainings are informed by self-control theory (Hirschi, 1969) and a social-interactional perspective (Patterson et al., 1989), both of which point to the importance of early interactions with caregivers. Techniques that are emphasized in this type of level 3

enhancement offered to teachers and/or parents include effective limit setting, appropriate discipline strategies, and rewarding positive behavior.

The impetus for offering behavior management training to teachers has come from the high frequencies of externalizing behavior problems that early childhood teachers report and from their identification of the ability to manage challenging child behavior as one of their top training needs (Snell et al., 2012; Yoshikawa & Knitzer, 1997; Yoshikawa & Zigler, 2000). Parent behavior management training programs operating outside of the context of ECE have a longer history (Piquero & Jennings, 2012), but they have only more recently been evaluated as enhancements within ECE settings.

The Incredible Years program is an example that has been used with both parents and teachers. In this enhancement to ECE services, parents participate in 12-week parent groups and are taught behavior management techniques through viewing video vignettes and having group discussions. Teachers participate in six workshops that employ the same video-vignette method to teach classroom-wide positive management and discipline strategies (Webster-Stratton, Reid, & Hammond, 2001).

Though individual level 3 programs have been evaluated and summarized by others, questions remain about the average benefit of these level 3 programs over level 2 programs, as well as whether the two conceptually different types of level 3 programs vary in their impacts (Mattera, Lloyd, Fishman, & Bangser, 2013). In this paper, our third and fourth research questions attempt to address these gaps in the literature. Specifically, we ask: Are level 3 ECE programs with a clear and intensive focus on social and emotional development associated with fewer externalizing behavior problems than level 2 programs? And if so, is child social skills training or caregiver behavior management training more effective?

3. Summary of research questions

This is the first meta-analysis to explicitly examine the relative effects of different practices used within ECE settings on externalizing behavior problems in young children. Specifically, we examine the relative benefits of three levels of practice, each with increasing specificity and intensity aimed at children's social and emotional development (see Fig. 1). To summarize, we address the following four research questions:

1. What is the average treatment effect of ECE on children's externalizing behavior problems?
2. Are level 2 ECE programs that report a clear but broad focus on social and emotional development associated with fewer externalizing behavior problems than level 1 programs that do not articulate that focus?
3. Are level 3 ECE programs with a clear and intensive focus on social and emotional development associated with fewer externalizing behavior problems than level 2 programs?
4. If so, is child social skills training or caregiver behavior management training more effective?

4. Methods

To answer our research questions, we draw upon a comprehensive meta-analytic database that includes evaluations of ECE programs implemented in the United States and published between 1960 and 2007. In this paper, we limited our analyses to evaluations of ECE programs for children from birth to five years old that included measures of externalizing behavior problems. Our meta-analytic approach transformed estimates of externalizing behavior problems into a common metric, an "effect size," that represents the difference between a treatment group and a control group or a treatment group and an alternative treatment group. These effect sizes, which are expressed as fractions of standard deviations, were then used to estimate average effects across evaluations and to explore factors that might maximize or minimize a program's effectiveness. We followed standard procedures for conducting meta-analyses, which include (a) literature search, (b) data evaluation, and (c) data analysis (Cooper, Hedges, & Valentine, 2009).

4.1. Literature search

In this study, we utilized a subset of studies from a larger meta-analytic database being compiled by the National Forum on Early Childhood Policy and Programs (http://developingchild.harvard.edu/activities/forum/meta_analytic_database/). The starting point for the Forum's database was a bibliography of evaluations of ECE programs for 3- to 5-year-old children conducted between 1960 and 2003, compiled by Abt Associates and the National Institute for Early Education Research (NIEER) (Camilli et al., 2010; Jacob, Creps, & Boulay, 2004; Layzer, Goodson, Bernstein, & Price, 2001). The Forum's database updated Abt's list to 2007, added programs for children birth to age 3, and employed more rigorous methodological inclusion criteria (see Data evaluation section).

The Forum identified additional reports for inclusion by conducting keyword searches in ERIC, PsychINFO, EconLit, and Dissertation Abstract databases; manually searching the websites of policy institutes (e.g., RAND Corporation, Mathematica, NIEER) and state and federal departments (e.g., U.S. Department of Health and Human Services); and collecting references mentioned in included studies and other key reviews. In total, the Forum's search resulted in 10,309 reports for possible inclusion, in addition to the reports identified previously by Abt and NIEER. It is important to note that in traditional bibliographic searches, the goal is typically to maximize efficiency by minimizing both false positives (identified evaluations that are irrelevant) and false negatives (relevant evaluations not identified). In meta-analysis, the main objective is to reduce the number of false negatives (relevant evaluations not identified); therefore, the meta-analyst must be willing to cast a wider net and examine more false positives (Cooper et al., 2009).

4.2. Data evaluation

The next step in the meta-analysis process was to assess reports in order to identify high-quality evaluations with similar treatment and comparison groups (Borenstein, Hedges, Higgins, & Rothstein, 2009). To be included in the Forum's database, evaluations had to have: a) a comparison group (either an observed control or alternative treatment group); b) at least 10 participants in each condition; and c) attrition rates of less than 50% between the initiation of treatment and measurement. Included evaluations had to be experimental or quasi-experimental, using one of the following designs: random assignment, regression discontinuity, fixed effects (individual or family), residualized or other longitudinal change models, difference in difference, instrumental variables, propensity score matching, or interrupted time series. Quasi-experimental evaluations not using one of the former analytic strategies were also included if they had a comparison group and either a) pre-and post-test information on the outcome of interest or b) demonstrated adequate comparability of groups on baseline characteristics. Studies in which the intervention was designed for children with diagnosed behavioral, emotional, or medical disorders or learning disabilities were excluded. These methodological criteria are more rigorous than those applied by Abt Associates and NIEER. For example, the Forum's database excludes all pre to post only (no comparison group) evaluations, as well as regression-based studies in which the baseline equivalence of treatment and control groups was not investigated.

The majority (91%) of the 10,309 ECE reports identified by the National Forum on Early Childhood Policy and Programs' research team's search were excluded after reading the abstracts, indicating that they did not meet inclusion criteria for obvious reasons (e.g., they were not quantitative evaluations or did not have a comparison group). Of the evaluation reports that made it to the full-text screening phase, reasons for exclusion most often included not meeting the research design criteria, with a small number of reports being excluded due to other eligibility criteria (e.g., they only reported results for students with disabilities). The high percentage of excluded reports is typical in meta-analytic searches and reflects the need to cast a wide net (Cooper et al., 2009), as discussed in our description of the literature search process. Once all reports were assessed, those that were screened-in were grouped into 273 studies.

Each of the 273 studies was then coded by a team of nine doctoral-level research assistants in order to document information about study design, program and sample characteristics, and information needed to compute effect sizes. Prior to coding independently, the Forum's meta-analytic research team implemented a training process that included practice coding, assessing reliability, and holding regular meetings (Cooper et al., 2009). Specifically, research assistants were trained during a 3 to 6 month process, during which time an overview of the project was provided; each item in the codebook was discussed; a manualized effect size training was completed; and a sample of studies was coded alongside an experienced coder. Trained coders were then required to achieve an interrater agreement of 1.00 for effect sizes and .80 for all other study information with a master set of codes, based on the procedure used in Abt's meta-analytic database (see Camilli et al., 2010). The master set of codes was created by two original developers of the codebook. The range of interrater agreements for all study information was .87–.96 ($M = .87$).

Throughout coding, discrepancies and questions were resolved through weekly meetings between coders and principal investigators, and decisions were kept in an annotated codebook to ensure that decisions made about any ambiguities during these meetings were followed throughout the coding process (Cooper et al., 2009). Two additional steps were taken to ensure high-quality data. First, at the conclusion of the construction of the database, data entry was checked through systematic data exploration and cleaning. This process included checking outliers; confirming skip patterns, and examining missing values. Second, primary independent variables used in the present study were independently coded for all included studies by the first author. This agreement rate was nearly perfect (.97).

4.3. Overall database and analytic sample

The resulting database of the National Forum on Early Childhood Policy and Programs is organized into a three-level hierarchy: study, contrast, and effect size. A "study" is defined as a collection of comparisons in which the treatment groups are drawn from the same broadly-defined pool of subjects ($N = 273$). Each study also produces a number of "contrasts," defined as comparisons of groups that experienced different conditions within a study ($N = 1242$). For example, a study might provide one group of participants with level 2 ECE services, one group of participants with level 3 ECE services, and one group of participants with no services. This would represent three contrasts: level 1 ECE versus no services, level 2 ECE versus no services, and level 1 ECE versus level 2 ECE. The former two contrasts are labeled as "treatment versus control" contrasts since the measured difference is between a group who was offered services and a group who was offered no services. The latter contrast is labeled as a "treatment versus alternative treatment" contrast since the measured difference is between two groups who were offered different services. Finally, effect sizes are the standardized differences on a set of outcome measures ($N = 15,804$). Effect sizes (Hedges' g) were computed using the software package Comprehensive Meta-Analysis (Borenstein, Hedges, Higgins, & Rothstein, 2005) and are expressed as a fraction of a standard deviation. A standardized mean difference is useful for making the different instruments used within and across studies comparable (Bushman & Wang, 2009). Hedges' g adjusts the standardized mean difference (Cohen's d) to account for bias when sample sizes are small.

For the present study, we used a subset of the Forum's database and only included studies that measured externalizing behavior problems. We included both treatment versus control contrasts and treatment versus alternative treatment contrasts. Each of the included treatment versus control contrasts compared a group of participants who was offered ECE services to a group who was not offered ECE services. These contrasts were used to answer our first two research questions (RQ1: What is the average treatment effect of ECE on children's externalizing behavior problems?; RQ2: Are level 2 ECE programs that report a clear but broad focus on social and

emotional development associated with fewer externalizing behavior problems than level 1 programs that do not articulate that focus?). None of the treatment versus control contrasts included level 3 program enhancements (child social skills training or caregiver behavior management). To examine the effects of these types of programs, we turned to a sample of treatment versus alternative treatment comparisons, which were used to answer research questions 3 and 4 (RQ3: Are level 3 ECE programs with a clear and intensive focus on social and emotional development associated with fewer externalizing behavior problems than level 2 programs?; RQ4: If so, is child social skills training or caregiver behavior management training more effective?). Our resulting sample represented 143 effect sizes nested in 55 contrasts and 31 studies. Of the 55 contrasts, 39 were treatment versus control comparisons. The remaining 16 were treatment versus alternative treatment contrasts.

4.4. Measures

4.4.1. Dependent variable

We examined program impacts on 1) externalizing behavior problems, conceptualized as aggressive, disruptive, impulsive, or hyperactive behavior (Achenbach & Edelbrock, 1978; Campbell, 1995; Hinshaw, 2002; Zachrisson et al., 2013) and 2) later measures of corresponding antisocial behaviors (e.g., ever committed purposeful injury; ever arrested for a violent offense). We excluded measures that combined reports of externalizing behavior problems with other social and emotional domains. For example, we excluded measures of the overall Child Behavior Checklist (Achenbach, 1992) because it assesses both internalizing and externalizing behavior. However, we included the Child Behavior Checklist's externalizing behavior problem measure and its subscales.

Ages of children at the time of measurement ranged from 18 months to 40 years. For children from 18 months to 3 years old, the most common instruments were subscales of the Child Behavior Checklist (13/23 effect sizes) and observations of behavior problems (e.g., having tantrums, throwing toys) in laboratory-based tasks (9/23 effect sizes). Instruments for children ages 4 to 5 were much more varied, but the most common instruments included subscales of the Child Behavior Checklist (13/76), the Social Skills Rating System (11/76), the Preschool and Kindergarten Behavior Scale (6/76), and the Eyberg Child Behavior Inventory (4/76). For children older than 5 years of age, the most common instruments were counts of aggressive behaviors (22/44) and the presence of antisocial behaviors (11/44).

Effect sizes were coded so that negative numbers represent lower levels of externalizing behavior problems. In other words, a negative effect size suggests that the intervention had a beneficial effect on externalizing behavior problems. For treatment versus control contrasts, the range of effect sizes was -1.09 to 1.01 . For treatment versus alternative treatment contrasts, the range was -1.59 to $.68$.

4.4.2. Independent variables: three levels of practice

As part of our protocol, researchers coded the central stated purposes of each ECE program. Up to four goals could be coded, and only goals that were clearly stated in the reports were marked. All ECE programs used in our analyses included a stated goal to improve child cognitive or school outcomes. If a report failed to also include a stated goal to improve social and emotional development (including child behavior, emotional regulation, or relational skills), then the program was coded as a level 1 ECE program. If a report did include a stated goal to improve social and emotional behavior, but did not describe a level 3 enhancement, then the program was coded as a level 2 ECE program.² Finally, if the report included a stated goal to improve social and emotional development *and* described child social skills training or caregiver behavior management training, then the program was coded as a level 3 ECE program. Within level 3 ECE programs, a variable was created to designate if the program included child social skills training. Programs were coded as including child social skills training enhancements if there was a program enhancement focused on teaching children appropriate behaviors (e.g., making friends, accepting others' rights, expressing needs and emotions) and/or targeting the basic cognitive skills needed for effective social problem solving (e.g., perspective taking, self-control, and anger management). Another variable was created to designate if the program included caregiver behavior management training. Programs were coded as including caregiver behavior management training if they included an enhancement focused on giving teachers or parents techniques for behavior management, such as effective limit setting, appropriate discipline strategies, and rewarding positive behavior. See [Appendices A, B, and C](#) for descriptions of programs in each level.

4.4.3. Independent variables: covariates

Research questions one (What is the average treatment effect of ECE on children's externalizing behavior problems?) and three (Are level 3 ECE programs with a clear and intensive focus on social and emotional development associated with fewer externalizing behavior problems than level 2 programs?) were addressed through meta-analyses of conditions that were experimentally or quasi-experimentally evaluated; therefore, we did not need to include covariates in the models that attended to these questions. However, in order to answer research questions two (Are level 2 ECE programs that report a clear but broad focus on social and emotional development associated with fewer externalizing behavior problems than level 1 programs that do not articulate that focus?) and four (Is child social skills training or caregiver behavior management training more effective?), we included a set of covariates in our models. Covariates were used in these analyses since evaluations did not experimentally or quasi-experimentally examine differences

² It should be noted that programs were not designated as level 2 programs based on an objective measure of the proportion of instruction dedicated to externalizing behavior. Rather, our judgment of a program's broad focus on social and emotional development is based on the assumed relationship between stated goals and program content. This limitation is an important one in meta-analysis – the need to rely only on what is reported in prior studies.

between program types. Rather, we utilized variability across evaluations, controlling for potentially confounding variables, to answer these research questions.

We included in our regression models a set of differences in program, participant, and study design characteristics based on the ECE literature and prior relevant meta-analytic studies (e.g., Shager et al., 2013). As a program characteristic, we included a continuous variable indicating the total length of the program, measured in months for treatment versus control contrasts (which were longer, on average) and in weeks for treatment versus alternative treatment contrasts (which were shorter, on average). We also included a continuous measure indicating program dosage, expressed as hours per week. At the participant level, we included a continuous variable indicating the average age of participant children at the time of measurement and a dichotomous variable noting if the majority of participants were African American or Latino, with the omitted category representing participants that were majority non-Latino white or other non-reported groups.³

In order to control for variation in the quality of study design, we included an index ranging from zero to three, with higher values representing higher quality studies. The index was created by summing across three dichotomous measures: 1) the study used random assignment, 2) the study had less than 25% attrition in treatment and comparison groups at the time of follow-up, and 3) coders did not observe any evidence of systematic bias in the evaluation or study methods (i.e., attrited treatment subjects were excluded from analyses; degree of volunteering for the program was different for the experimental and control groups). We also included a dichotomous variable indicating if the study was published in a peer-reviewed journal to account for the possibility that larger and more significant findings are more likely to be published in such outlets. Similarly, a variable was included that identified studies with an active control group (sought out ECE services out of their own volition), a characteristic thought to be associated with smaller effect sizes. Finally, we included a set of dichotomous variables to describe if the measure was taken during treatment, at the end of treatment, or at follow-up (omitted).⁴

In some cases, reports in our sample failed to provide the necessary information needed to code one or more of the following covariates: length of program, hours per week, average age of participant at time of measurement, and whether the majority of participants were African American or Latino. Rates of missing values for these four covariates ranged from 5.2% to 31.3%. Missing data on covariates is a common problem in meta-analytic studies because the conventions and norms regarding information reported in the evaluation studies varies so much across journals, disciplines, and time. Multiple imputation (MI) is generally the preferred method for accounting for missingness in studies based on individuals, families, and schools because each case with missing data on some variables typically has non-missing data on a number of correlated variables that can be used to generate the imputation. In contrast, meta-analytic studies, including our own, often lack sufficient observed characteristics that provide a strong basis for imputation. In addition, using MI in multi-level models can cause other problems (e.g., unstable estimates). Therefore, to indicate missingness in our analyses, we instead used dummy variable adjustments (Allison, 2002; Puma, Olsen, Bell, & Price, 2009). For example, if a contrast did not provide information on the length of the program, a code of zero replaced the missing value on this variable and a second dummy variable was created to indicate missingness on length of program. Both the length of program predictor variable (including new zero codes) and the dummy coded variable indicating missingness were included simultaneously in analyses. The same approach was used for the remaining covariates with missing data. This method has been recommended as an alternative to multiple imputation when covariates are missing in experimental studies and is superior to other common techniques, such as mean value imputation and case deletion (Puma et al., 2009).

4.5. Analytic approach

We first examined contrast and effect size level statistics separately for our analytic samples of treatment versus control and treatment versus alternative treatment comparisons. Then, we conducted a series of multi-level HLM models to answer our four research questions. The nested nature of the effect-size data (i.e., effect sizes clustered within contrasts, which in turn are clustered within studies) gave rise to the need to adjust statistically for the non-independence of the effect size estimates. For the purposes of these analyses, we estimated a two-level model, with level 1 reflecting effect sizes and level 2 reflecting contrasts. We do this for both theoretical and practical reasons. First, the most typical circumstance giving rise to multiple contrasts is multiple treatment arms with separate groups of children. Therefore, we expected effect sizes to be more similar within contrasts than within studies. Indeed, the intraclass correlation for effect sizes within contrasts ($ICC = .44$) was slightly larger than the intraclass correlation for effect sizes within studies ($ICC = .37$). Second, more than half of the included studies included only one contrast, thus the data would not support a three-level model.

Using a multi-level HLM framework (Raudenbush & Bryk, 2002), the level-1 model (effect size level) is:

$$ES_{ij} = \pi_{0j} + \pi_{1j}x_{1ij} + \dots + \pi_{kj}x_{kij} + e_{ij} \quad (1)$$

³ Since meta-analysis relies on information reported by prior studies, we were not able to capture more nuanced ethnic and racial group differences.

⁴ Some covariates were excluded in analyses to answer research question four because they weren't relevant when examining treatment versus alternative treatment contrasts. For example, the "active control" variable was no longer applicable. Other measures excluded from research question four analyses included hours per week (due to the lack of consistency in studies' reporting of this) and when the measure was taken (since nearly all of the measures were taken at the end of treatment).

where effect size i in contrast j is modeled as a function of the intercept (π_{0j}), which represents the average (covariate adjusted) effect size for all contrasts; a set of estimated effects of k independent variables measured at the effect size level ($\pi_{1j}x_{1ij} + \dots + \pi_{kj}x_{kij}$); and a within-contrast error term (e_{ij}). The level-2 equation (contrast level) models the intercept as a function of the grand mean effect size (β_{00}), p independent variables measured at the contrast level ($\beta_{01}x_{1j} + \dots + \beta_{0p}x_{pj}$) and a between-contrast random error term (u_{0j}):

$$\pi_{0j} = \beta_{00} + \beta_{01}x_{1j} + \dots + \beta_{0p}x_{pj} + u_{0j}. \quad (2)$$

This “mixed effects” model, which can also be expressed in one equation by substituting (2) into (1), assumes that there are two sources of variation in the effect size distribution, beyond subject-level sampling error: 1) the “fixed” effects of effect size and contrast level variables that measure key features of the program design, contrast-level study methods, effect size characteristics and other covariates; and 2) remaining “random” unmeasured sources of variation between (u_{0j}) and within contrasts (e_{ij}), which are assumed to be homogeneous both within and between model levels. Similar to other meta-analyses with multi-level data, we conducted our analyses using the PROC MIXED Procedure in SAS and weighted effect sizes by the inverse of the variance of each effect size multiplied by the inverse of the number of effect sizes per contrast (Camilli et al., 2010; Cooper, Hedges, & Valentine, 2009; de la Torre, Camilli, Vargas, & Vernon, 2007; Kuppens, Laurent, Heyvaert, & Onghena, 2013; Lipsey & Wilson, 2001).

We began with an intercept-only multi-level regression model (no predictors) of treatment versus control contrasts in order to answer research question one (What is the average treatment effect of ECE on children's externalizing behavior problems?). In order to answer research question 2 (Are level 2 ECE programs that report a clear but broad focus on social and emotional development associated with fewer externalizing behavior problems than level 1 programs that do not articulate that focus?), we added in a variable designating if a program was a level 2 program (versus level 1), followed by a model with covariates. To answer research questions 3 and 4, we turned to the treatment versus alternative treatment contrasts. To answer research question 3 (Are level 3 ECE programs with a clear and intensive focus on social and emotional development associated with fewer externalizing behavior problems than level 2 programs?) we ran an intercept-only multi-level regression model (no predictors). To answer research question 4 (Is child social skills training or caregiver behavior management training more effective?), we ran intercept-only models separately for each type of approach to determine which approach(es) resulted in significant and meaningful benefits compared to ECE programs with more general approaches to improving behavior. We also ran a model with child social skills training as the predictor, followed by a model that added in covariates, in order to see if level 3 ECE programs with child social skills training enhancements resulted in significantly different effects than those with caregiver behavior management training enhancements.

Following our primary analyses, we carried out several additional analyses to test the sensitivity of our findings to alternate model specifications. We also explored publication bias. First, we created two funnel plots: one for treatment versus control contrasts and one for treatment versus alternative treatment contrasts. Funnel plots are visual representations of the data in which mean effect sizes are plotted against some measure of sample size (e.g., standard error) (Sterne & Egger, 2001). Second, we computed fail-safe N s for our two primary significant findings (Rosenthal, 1979). Fail-safe N s are a calculation of the number of non-significant effects that would need to be added to the dataset to nullify findings.

5. Results

5.1. Descriptive Statistics

In Table 1, we present contrast- and effect size-level descriptive statistics separately for our analytic samples of treatment versus control and treatment versus alternative treatment comparisons. Of the treatment versus control contrasts (those that compared groups of participants who were offered ECE services to groups not offered ECE services), 28% were from level 1 programs and 72% were level 2 programs. The average length of the programs evaluated was 26.39 months, and the average number of hours per week of provided ECE was 15.10. On average, 86% of children were low-income. African American and Latino children represented more than half of the participants in 41% of the contrasts. In the treatment versus alternative treatment contrasts (those that compared level 3 programs to level 2 programs), 50% of the enhancements being evaluated included child social skills training, while 56% included caregiver behavior management training. These program enhancements were, on average, 19.75 weeks long. An average of 88% of children were low-income, and African American and Latino children represented a majority of participants in 50% of the contrasts.

5.2. Average treatment effect of ECE on children's externalizing behavior problems (RQ1)

Results from an intercept-only model derived a weighted grand mean effect size of $-.03$ ($SE = .04$) for externalizing behavior problems, which was not significantly different from zero. This suggests that, on average, participants enrolled in the ECE programs did no better or worse on measures of externalizing behavior problems than participants not enrolled in ECE programs.

5.3. Level 2 vs. level 1 ECE programs (RQ2)

In Table 2, we display results of our analyses examining whether level 2 programs were associated with different effects on children's externalizing behavior problems than level 1 programs. Bivariate analyses showed that level 1 programs, or those without

Table 1

Descriptive statistics.

| | Treatment vs. control | | Treatment vs. alternative treatment | |
|---|----------------------------------|------------------------------|-------------------------------------|------------------------------|
| | <i>n</i> = 19 studies | | <i>n</i> = 12 studies | |
| | Mean (<i>SD</i>) or percentage | | Mean (<i>SD</i>) or percentage | |
| | Effect size level <i>n</i> = 96 | Contrast level <i>n</i> = 39 | Effect size level <i>n</i> = 47 | Contrast level <i>n</i> = 16 |
| Level 1 program | 24% | 28% | – | – |
| Level 2 program | 76% | 72% | – | – |
| Level 3 program: Child Social Skills Training | – | – | 38% | 50% |
| Level 3 program: Caregiver Behavior Management Training | – | – | 66% | 56% |
| Program length | 38.37 (21.07) months | 26.39 (18.68) months | 21.4 (8.99) weeks | 19.75 (9.60) weeks |
| ECE hours per week | 22.18 (13.03) | 15.10 (11.22) | – | – |
| Child age at follow up (in months) | 81.14 (70.82) | 60.50 (43.80) | 55.57 (5.08) | 55.84 (5.02) |
| Average % low income | 90.37 (25.52) | 85.96 (31.28) | 86.39 (17.92) | 88.13 (16.52) |
| Average % African American | 65.82 (32.00) | 42.61 (27.70) | 39.11 (26.50) | 49.89 (30.06) |
| Average % Latino | 11.53 (16.07) | 15.50 (17.40) | 14.56 (10.79) | 14.67 (14.06) |
| Majority of participants African American or Latino | 68% | 41% | 40% | 50% |
| High quality study index | 1.80 (.66) | 1.72 (.82) | 1.36 (.79) | 1.34 (.87) |
| Peer refereed publication | 40% | 30% | 68% | 63% |
| Active control group | 34% | 39% | – | – |
| Measure taken during treatment | 14% | 8% | 2% | 3% |
| Measure taken at the end of treatment | 28% | 43% | 92% | 95% |
| Measure taken at follow-up | 58% | 49% | 6% | 2% |
| Missing one or more covariates | 46% | 67% | 21% | 38% |

a clear focus on social and emotional development, had no significant effects on externalizing behavior problems relative to control groups ($ES = .13$ *SD*, $SE = .07$, $p < .10$). On the other hand, level 2 programs, or those with a clear but broad focus on social and emotional development, were significantly associated with modest decreases in externalizing behavior problems relative to control groups ($ES = -.10$ *SD*, $SE = .04$, $p < .05$). In short, these results suggest that level 2 programs were significantly better at reducing externalizing behavior problems than level 1 programs ($ES = -.23$ *SD*, $SE = .09$, $p < .01$) (see Fig. 2). When we added the full set of program, demographic, and study design covariates to the model, we continued to observe a significantly larger reduction in externalizing behavior problems for level 2 programs versus level 1 programs ($ES = -.24$ *SD*, $SE = .11$, $p < .05$).

5.4. Level 3 vs. Level 2 ECE programs (RQ3)

Looking across treatment versus alternative treatment contrasts, we found a weighted grand mean effect size of $-.26$ ($SE = .11$; $p < .05$) in an intercept only model. This suggests that, on average, supplementing level 2 ECE programs with child social skills training

Table 2

Multi-level regression models predicting effect sizes for measures of externalizing behavior problems for level 2 versus level 1 programs.

| | No covariates | With covariates |
|---|--------------------------|-----------------|
| Intercept | .134 (.073) ⁺ | .162 (.231) |
| Level 2 program | -.231 (.085)** | -.243 (.110)* |
| Program length | | .002 (.003) |
| ECE hours per week | | .002 (.006) |
| Child age at follow up (in months) | | -.001 (.001) |
| Majority of participants African American or Latino | | .104 (.149) |
| High quality study index | | .015 (.042) |
| Peer refereed publication | | -.025 (.049) |
| Active control group | | -.162 (.114) |
| Measure taken during treatment | | .007 (.059) |
| Measure taken at the end of treatment | | -.004 (.042) |

Note. ⁺ $p < 0.10$; * $p < 0.05$; ** $p < 0.01$; Standard errors are in parentheses.

Note. The intercept represents the effect of level 1 programs relative to control conditions. The “Level 2 program” coefficient represents the difference in effects between level 1 and level 2 programs.

Note. Coefficients for missing data dummy variables are not displayed.

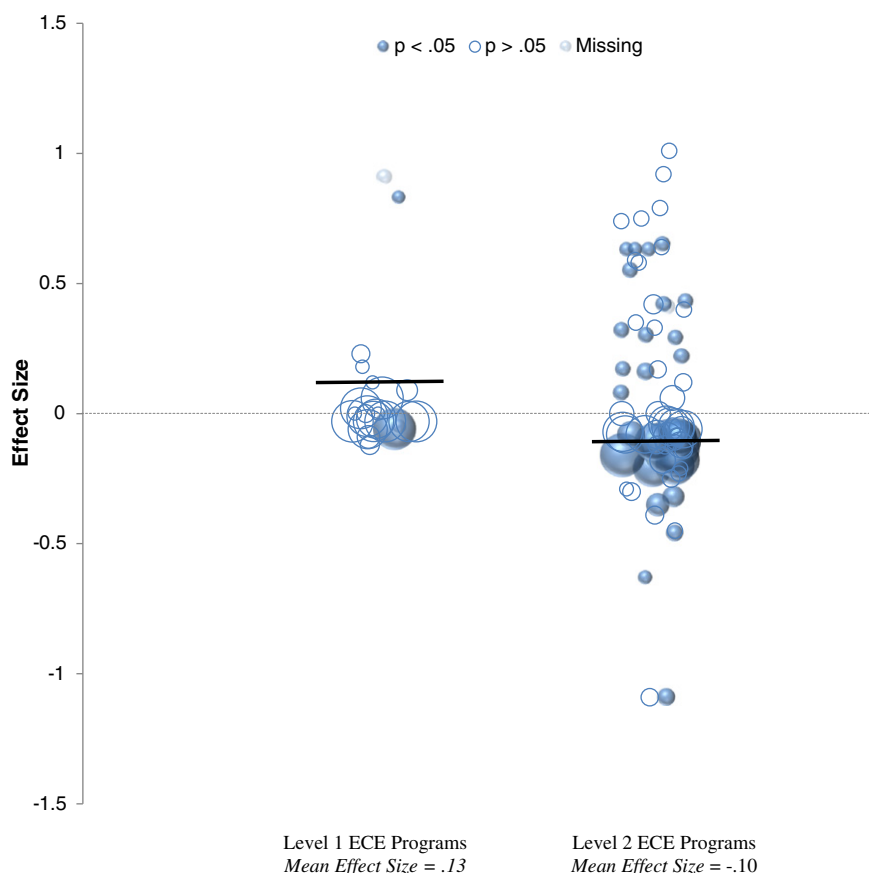


Fig. 2. Externalizing behavior problem effect sizes from level 1 and level 2 ECE programs. Dots represent individual effect sizes, dot sizes represent the square root of samples sizes, and bars represent regression adjusted mean effect sizes for each category. Negative effect sizes represent fewer externalizing behavior problems.

and/or caregiver behavior management training was associated with .26 standard deviations fewer externalizing behavior problems compared to level 2 programs.

5.5. Child social skills training vs. caregiver behavior management training (RQ4)

To explore the relative effectiveness of the two different approaches, we first examined intercept-only models for each. Results are presented in Table 3. Our multi-level intercept-only models suggest that the addition of a child social skills training program enhancement was associated with a reduction of externalizing behavior problems by a standard deviation of .50 relative to level 2 programs ($SE = .03$; $p < .05$). On the other hand, the addition of a caregiver behavior management training program enhancement was not associated with significant reductions in externalizing behavior problems relative to level 2 programs ($ES = -.10$ $SD, SE = .11$). We subsequently employed bivariate and controlled models in order to test whether there was a significant difference between level 3 ECE programs with a child social skills training enhancement and those that only provided caregiver behavior management training (see Table 4). Bivariate analyses revealed a statistically significant and negative (less externalizing behavior problems)

Table 3

Multi-level intercept-only models examining effect sizes for measures of externalizing behavior problems in programs that offered child social skills training and caregiver behavior management training.

| | Child social skills training | Caregiver behavior management training |
|-----------|------------------------------|--|
| Intercept | -.501 (.026)* | -.096 (.106) |

Note. † $p < 0.10$; * $p < 0.05$; ** $p < 0.01$; standard errors are in parentheses.

Note. The intercept in the first column represents the effect of level 3 programs with child social skills training relative to level 2 programs. The intercept in the second column represents the effect of level 3 programs with caregiver behavior management training relative to level 2 programs.

Table 4

Multi-level regression models examining effect sizes for measures of externalizing behavior problems comparing programs that offered child social skills training to those that only offered caregiver behavior management training.

| | No Covariates | With Covariates |
|---|---------------|-----------------|
| Intercept | -.017 (.139) | -.161 (.338) |
| Child social skills training | -.457 (.033)* | -1.045 (.359)* |
| Program length | | .031 (.019) |
| Child age at follow up (in months) | | .002 (.006) |
| Majority of participants African American or Latino | | -.151 (.267) |
| High quality study index | | -.293 (.173) |
| Peer refereed publication | | .139 (.317) |

Note. [†] $p < 0.10$; * $p < 0.05$; ** $p < 0.01$; standard errors are in parentheses.

Note. The intercept represents the effect of level 3 programs with caregiver management training relative to level 2 programs. The “child social skills training” coefficient represents the difference in effects between level 3 programs with caregiver management training and level 3 programs with child social skills training.

Note: Coefficients for missing data dummy variables are not displayed.

relationship between level 3 programs having child social skills training compared to level 3 programs having only caregiver behavior management training ($ES = -.46$, $SE = .03$, $p < .05$). These results are pictorially displayed in Fig. 3. Adding a full set of covariates did not substantively change these relationships.

5.6. Robustness checks

5.6.1. Treatment vs. control analyses

For our treatment versus control comparisons, we carried out several additional analyses to test the sensitivity of our findings that level 2 programs were associated with fewer externalizing behavior problems than level 1 programs (see Table 5).

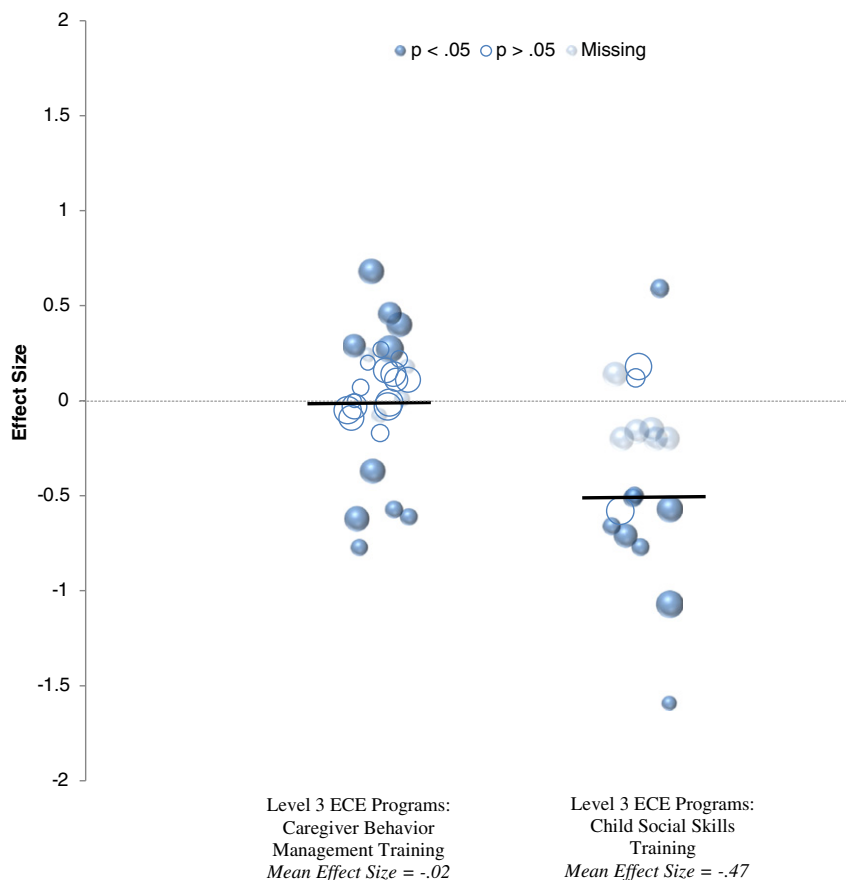


Fig. 3. Externalizing behavior problem effect sizes from level 3 ECE programs that offered child social skills training and those that only offered caregiver behavior management training. Dots represent individual effect sizes, dot sizes represent the square root of samples sizes, and bars represent regression adjusted mean effect sizes for each category. Negative effect sizes represent fewer externalizing behavior problems.

Table 5

Robustness checks: Multi-level regression models predicting effect sizes for measures of externalizing behavior problems for level 2 versus level 1 programs.

| | 10 h/wk + | 4 months long + | Perry Preschool excluded | 1960s studies excluded | Follow-up within 5 years post-treatment | Follow-up within 3 years post-treatment | Follow-up within 1 year post-treatment | High-quality measures |
|-----------------|---------------|--------------------|--------------------------------|---------------------------|---|---|--|--------------------------|
| Intercept | .079 (.084) | .006 (.032) | -.007 (.030) | .002 (.030) | -.025 (.028) | -.025 (.029) | -.037 (.024) | -.027 (.026) |
| Level 2 program | -.177 (.100)+ | -.120 (.036)** | -.105 (.034)** | -.118 (.034)** | -.092 (.032)** | -.093 (.032)** | -.088 (.027)** | -.095 (.029)** |

Note. + $p < 0.10$; * $p < 0.05$; ** $p < 0.01$; standard errors are in parentheses.

Note. The intercept represents the effect of level 1 programs relative to control conditions. The "Level 2 program" coefficient represents the difference in effects between level 1 and level 2 programs.

5.6.1.1. Sample. In our primary analyses, we included ECE programs that varied in the number of hours per week and the length of the program, and controlled for this variation using covariates. However, since it is possible that low-intensity ECE programs (e.g., summer Head Start) might be substantively different from other programs in our sample, we also tested a model that limited the sample to programs that were more than 10 hours a week and a model that limited the sample to programs that were greater than 4 months in length. The pattern of results remained unchanged; however, in the latter model, level 2 programs were only significantly different from level 1 programs at a trend level ($p < .10$).

We were also concerned that our findings might be driven by the Perry Preschool study, which is a level 2 program and is often cited for its robust findings regarding reductions in delinquency and adult crime. Indeed, Perry Preschool's effect sizes included two of the largest in our sample. However, when we removed Perry Preschool from our sample, our pattern of results remained. We further excluded all studies conducted in the 1960's and again found a consistent pattern of results.

5.6.1.2. Length of follow-up. In our main analyses, we display findings that included measures taken at any time point that was measured, including during treatment, at the end of treatment, and at follow-up. Our pattern of findings continued to indicate an advantage for level 2 programs over level 1 programs when we limited our sample to measures taken between two-thirds of the program and up to five years after program completion; taken between two-thirds of the program and up to three years after program completion; and taken between two-thirds of the program and up to 1 year after program completion.

5.6.1.3. Measurement quality. For some measures in our sample, there was little information provided about their development or psychometric properties. Prior research has suggested that researcher developed measures and measures with low reliability can yield different effect sizes than reliable measures and measures that are standardized (Rosenshine, 2010; Shager et al., 2013). To test whether our findings were sensitive to measurement quality, we restricted our sample to measures that were either highly reliable (reliability $> .9$) or nationally normed. Our pattern of results remained unchanged.

5.6.1.4. Treatment vs. alternative treatment analyses. Our treatment versus alternative treatment comparisons had less variability in sample, length of follow-up, and measurement quality, rendering parallel robustness checks unnecessary. However, in our final presented analyses, we ultimately excluded three studies (including 5 contrasts and 60 effect sizes) because it was not clear whether or not their ECE counterfactual or comparison conditions were level 2 programs, and we were interested primarily in how level 2 programs could be further improved. This set-up provided the most coherent set of analyses. Yet, we were still interested in whether the addition of these three excluded studies would change our findings. The studies included an additional evaluation of the Incredible Years Basic Program; an evaluation of preschool with RECAP ("Reaching Educators, Children, and Families"); and the interpersonal and cognitive problem solving intervention developed by Shure and Spivack (1980, 1982). With these studies added to analyses, child social skills training remained the only program enhancement that was significant relative to level 2 programs in intercept-only models. In the regression analyses comparing child social skills training to caregiver behavior management training, the coefficient for child social skills training became slightly larger and more significant ($p < .01$).

5.7. Publication bias

The funnel plot for treatment versus control contrasts is presented in Fig. 4, and the one for treatment versus alternative treatment contrasts is presented in Fig. 5. Each plot produced reasonably symmetric funnel patterns, and no significant asymmetry was detected using Egger's regression test ($t(37) = 1.70, p = .10$ for treatment versus control contrasts; $t(14) = 1.34, p = .20$ for treatment versus alternative treatment contrasts). In computing fail-safe Ns, we found that 369 contrasts with non-significant effects would need to be added to the dataset to nullify the beneficial effect of level 2 programs over level 1 programs. Seventy-three contrasts with non-significant effects would need to be added to the dataset to nullify the beneficial effect of child social skills training. Both findings are robust against publication bias based on Rosenthal's guidelines that the fail-safe N is unlikely when it is greater than or equal to five times the number of contrasts plus 10.

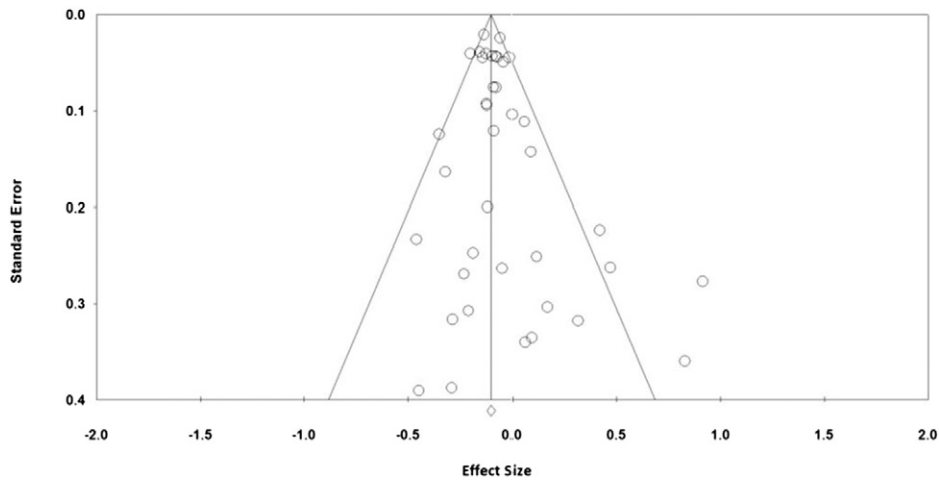


Fig. 4. Funnel plot of mean effect sizes in treatment versus control contrasts.

6. Discussion

In this meta-analytic paper, we set out to examine the relative effects of different practices used within ECE settings on externalizing behavior problems, using a meta-analysis of rigorously evaluated ECE programs over the last several decades. Since meta-analyses rely on information presented in existing studies, they are particularly well-suited for identifying overarching patterns that can help shape future work. Specifically, we examined three levels of ECE practice, each with increasing specificity and intensity aimed at children's social and emotional development. In short, we found that each successive level of programs did a better job than the prior level at reducing externalizing behavior problems (see Fig. 6 for a summary of key findings).

Level 1 programs, or those without a clear focus on social and emotional development, were not significantly associated with externalizing behavior problems relative to control groups. On the other hand, Level 2 programs, or those with a clear but broad focus on social and emotional development, were associated with significant, modest reductions in externalizing behavior problems relative to control groups. These findings suggest that under conditions aiming to actively support social and emotional development, ECE can be an effective mechanism for reducing externalizing behavior problems. Although it would be instructive to statistically untangle the mechanisms through which a broad focus on social and emotional development translated into improved outcomes in level 2 classrooms, only a few studies in our database explicitly delineated how this objective was infused into regular practices and processes. In many cases, level 2 programs utilized a global curriculum, such as High Scope or Creative Curriculum, that incorporated content across social, emotional, cognitive, and physical development.

Our findings also suggest that level 2 ECE programs may be able to significantly and meaningfully build upon their impacts through the addition of program enhancements that more intensively target children's social and emotional development. These benefits were largely driven by programs targeting children's social skills (what we refer to as Level 3a in our conceptualization). In our

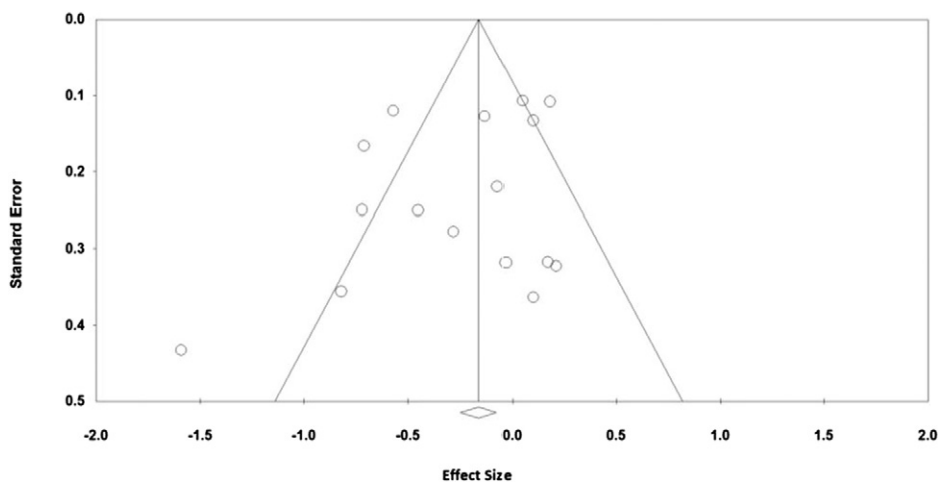


Fig. 5. Funnel plot of mean effect sizes in treatment versus alternative treatment contrasts.

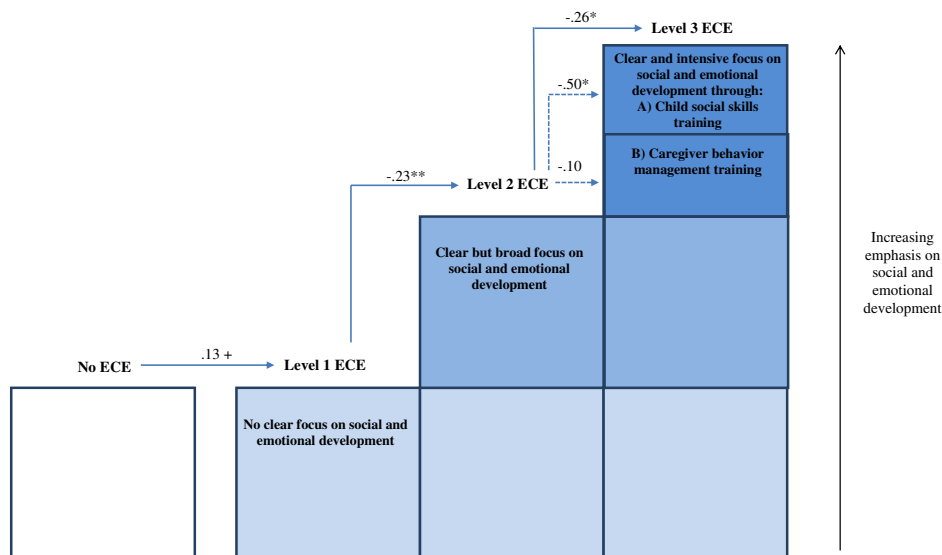


Fig. 6. ECE and the prevention of externalizing behavior problems: summary of key findings. Negative effect sizes represent fewer externalizing behavior problems. $+p < 0.10$; $*p < 0.05$; $**p < 0.01$.

database, social skills training enhancements 1) focused on giving children skills for interacting with others, such as making friends, accepting others, or expressing individual needs, and 2) targeted basic cognitive skills for social problem solving, such as perspective taking and self-control. Child social skills trainings were implemented through developmentally focused and domain specific curricula, meaning they sought to build skills within the bounded realm of social skills rather than through a broader focus encompassing multiple domains of development (Clements, 2007; Yoshikawa et al., 2013). The addition of a child social skills training enhancement to a level 2 program (making it a level 3a program), resulted in half of a standard deviation reduction in externalizing behavior problems relative to level 2 programs, which is nearly twice as large as the effect found in a previous study for high-quality social and emotional learning programs implemented in primary and secondary schools (Durlak, Dymnicki, Taylor, Weissberg, & Schellinger, 2011).

In contrast to the findings just described, caregiver behavior management trainings (what we refer to as level 3b) were not significantly related to additional benefits for externalizing behavior problems relative to level 2 programs. These results are consistent with extensive research that has found that changing adult behavior, especially habitual patterns of caregiver–child interactions, is difficult (Halpern, 2000; Grindal et al., under review). Furthermore, both parent and teacher behavior management training programs that are added onto ECE programs rest on the assumption that parents and teachers will be able to apply the skills they learn in their training sessions in a variety of situations and contexts after the trainings have ended. Child social skills training may be more effective than caregiver behavior management trainings in part because the former directly provide a vehicle for children to acquire a range of critical capabilities for learning how to calm down, how to recognize their own emotions and those of others, and how to negotiate conflict – all of which they may be more likely to be able to carry with them across settings. It may also be that early social skills are particularly salient factors in protecting children against heightened externalizing behavior problems, whereas behavior management strategies employed by parents and teachers are less powerful protective factors in the early childhood years. In sum, our findings largely support our initially proposed framework. At the third level, we find that child social skills training programs are particularly promising.

Still, a number of limitations of our analyses warrant further discussion. First, we were only able to analyze data from programs that have been rigorously evaluated and that measured externalizing behavior problems as an outcome. Consequently our analyses only represent a subset of ECE programs. Second, although we found that child social skills training produced the most promising reductions in externalizing behavior problems, the limitations of using meta-analytic data precluded our ability to identify the features that were most effective. For example, recent research has shown that in-service, observation coaching and mentoring can boost the effects of a curriculum above and beyond didactic training alone (Neuman & Cunningham, 2009). Our database does not provide information concerning the specific levels of coaching versus didactic training provided to teachers in each program. Third, we focused on externalizing behavior problems as opposed to examining a broader set of behavior problems.

One recent large-scale study asking questions similar to the questions examined in our meta-analysis has attempted to address some of these limitations of the broader literature. Specifically, the Head Start CARES study simultaneously evaluated the effects of a child social skills training enhancement (Head Start plus Preschool PATHS) and a caregiver behavior management enhancement (Head Start plus Incredible Years) relative to Head Start as usual (see Domitrovich et al., 2007 and Webster-Stratton, Reid, & Hammond, 2001 for details on PATHS and Incredible Years). Both enhancements included comparable levels of coaching support, and the evaluation measured a broad set of social and emotional outcomes. In spite of examining similar questions, the Head Start CARES study came to somewhat different conclusions than our meta-analysis (Morris et al., 2014). Neither enhancement had effects

on externalizing behavior problems when measured in spring of the preschool year relative to level 2 programs. However, the enhancements did impact other social and emotional outcomes, including knowledge of emotions, social problem skills, and related social behavior, with the child social skills training enhancement producing a wider and more consistent set of impacts (Morris et al., 2014). It is possible that this study came to somewhat different conclusions because of the challenges in scaling enhancements to larger populations. It is also possible that the early effects on social and emotional outcomes will translate into reductions in externalizing behavior problems as children get older.

Future research should continue replicating our analyses with more contemporary national ECE studies, explore the circumstances under which child social skills trainings are more or less effective at preventing externalizing behavior problems, and examine how much implementation support and dosage is needed to achieve the greatest benefits at scale.

In spite of its limitations, the present study has the benefit of looking across a number of high-quality evaluations over several decades. As ECE programs continue to expand nationwide, our results suggest that strategies employed to reduce externalizing behavior problems should be given greater attention. Federal and state ECE programs should carefully consider whether chosen curricula include social and emotional development as an aim. Individual ECE preschool programs with concerns about externalizing behavior problems in children should also consider adopting an existing evidence-based social skills training program.

Though social and emotional development has been given less attention than cognitive gains in the recent literature on ECE and school readiness, the early prevention of externalizing behavior problems is critical for promoting later school and life success. This meta-analytic study suggests that when ECE programs implement intentional interventions focused explicitly on social and emotional development, they can make a difference.

Role of funding source

We are grateful to the following funders of the National Forum on Early Childhood Policy and Programs: the Birth to Five Policy Alliance, the Buffett Early Childhood Fund, Casey Family Programs, the McCormick Tribune Foundation, the Norlien Foundation, Harvard University, and an Anonymous Donor. We are also grateful to the Institute of Education Sciences, US Department of Education for supporting this research (#R305A110035), and to Abt Associates, Inc. and the National Institute for Early Education Research for making their data available to us.

Appendix A

Level 1 ECE programs: No clear focus on social and emotional development

| Start year | Program name | Program description | Preschool versus infant/toddler program | Effect size min and max |
|-------------------|---|--|---|-------------------------|
| 1980 ^a | Infant Care | This study evaluated use of general infant ECE. There was no uniformly stated goal to improve social and emotional development or requirement to use a particular curriculum. | Infant/toddler | .00 to .18 |
| 1983 ^a | Quality Toddler ECE | Children were enrolled in ECE centers meeting specific requirements for adult to child ratios, staff formal training, and low staff turnover. There was no uniformly stated goal to improve social and emotional development or requirement to use a particular curriculum. | Infant/toddler | -.05 to .91 |
| 1989 | Quality Infant and Toddler ECE | Children were enrolled in ECE centers meeting NAEYC's 1984 criteria for staff-child interactions, staff qualifications, and staffing patterns. There was no uniformly stated goal to improve social and emotional development or requirement to use a particular curriculum. | Infant/toddler | .86 |
| 1991 | Comprehensive Child Development Program | This program provided comprehensive services to families. Case managers provided some services directly to parents and also brokered arrangements with local providers for additional services. Parent-based services included prenatal care, health care, nutrition, referral to education, employment counseling, and assistance in securing income support, health care, and housing. The core services provided to children included health screening, treatment and referral, immunizations, and ECE services. The primary stated goals of the program were to improve child cognitive and school outcomes, improve child health, improve parent outcomes, and improve children's settings. | Both | -.06 to .06 |
| 1991 | North Carolina Pre-Kindergarten | This state pre-k study was made up a variety of ECE programs, most of which provided a combination of ECE and parenting education. There was no uniformly stated goal to improve social and emotional development or requirement to use a particular curriculum. | Preschool | -.12 to .09 |
| 1997 | Pre-Kindergarten in ECLS-B (591) | This study used a nationally representative dataset to examine the effects of participating in school-based and publicly-funded ECE programs in the year prior to kindergarten. There was no uniformly stated goal to improve social and emotional development or requirement to use a particular curriculum. | Preschool | -.06 |

Appendix A (continued)

| Start year | Program name | Program description | Preschool versus infant/toddler program | Effect size min and max |
|------------|---|--|---|-------------------------|
| 1999 | Even Start Literacy Program Preschool Component (737) | Even Start was a family literacy initiative with the primary goal of improving the academic achievement of low-income children and their parents, with a focus on reading. They attempted to accomplish this goal through offering a combination of ECE services, adult education, parenting education, and parent-child joint activities. | Preschool | .00 to .23 |

^aEstimated by subtracting three years from the publication date.

Appendix B

Level 2 ECE programs: clear but broad focus on social and emotional development

| Start year | Program name | Program description | Preschool versus infant/toddler program | Effect size min and max |
|-------------------|---|--|---|-------------------------|
| 1962 | Perry Preschool | Perry Preschool utilized the High/Scope curriculum, which aims to improve child outcomes across a range of developmental domains, including social and emotional development. | Preschool | −1.09 to .16 |
| 1962 | Cambridge, Massachusetts Preschool Intervention | This intervention aimed to improve intelligence and social skills. | Preschool | −.29 to −.21 |
| 1965 | Lubbock, Texas Summer Head Start | Head Start's stated purpose is to promote school readiness by enhancing children's cognitive, social, and emotional development. How Head Start's social and emotional goals translate into practice varies from site to site. | Preschool | .06 |
| 1966 | Head Start Retrospective Study | Head Start's stated purpose is to promote school readiness by enhancing children's cognitive, social, and emotional development. How Head Start's social and emotional goals translate into practice varies from site to site. | Preschool | −.45 to .42 |
| 1969 | Syracuse University Family Development and Research | This program implemented a curriculum based on the work of Piaget and Erikson that broadly targeted children's language, cognitive, and social/ emotional skills. They aimed to, "give very little children experiences and choices that would help them to develop a concern for their needs and rights in relation to the needs and rights of others." | Both | −.63 to .63 |
| 1972 | Abecedarian Project | Abecedarian combined an emphasis on language and cognitive skills with a focus on supporting developing social skills. | Both | −.30 to 1.01 |
| 1973 | Child and Family Resource Program (CFRP) | CFRP's stated goal was to, "assist parents to promote the total development (emotional, cognitive, language, and physical) of infants and toddlers." They provided children opportunities to gain social skills in the center-based settings and provided parents with support in learning relevant parenting skills. | Both | .00 |
| 1992 | Kentucky Educational Reform Act Preschool Program | Kentucky preschools were required to utilize a clear curriculum that promoted both intellectual and social and emotional development. Most sites utilized High/Scope or Creative Curriculum, which both aim to improve child outcomes across a range of developmental domains, including social and emotional development. | Preschool | −.35 to −.32 |
| 1996 | Early Head Start Research and Evaluation Project | Early Head Start, like Head Start, aims to promote school readiness by enhancing children's cognitive, social, and emotional development. How Early Head Start's social and emotional goals translate into practice varies from site to site. | Infant/toddler | −.18 to −.05 |
| 1997 | Early Childhood Longitudinal Study – Kindergarten Cohort Head Start Study | Head Start's stated purpose is to promote school readiness by enhancing children's cognitive, social, and emotional development. How Head Start's social and emotional goals translate into practice varies from site to site. | Preschool | −.16 to −.13 |
| 2002 | National Head Start Impact Study, First Year | Head Start's stated purpose is to promote school readiness by enhancing children's cognitive, social, and emotional development. How Head Start's social and emotional goals translate into practice varies from site to site. In this national study, most sites utilized either High/Scope or Creative Curriculum, which both aim to improve child outcomes across a range of developmental domains, including social and emotional development. | Preschool | −.21 to −.04 |
| 2003 ^a | California Head Start Follow-Up | Head Start's stated purpose is to promote school readiness by enhancing children's cognitive, social, and emotional development. How Head Start's social and emotional goals translate into practice varies from site to site. In this study, all of the Head Start sites utilized High/ Scope, which aims to improve child outcomes across a range of developmental domains, including social and emotional development. | Preschool | −.17 |

^aEstimated by subtracting three years from the publication date.

Appendix C

Level 3 ECE programs: clear and intensive focus on social and emotional development through child social skills training and caregiver behavior management training

| Start year | Name | Description of child social skills training | Description of caregiver behavior management training | Description of counterfactual | Preschool versus infant/toddler program | Effect size min and max |
|-------------------|--|--|---|--|---|-------------------------|
| 1993 | Standard Head Start <i>plus</i> PARTNERS Parent and Teacher Training | | Parents participated in 8 to 9 weeks of parent groups, during which they discussed short videotapes modeling parent–child interactions. Topics included using praise and encouragement, effective limit setting, and handling misbehavior. Teachers received 2-day workshops that mirrored parent groups in content. | Standard Head Start | Preschool | –.05 to .27 |
| 1994, 1995, 1996 | Standard Head Start <i>plus</i> Al's Pals Curriculum | Al's Pals Curriculum is a resiliency- based curriculum designed to provide real-life situations that introduce children to health-promoting concepts and build prosocial skills. Two lessons lasting 15–20 min were taught each week over a 23-week period and focused on topics such as understanding feelings, expressing feelings appropriately, caring about others, accepting differences, establishing and maintaining social relationships, brainstorming ideas, thinking flexibly, using self-control, and solving problems. | | Standard Head Start | Preschool | – 1.07 to – .57 |
| 1995 ^a | Center Care <i>plus</i> Taking Part Curriculum | Taking Part is a 15-week social skills curriculum. Children were given lessons and scaffolded in order to teach skills such as listening to others, speaking kindly, developing group play skills, and resolving conflicts. | | Center care in which children were taught the same social skills, but without a structured curriculum. | Preschool | – 1.59 |
| 1996 | Standard Head Start <i>plus</i> Second Step Curriculum | Second Step is a violence-prevention curriculum focused on teaching empathy, perspective taking, impulse control, and problem solving skills. This 28-week curriculum was implemented during circle time. Children were shown cards depicting various emotions and social dilemmas that facilitated discussions and role playing. | | Standard Head Start | Preschool | –.51 to .59 |
| 1997 | Standard Head Start <i>plus</i> Incredible Years Parent and Teacher Training | | Parents participated in 12-week parent groups (the “Basic Incredibly Years Program”) that taught positive discipline strategies, effective parenting skills, and strategies for coping with stress. Groups viewed videotapes of model parenting skills and discussed parent–child interactions with the support of a group leader. Head Start teachers and teaching assistants participated in six monthly 1-day workshops. Workshops emphasized classroom-wide positive management and discipline strategies. The same video-vignette method was used. | Standard Head Start | Preschool | –.62 to .68 |

| | | | | | | |
|-------------------|---|--|---|---|-----------|----------------------------|
| 1997 | Standard Head Start <i>plus</i> Self-Administered Parent and Teacher Training (manual-based) | | Parents and teachers were instructed to read the manual and implement the outlined procedures. Manual components included setting behavioral goals, rewarding positive behavior, ignoring techniques, and using time-out procedures. Consultants were available for questions but provided minimal up-front direction on using the manual. | Standard Head Start | Preschool | .00 to .20 |
| 1997 | Standard Head Start <i>plus</i> Self-Administered Parent and Teacher Training (video-based) | | Parents and teachers were instructed to watch the videos and implement procedures outline in the accompanying manual. The videotape program topics included praise and rewards, limit setting, and how to handle misbehavior. Consultants were available for questions but provided minimal up-front direction on using the videotape series. | Standard Head Start | Preschool | .07 to .27 |
| 1999 | High/Scope Preschool <i>plus</i> Making Choices Curriculum and Strong Families Parent Program | The Making Choices curriculum was comprised of 30 lessons structured around learning to collaborate, recognize emotional cues, use self-talk, distinguish between anti- and pro-social behaviors, and predict consequences of actions. | The Strong Families portion of the intervention aimed to increase positive, effective parenting skills and to decrease the frequency of coercive parenting styles. | High/Scope preschool | Preschool | -.77 to -.66 |
| 2000 | Standard Head Start <i>plus</i> Effective Black Parenting Program (EBP) | | EBP is a culturally adapted program for parents of African American children to build participants' parenting skills. The program included eight 3-hour sessions aimed at helping parents reduce behavior problems. | Standard Head Start | Preschool | -.61 to -.17 |
| 2003 ^a | Standard Head Start <i>plus</i> Good Behavior Game | | In this intervention, teachers were trained to use the "Good Behavior Game," a classroom-wide behavioral management strategy that utilized reinforcements natural to the classroom environment for the inhibition of negative behavior. | Standard Head Start | Preschool | -.08 to .01 |
| 2003 ^a | Standard Head Start <i>plus</i> Behavioral Consultation | | This intervention included 6 weeks of behavioral consultation for teachers. In consultations, teachers were trained to identify behavior problems, design intervention plans, implement plans, and evaluate implementation. | Standard Head Start | Preschool | .18 to .24 |
| 2003 ^a | Standard Head Start <i>plus</i> Promoting Alternative Thinking Strategies (PATHS) Curriculum | The PATHS curriculum used in this study included 30 lessons delivered weekly during circle time. Thematic units included lessons on compliments, basic and advanced feelings, a self-control strategy, and problem solving. Detailed extension activities were also included in the curriculum. | | Standard Head Start | Preschool | -.20 to .14 |
| 2004 ^a | Standard Head Start <i>plus</i> Head Start REDI | The Preschool PATHS Curriculum was used to promote children's social-emotional skills through 33 lessons delivered at circle time, as well as through extension activities. The PATHS curriculum targeteded prosocial friendship skills; emotional understanding and emotional expression; self-control; and problem solving skills, including interpersonal negotiation and conflict resolution. Language and emergent literacy skills were also targeted through an interactive reading program, "sound games," and print center activities. | | Standard Head Start implementing either High/Scope or Creative Curriculum | Preschool | .18 |
| 2005 | Standard Head Start <i>plus</i> Child-Teacher Relationship Training (CTRT) | | CTRT, which was based on parent-child relationship therapy, trained teachers how to create non-judgmental, understanding, and accepting environments; recognize and respond to children's feelings; and set appropriate limits. Teachers were trained in 10 weekly sessions during Phase I, followed by 10 weekly in-class coaching and modeling in PHASE II. | Standard Head Start | | -.77 to .22 -.77 to .22 |

^aEstimated by subtracting three years from the publication date.

References

- Achenbach, T. M. (1992). *Manual for the child behavior checklist*. Burlington, VT: University of Vermont Department of Psychiatry.
- Achenbach, T. M., & Edelbrock, C. S. (1978). The classification of child psychopathology: A review and analysis of empirical efforts. *Psychological Bulletin*, *85*, 1275–1301.
- Allison, P. D. (2002). *Missing data*. Thousand Oaks, CA: Sage.
- Borenstein, M., Hedges, L., Higgins, J., & Rothstein, H. (2005). *Comprehensive meta-analysis, version 2*. Englewood, NJ: Biostat.
- Borenstein, M., Hedges, L., Higgins, J., & Rothstein, H. (2009). *Introduction to meta-analysis*. Chichester, UK: Wiley.
- Bronfenbrenner, U. (1977). Toward an experimental ecology of human development. *American Psychologist*, *32*, 512–531.
- Bub, K. L., McCartney, K., & Willett, J. B. (2007). Behavior problem trajectories and first grade cognitive ability and achievement skills: A latent growth curve analysis. *Journal of Educational Psychology*, *3*, 653–670.
- Bulotsky, R. J., & Fantuzzo, J. W. (2011). Preschool behavior problems in classroom learning situations and literacy outcomes in kindergarten and first grade. *Early Childhood Research Quarterly*, *26*, 61–73.
- Burgess, K., Chien, N., Morrissey, T., & Swenson, K. (2014). *Trends in the use of early care and education, 1995–2011: Descriptive analysis of child care arrangements from national survey data*. Washington, DC: U.S. Department of Health and Human Services, Office of the Assistant Secretary for Planning and Evaluation.
- Bushman, B. J., & Wang, M. C. (2009). Vote-counting procedures in meta-analysis. In H. Cooper, & L. V. Hedges (Eds.), *The handbook of research synthesis and meta-analysis* (pp. 207–220) (2nd ed.). New York, NY: Russell Sage Foundation.
- 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.
- Campbell, S. (1995). Behavior problems in preschool children: A review of recent research. *Journal of Child Psychology and Psychiatry*, *36*, 113–149.
- Campbell, S., Shaw, D. S., & Gilliom, M. (2000). Early externalizing behavior problems: Toddlers and preschoolers at risk for later maladjustment. *Developmental Psychopathology*, *12*, 467–488.
- Campbell, S., Spieker, S., Burchinal, M., Poe, M. D., & NICHD Early Child Care Research Network (2006). Trajectories of aggression from toddlerhood to age 9 predict academic and social functioning through age 12. *Journal of Child Psychology and Psychiatry*, *47*, 791–800.
- Clements, D. H. (2007). Curriculum research: Toward a framework for “research-based curricula.” *Journal for Research in Mathematics Education*, *38*, 35–70.
- Coley, R. L., Votruba-Drzal, E., Miller, P. L., & Koury, A. (2013). Timing, extent, and type of child care and children's behavioral functioning in kindergarten. *Developmental Psychology*, *49*, 1859–1873.
- Cooper, H., Hedges, L. V., & Valentine, J. C. (2009). *The handbook of research synthesis and meta-analysis* (2nd ed.). New York, NY: Russell Sage Foundation.
- Crick, N. R., & Dodge, K. A. (1994). A review and reformulation of social-information-processing mechanisms in children's social adjustment. *Psychological Bulletin*, *115*, 74–101.
- de la Torre, J., Camilli, G., Vargas, S., & Vernon, R. F. (2007). Illustration of a multilevel model for meta-analysis. *Measurement and Evaluation in Counseling and Development*, *40*, 169–180.
- Domitrovich, C. E., Cortes, R. C., & Greenberg, M. T. (2007). Improving young children's social and emotional competence: A randomized trial of the preschool “PATHS” curriculum. *The Journal of Primary Prevention*, *28*, 67–91.
- Duncan, G. J., & Magnuson, K. A. (2011). The nature and impact of early achievement skills, attention skills, and behavior problems. In G. J. Duncan, & R. J. Murnane (Eds.), *Whither opportunity: Rising inequality, schools, and children's life chances* (pp. 47–69). New York, NY: Russell Sage.
- Durlak, J. A., Dymnicki, A. B., Taylor, R. D., Weissberg, R. P., & Schellinger, K. B. (2011). The impact of enhancing students' social and emotional learning: A meta-analysis of school-based universal interventions. *Child Development*, *82*, 405–432.
- Fantuzzo, J. W., Bulotsky-Shearer, R., Frye, D., McDermott, P. A., McWayne, C., & Perlman, S. (2007). Investigation of social, emotional, and behavioral dimensions of school readiness for low-income, urban preschool children. *School Psychology Review*, *36*, 44–62.
- Farrington, D. P., & Welsh, B. P. (2007). *Saving children from a life of crime: Early risk factors and effective interventions*. Oxford, UK: Oxford University Press.
- Fox, L., Dunlap, G., Hemmeter, M. L., Joseph, G., & Strain, P. (2003). The Teaching Pyramid: A model for supporting social competence and preventing challenging behavior in young children. *Young Children*, *58*, 48–53.
- Gormley, W. T., Phillips, D. A., Newmark, K., Welti, K., & Adelstein, S. (2012). Social-emotional effects of early childhood education programs in Tulsa. *Child Development*, *82*, 2095–2109.
- Grindal, T., Bowne, J., Yoshikawa, H., et al. (2015). *The added impact of parenting education in early childhood education programs: A meta-analysis*. (under review).
- Halpern, R. (2000). Early intervention for low-income children and families. In J. P. Shonkoff, & S. J. Meisels (Eds.), *Handbook of early childhood intervention* (pp. 361–386). New York, NY: Cambridge University Press.
- Hinshaw, S. P. (2002). Process, mechanism, and explanation related to externalizing behavior in developmental psychopathology. *Journal of Abnormal Child Psychology*, *30*, 431–446.
- Hirschi, T. (1969). *Causes of delinquency*. London, England: University of California Press.
- High Scope Educational Research Foundation (2014). Retrieved from <http://www.highscope.org/Content.asp?ContentId=63>.
- Houts, R. M., Caspi, A., Pianta, R. C., Arseneault, L., & Moffitt, T. E. (2010). The challenging pupil in the classroom: The effect of child on teacher. *Psychological Science*, *12*, 1802–1810.
- Howes, C., Burchinal, M., Pianta, R., et al. (2008). Ready to learn? Children's pre-academic achievement in pre-kindergarten classrooms. *Early Childhood Research Quarterly*, *23*, 27–50.
- Huaqing Qi, C., & Kaiser, A. P. (2003). Behavior problems of preschool children from low-income families: Review of the literature. *Topics in Early Childhood Special Education*, *23*, 188–216.
- Jacob, R. T., Creps, C. L., & Boulay, B. (2004). *Meta-analysis of research and evaluation studies in early childhood education*. Cambridge, MA: Abt Associates Inc.
- Keenan, K., & Wakschlag, L. (2000). More than terrible twos: The nature and severity of disruptive behavior problems in clinically-referred preschool children. *Journal of Abnormal Child Psychology*, *28*, 33–46.
- Kuppens, S., Laurent, L., Heyvaert, M., & Onghena, P. (2013). Associations between parental psychological control and relational aggression in children and adolescents: A multilevel and sequential meta-analysis. *Developmental Psychology*, *49*, 1697–1712.
- Layzer, J. I., Goodson, B. D., Bernstein, L., & Price, C. (2001). *National evaluation of family support programs, volume A: The meta-analysis, final report*. Cambridge, MA: Abt Associates Inc.
- Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks, CA: Sage Publications.
- Loeb, S., Bridges, M., Bassok, D., Fuller, B., & Rumberger, R. W. (2007). How much is too much? The influence of preschool centers on children's social and cognitive development. *Economics of Education Review*, *26*, 52–66.
- Loeb, S., Fuller, B., Kagan, L., & Carrol, B. (2004). Child care in poor communities: Early learning effects of type, quality, and stability. *Child Development*, *1*, 47–65.
- Loeber, R., Farrington, D. P., & Waschbusch, D. A. (1998). Serious and violent juvenile offenders. In R. Loeber, & D. P. Farrington (Eds.), *Serious and violent juvenile offenders* (pp. 13–29). Thousand Oaks, CA: Sage.
- Losel, F., & Bender, D. (2003). Resilience and protective factors. In D. P. Farrington, & J. W. Coid (Eds.), *Prevention of adult antisocial behavior*. Cambridge, UK: Cambridge University Press.
- Losel, F., & Bender, D. (2006). Risk factors for serious and violent antisocial behavior in children and youth. In A. Hagell, & R. Jeyarajah-Dent (Eds.), *Children who commit acts of serious interpersonal violence: Messages for best practice*. London, UK: Jessica Kingsley Publishers.
- Losel, F., & Bender, D. (2012). Child social skills training in the prevention of antisocial development and crime. In D. Farrington, & B. C. Welsh (Eds.), *The Oxford handbook of crime prevention* (pp. 102–129). Oxford, UK: Oxford University Press.
- Magnuson, K., Ruhm, C., & Waldfogel, J. (2007). The persistence of preschool effects: Do subsequent classroom experiences matter? *Early Childhood Research Quarterly*, *22*, 18–38.

- Mattera, S., Lloyd, C. M., Fishman, M., & Bangser, M. (2013). *A first look at the Head Start CARES demonstration: Large-scale implementation of programs to improve children's social-emotional competence*. Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.
- Moffit, T. E. (1993). Adolescence-limited and life-course-persistent antisocial behavior: A developmental taxonomy. *Psychology Review*, 4, 674–701.
- Morris, P., Mattera, S. K., Castells, N., Bangser, M., Bierman, K., & Raver, C. (2014). *Impact findings from the Head Start CARES Demonstration: National evaluation of three approaches to improving preschoolers' social and emotional competence*. Washington, DC: Office of Planning Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.
- National Association for the Education of Young Children, & National Council of Teachers of Mathematics (2002). Math experiences that count! *Young Children*, 57, 60–62.
- National Scientific Council on the Developing Child (2007). *The science of early childhood development: Closing the gap between what we know and what we do*.
- Nelson, G., Westhues, A., & MacLeod, J. (2003). A meta-analysis of longitudinal research on preschool prevention programs for children. *Prevention and Treatment*, 6, 1–35.
- Neuman, S. B., & Cunningham, L. (2009). The impact of a practice-based approach to professional development: Coaching makes a difference. *American Educational Research Journal*, 46, 532–566.
- Patterson, G. R., DeBaryshe, B. D., & Ramsey, E. (1989). A developmental perspective on antisocial behavior. *American Psychologist*, 44, 329–335.
- Patterson, G. R., Reid, J. B., & Dishion, T. J. (1992). *Antisocial boys: A social interactional approach*. Eugene, OR: Castalia.
- Piquero, A. R., & Jennings, W. G. (2012). Parent training and the prevention of crime. In D. P. Farrington, & B. C. Welsh (Eds.), *The Oxford handbook of crime prevention* (pp. 89–101). New York, NY: Oxford University Press.
- Puma, M. J., Olsen, R. B., Bell, S. H., & Price, C. (2009). *What to do when data are missing in group randomized trials*. Cambridge, MA: Abt Associates.
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods, Vol. 1*. Thousand Oaks, CA: SAGE Publications, Incorporated.
- Raver, C. C. (2002). Emotions matter: Making the case for the role of young children's emotional development for early school readiness. *Social Policy Report*, 16, 3–19.
- Rosenshine, B. (2010). Researcher-developed tests and standardized tests: A review of ten meta-analyses. *Paper presented at the Society of Research on Educational Effectiveness conference, Washington, DC*.
- Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin*, 86, 638–641.
- Schindler, H. S., & Yoshikawa, H. (2012). Preventing crime in the preschool years. In D. P. Farrington, & B. C. Welsh (Eds.), *The Oxford handbook of crime prevention* (pp. 70–88). New York, NY: Oxford University Press.
- Schweinhart, L. J., Montie, J., Xiang, Z., Barnett, W. S., Belfield, C. R., & Nores, M. (2005). *Lifetime effects: The High/Scope Perry Preschool study through age 40*. Ypsilanti, MI: High/Scope Press.
- Shager, H. M., Schindler, H. S., Magnuson, K. A., Duncan, G. J., Yoshikawa, H., & Hart, C. M. D. (2013). Can research design explain variation in Head Start research results? A meta-analysis of cognitive and achievement outcomes. *Education Evaluation and Policy Analysis*, 35, 76–95.
- Shaw, D. S., Lacourse, E., & Nagin, D. S. (2005). Developmental trajectories of conduct problems and hyperactivity from ages 2 to 10. *Journal of Child Psychology and Psychiatry*, 46, 931–942.
- Shure, M. B., & Spivack, G. (1980). Interpersonal problem solving as a mediator of behavioral adjustment in preschool and kindergarten children. *Journal of Applied Developmental Psychology*, 1, 29–44.
- Shure, M. B., & Spivack, G. (1982). Interpersonal problem-solving in young children: A cognitive approach to prevention. *American Journal of Community Psychology*, 10, 341–356.
- Smith, J. D., Dishion, T. J., Shaw, D. S., Wilson, M. N., Winter, C., & Patterson, G. (2014). Coercive family process and early onset conduct problems from age 2 to school entry. *Development and Psychopathology*, 26, 917–932.
- Snell, M. E., Voorhees, M. D., Berlin, R. A., Stanton-Chapman, T. L., Hadden, S., & McCarty, J. (2012). Use of interview and observation to clarify reported practices of Head Start staff concerning problem behavior implications for programs and training. *Journal of Positive Behavior Interventions*, 14, 108–117.
- Snow, C. E., Burns, M. S., & Griffin, P. (Eds.). (1998). *Preventing reading difficulties in young children*. Washington, DC: National Research Council, National Academy Press.
- Sterne, J. A., & Egger, M. (2001). Funnel plots for detecting bias in meta-analysis: Guidelines on choice of axis. *Journal of Clinical Epidemiology*, 54, 1046–1055.
- Tolan, P. H., Dodge, K., & Rutter, M. (2013). Tracking the multiple pathways of parent and family influence on disruptive behavior disorders. In P. H. Tolan, & B. L. Leventhal (Eds.), *Disruptive behavior disorders* (pp. 161–191). New York, NY: Springer.
- Tremblay, R. E. (2010). Developmental origins of disruptive behavior problems: The "original sin" hypothesis, epigenetics, and their consequences for prevention. *Journal of Child Psychology and Psychiatry*, 51, 341–367.
- Tremblay, R. E., Japel, C., Perusse, D., et al. (1999). The search for the age of 'onset' of physical aggression: Rousseau and Bandura revisited. *Criminal Behaviour and Mental Health*, 9, 8–23.
- Tremblay, R. E., Nagin, D. S., Seguin, J. R., et al. (2004). Physical aggression during early childhood: Trajectories and predictors. *Pediatrics*, 114, 43–50.
- U.S. Department of Health and Human Services (2005). *Head start impact study: First year findings*. Washington, DC: U.S. Department of Health and Human Services.
- Vermeer, H. J., & van IJzendoorn, M. H. (2006). Children's elevated cortisol levels at daycare: A review and meta-analysis. *Early Childhood Research Quarterly*, 3, 390–401.
- Webster-Stratton, C., Reid, M. J., & Hammond, M. (2001). Preventing conduct problems, promoting social competence: A parent and teacher training partnership in Head Start. *Journal of Clinical Child Psychology*, 30, 283–302.
- Yoshikawa, H., & Knitzer, J. (1997). *Lessons from the field: Head Start mental health strategies to meet changing needs*. New York, NY: National Center for Children in Poverty.
- Yoshikawa, H., Schindler, H., & Caronongan, P. (2009). *Prevention of mental health disorders, delinquency, and problem behaviors through intervention in infancy and early childhood*. Washington, DC: Institute of Medicine.
- Yoshikawa, H., & Zigler, E. (2000). Mental health in Head Start: New directions for the twenty-first century. *Early Education and Development*, 11, 247–264.
- Yoshikawa, H., Weiland, C., Brooks-Gunn, J., et al. (2013). *Investing in our future: The evidence base on preschool education*. Society for Research in Child Development.
- Zachrisson, H. D., Dearing, E., Lekhal, R., & Toppelberg, C. O. (2013). Little evidence that time in child care causes externalizing problems during early childhood in Norway. *Child Development*, 60, 1–19.
- Zigler, E., & Styfco, S. J. (1995). *Head Start and beyond: A national plan for extended childhood intervention*. New Haven, CT: Yale University Press.