Meta-Analysis of the Effects of Early Education Interventions on Cognitive and Social Development

GREGORY CAMILLI

University of Colorado

SADAKO VARGAS
SHARON RYAN
W. STEVEN BARNETT

Rutgers, The State University of New Jersey

Background/Context: *There is much current interest in the impact of early childhood education programs on preschoolers and, in particular, on the magnitude of cognitive and affective gains.*

Purpose/Objective/Research Question/Focus of Study: *Because this new segment of public education may require substantial resources, accurate descriptions are required of the potential benefits and costs of implementing specific preschool programs. To address this issue comprehensively, a meta-analysis was conducted for the purpose of synthesizing the outcomes of comparative studies in this area.*

Population/Participants/Subjects: *A total of 123 comparative studies of early childhood interventions were analyzed. Each study provided a number of contrasts, where a contrast is defined as the comparison of an intervention group of children with an alternative intervention or no intervention group.*

Intervention/Program/Practice: *A prevalent pedagogical approach in these studies was direct instruction, but inquiry-based pedagogical approaches also occurred in some interventions. No assumption was made that nominally similar interventions were equivalent.*

Research Design: *The meta-analytic database included both quasi-experimental and randomized studies. A coding strategy was developed to record information for computing study effects, study design, sample characteristics, and program characteristics.*

Teachers College Record Volume 112, Number 3, March 2010, pp. 579–620

0161-4681

Findings/Results: *Consistent with the accrued research base on the effects of preschool education, significant effects were found in this study for children who attend a preschool program prior to entering kindergarten. Although the largest effect sizes were observed for cognitive outcomes, a preschool education was also found to impact children's social skills and school progress. Specific aspects of the treatments that positively correlated with gains included teacher-directed instruction and small-group instruction, but provision of additional services tended to be associated with negative gains.*

Conclusions/Recommendations: *Given the current state of research on the efficacy of early childhood interventions, there is both good and bad news. The good news is that a host of original and synthetic studies have found positive effects for a range of outcomes, and this pattern is clearest for outcomes relating to cognitive development. Moreover, many promising variables for program design have been identified and linked to outcomes, though little more can be said of the link than that it is positive. The bad news is that there is much less empirical information in the studies examined available for designing interventions at multiple levels with multiple components.*

There is much current interest in the impact of early childhood education programs on preschoolers and, in particular, on the magnitude of cognitive and affective gains for children considered at risk of school failure in the early grades. Unlike other earlier policy efforts with similar aims (e.g., Head Start), many policy makers are not viewing publicly funded preschool solely as a compensatory program for children disadvantaged by poverty and other issues. Instead, they are increasingly supporting universal preschool programs as a means to capitalize on the learning that takes place in the early years for all children, based on research that demonstrates that participation in a universal preschool program improves children's academic achievement regardless of background or personal circumstances (Barnett, Brown, & Shore, 2004; Gormley, Phillips, & Gayer, 2008).

However, not all states offer preschool for all, choosing instead to target their programs toward the neediest segments of their student populations. In other states, the potential expansion of preschool programs has met some resistance. This is perhaps not surprising given that implementing a new segment of public education requires significant funding, and the research base on the implementation of preschool programs is not always clear. Researchers have focused on obtaining accurate descriptions of the potential benefits and costs of implementing specific preschool programs (e.g., Barnett, 1996), whereas other researchers have attempted to provide summaries of this research (e.g., White, 1985; White, Taylor, & Moss, 1992). Although a focus on the economics and impacts of providing preschool does help policy makers with questions as to why preschool should be funded and the issue of targeted versus universal programs, the more nuanced decisions about the kinds of services provided by the program (e.g., family supports, health), the curriculum offered, and the type of instruction to be employed remain less clear in this research base. To address this issue, a comprehensive data set was recently constructed for the meta-analysis of the outcomes of comparative studies to help determine the relations between various policy variables and program effects. The data set for the current study was prepared by Jacobs, Creps, and Boulay (2004).

Each study included in the database was designed using experimental principles: Preschoolers receiving a program of educational and other services were compared with a more or less similar group receiving either no intervention or an alternative intervention. Impacts were then estimated in a number of outcomes domains and to assess how program and population characteristics influenced these outcomes. Two previous meta-analyses (Gorey, 2001; Nelson, Westhues, & MacLeod, 2003) found positive and long-term effects for programmatic interventions on cognitive and social-emotional outcomes. In an analysis of 35 studies published in 1990 and 2000, Gorey reported relatively large effects sizes (*ES* = 0.7) for standardized tests of intelligence and academic achievement and that cognitive effects of relatively intense interventions remained large after 5–10 years (*ES* = 0.5 – 0.6). The relatively small number of studies included in this meta-analysis is potentially attributable to the requirement that studies either (a) be randomized, or (b) statistically control for preexisting differences.

For 34 studies reported between 1970 and 2000, Nelson et al. (2003) found a moderately large global impact of early childhood interventions in the preschool period (*ES* = .44), and these effects persisted through Grade 8 (*ES* = .08). The cognitive impact alone was somewhat larger over the K–8 period (*ES* = .22), and they found that cognitive impact was greatest for preschool programs with a direct teaching component. Positive effects on social-emotional impacts were also reported to endure through this period. Similarly to Gorey (2001), Nelson et al. found that more intensive treatments tended to have larger effects. The limited number of studies included in this meta-analysis is attributable to the selection criterion that a study have at least one follow-up assessment in elementary school or beyond. Selection criteria did not appear to include a provision for preexisting differences.

Three other recent studies in this area also deserve some consideration. The systematic review by Anderson et al. (2003) included 12 studies reporting cognitive outcomes for children aged 3–5. They obtained positive average effects for academic achievement (*ES* = .35), school readiness (*ES* = .38), and IQ (*ES* = .43). These average effect sizes were based on 10, 4, and 9 studies, respectively. Next, in the NICHD Early Childcare Research Network Study (2002), it was found that three core features of early child care—quantity, quality, and type—were related to children's school readiness and social behavior at age 4 1/2. In particular, higher quality care predicted higher entry-level academic skills and language performance. Yet this examination of about 1000 children represents high quality research within the framework of a single original study. Its significance is best understood within a series of longitudinal analyses of this sample examining multiple factors on child development. The third study in this series, by Karoly, Kilburn, and Cannon (2005), examined the costs and benefits of intervention programs for children from birth through 5 years old. They reviewed 20 early childhood programs with "well-implemented experimental designs or strong quasi-experimental designs" (p. xvii). For children aged 5–6, which they characterized as "near or in elementary school" (p. 66), they found an overall average effect size of $ES = .28$. However, the magnitude of the effect depended on the combination of program emphases (home visiting, parent education, or center-based education).

Though all the studies cited above reported positive effects, they are clearly based on different populations of students, different program approaches and philosophies, different time periods, and different coverage of studies. There is a need for a broader look at the efficacy of early intervention programs in a study that examines the research more broadly. The contribution of this new study is the larger collection of policy variables examined, including duration of treatment effect, types of programs and instructional practices, and the provision of services. In the following sections, the database is described and the analytic methods of the current study are provided, followed by results and policy implications. The topic of large systematic and prospective investigations is then considered.

DESCRIPTION OF THE DATA AND RESULTS

Meta-analysis is a method of statistically summarizing the results of quantitative measurements across many research studies. Cooper and Hedges (1994) described this method as consisting of five steps: problem formulation, data collection, data evaluation, analysis and interpretation, and public presentation. For the present study, the major focus is on the average impact of early education interventions in the cognitive outcome domain, which includes measures of intelligence and reading. Research in affective and school domains was examined, but fewer outcomes were reported. In addition to overall impact, the data are explored for specific characteristics of studies and programs that are associated with variations in outcomes.

A brief description of the selection criteria for studies follows. To be included in the meta-analysis database, quantitative studies of early childhood had to meet the following conditions: (1) The early childhood programs must be center based; educational interventions delivered only through home visits or in family childcare settings were excluded. (2) The early childhood programs/interventions must provide educational services to children directly; early childhood services provided solely via parent education were excluded, although parent involvement may be part of the program. (3) Interventions and programs must target children's cognitive and/or language development; programs may additionally target other aspects of children's development, but if cognitive/language outcomes are not central to the program, they were excluded. (4) Interventions and programs must provide services for at least 10 hours per week for 2 months; programs/interventions that are of lesser intensity or duration were excluded. (5) Programs that specifically target only special needs children were excluded. (6) Programs must have been implemented in the United States and reported no earlier than 1960. (7) Studies must have designs that include comparison groups in the form of a control (no treatment or intervention) or an alternative treatment.

These criteria were chosen, as implied, to focus on program-based intervention studies that were comparative in nature and substantial with respect to treatment intensity (time and duration). A variety of strategies and tools were used to screen relevant sources of studies, including research journals, books, technical reports, printed and computerized databases (e.g., ERIC), dissertations and theses, conference presentations and proceedings, foundation grants, and the like. The comprehensive list of sources used is given in Table 1.

A total of 161 studies were identified. Each study provided a number of contrasts, where a *contrast* is defined as the comparison of an intervention group of children with an alternative intervention or no intervention group. Individual studies typically contained more than one contrast because a study may examine more than one type of intervention, and most contrasts reported outcome measures in more than one outcome domain. Thus, three types of contrasts are possible: A treatment group was compared (1) with a group without uniform services, (2) with a group with services but not a uniform programmatic intervention, or (3) with a group receiving an alternative treatment. See Table 2 for a definition of key terms. Additional information was recorded that described other aspects of the treatment, such as the time of measurement (end of

Table 1. Strategies for Searching the Literature

Computer and/or Search of Electronic Databases
ERIC (Educational Resources Information Center database; includes Resources in Education and Current Index to Journals in Education)
Federal Research in Progress (FEDRIP)
PsycINFO Psychological Abstracts
Social SciSearch (Social Sciences Citation Index)
Dissertation Abstracts Online (Dissertation Abstracts International, Masters Abstracts)
Foundation Grants Index
Education Abstracts
Applied Social Sciences Index and Abstracts
National Technical Information Service (NTIS)
Child Development Abstracts and Bibliography
Social and Behavioral Science Documents (SBSD)
Social Sciences Literature Information System (SOLIS)
Review of Educational Research
Psychological Abstracts
Manual search of proceedings from relevant research conferences (e.g., AERA, SRCD, Head Start, NAEYC)
Footnote Chasing
References in journals from nonreview articles
References from nonreview articles not published in journals
References in review articles
References in books/book chapters
References listed on program/program model Web sites
Topical bibliographies compiled by others
Consultation
Communications with colleagues and clients
Attending meetings and conferences
Formal requests of scholars who are active in the field
Formal requests of foundations that fund research in the field
General requests to government agencies
Reviewing electronic networks

treatment, short-term, and long-term) and the type of instrument used to measure an outcome. Though the main outcomes of interest were cognitive, other types of outcome measures were also recorded, including children's social-emotional development and school progress. A select list of sources with references to all the included studies is available in the appendix. The full bibliography, which consists of 610 references, is available upon request.

For every dependent variable within a contrast, an effect size was calculated for each individual child outcome reported at each point in time. Originally, the database consisted of 8,168 effect sizes (or about 18 effect sizes per contrast), of which 2,409 came from low-quality or nonrelevant outcome measures. The latter were eliminated from further analyses. This resulted in a set of 123 studies, of which 76 contained only treatment-control (T/C) contrasts, 17 contained only treatment/alternative treatment (T/A) contrasts, and 30 contained both T/C and T/A contrasts. From these 123 studies, 2,243 effect sizes were calculated for T/C contrasts, and another 1,708 effect sizes were computed from T/A contrasts. After averaging across effect sizes within outcome categories with studies, 263 T/C contrasts and 149 T/A contrasts were obtained. Two contrasts lacked T/C or T/A designation. The T/C and T/A contrasts were analyzed separately.

DATA COLLECTION

A coding strategy was developed to record information for computing study effects, study design, sample characteristics, and program characteristics. The coding frame development and data collection activities were also carried out by Jacobs et al. (2004) in collaboration with the National Institute for Early Education Research at Rutgers University. Studies were coded using a formal protocol, and after training sessions, coders achieved an interrater reliability of .80 with a master coder. The first 10 studies were double-coded by independent analysts, as were 20% of the remaining studies. Differences in coding were resolved by a master coder. Data entry was also verified for approximately 10% of the studies. The coding protocol was a modified version of that used for the meta-analysis component of the National Evaluation of Family Support Programs (Layzer, Goodson, Bernstein, & Price, 2001). This protocol was customized for recording information on early childhood education interventions and expanded to include target variables that were found in previous meta-analyses in order to correlate with treatment outcomes in early childhood education.

Five types of information were collected at the contrast level: study design (e.g., group assignment); recruitment of participants (e.g., age groups targeted); description of the intervention (e.g., details of services children received, curriculum, and pedagogical approach); location of the intervention (e.g., urbanicity of the study sites); and various ratings of the quality of the implementation of the intervention. In particular, an overall quality rating of the study was constructed in which *high quality* was defined based on satisfying the following conditions: Treatment and comparisons groups were deemed equivalent at baseline; attrition bias was tested or corrected in some fashion; there was no evidence suggesting poor implementation of the intervention; and the information coded from a study was not considered inadequate.

Table 2. Key Meta-Analysis Report Terms

Program:	A center-based early childhood intervention for children aged 3–5 years that focuses on children's cognitive language development. Examples: Head Start, Abecedarian curriculum.
Study:	A research or evaluation project in which children enrolled in an early childhood program are compared with one or more other groups of children. Examples: the Abecedarian Study, the Perry Preschool Project.
Report:	A written description of a research study or studies and outcomes; the description may be published or unpublished. Example: journal article from *Child Development* or U.S. Department of Education report.
Contrast:	A comparison of one group of children receiving a particular treatment or intervention with another group of children. In this report, only contrasts between treatment and control groups were included. The contrast is the main unit of analysis for this report.
Individual Effect Size:	This is the difference between a treatment and control group at a given point of time on a given measure, expressed in standard deviation units.
Average Effect Size:	Any given contrast may have data on multiple individual effect sizes for each outcome domain and time point. To conduct analyses at the contrast level, an average effect size was created that is the mean of individual effect sizes for each outcome domain and time point for each contrast.

EFFECT SIZE COMPUTATION

Comparative outcomes were reported by a number of different statistics. These were converted to the effect size scale using a commercially available software package (Shadish, Robinson, & Lu, 1999). Effect size was computed using the standard formula

$$ES = c(m)\frac{\overline{X}_t - \overline{X}_c}{s_p}, \tag{1}$$

where $\overline{X}_t$ is the mean of the treatment group of size n_t on some measure, $\overline{X}_t$ is the mean of the comparison group of size n_c on that same measure, and s_p is an estimate of the pooled standard deviation. The correction factor is based on $m = n_t + n_c - 2$ degrees of freedom. The Hedges adjustment $c(m)$ results in effect sizes being shrunk toward zero, where the degree of shrinkage is inversely proportional to the number of children in a contrast.

The large-sample approximate variance of the effect size estimator is given by

$$var(ES) = \frac{n_t + n_c}{n_t \cdot n_c} + \frac{ES^2}{2(n_t + n_c)}. \qquad (2)$$

Histograms of raw effect sizes are given in Figures 1 and 2 for the T/C and T/A contrasts across dependent variable categories.[1]

Figure 1. Unaggregated effect sizes for contrasts of treatment and control

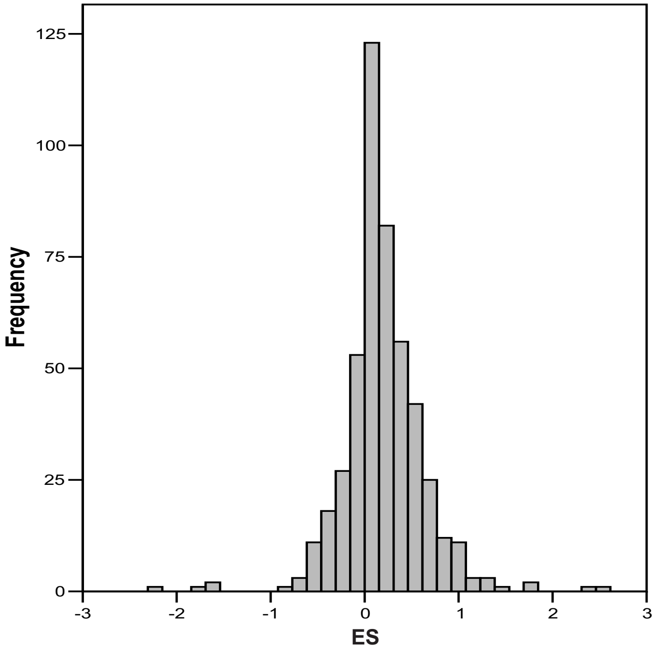
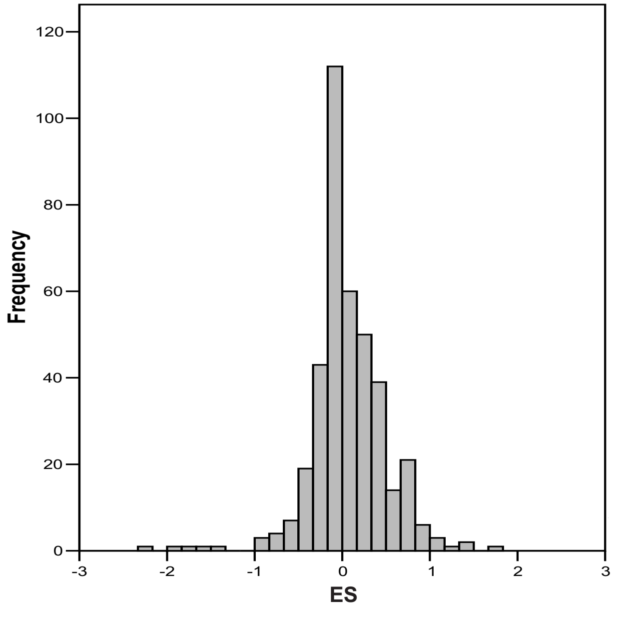


Figure 2. Unaggregated effect sizes for contrasts of treatment and alternative treatment



Not all conversions to the effect size scale provide equivalent results. Because of inadequate reporting in the original studies, some conversions will provide less adequate estimates of effect size. Whenever possible, the best estimates ("preferred") were obtained, but if this was not possible, rougher approximations ("acceptable") were recorded. Excluding "acceptable" effect sizes would exclude a great deal of valuable information. For example, suppose a study reported that certain outcomes were tested, specifically noting that no significant group differences were found, but reported no statistical information. To include this information, is it conventional to enter effect sizes of zero for reported measures into the database.

AVERAGING EFFECT SIZES

Many studies included outcome measures in more than one outcome domain. However, a study might also include several outcomes within an outcome domain. These effect sizes are clearly not statistically independent, as required by standard statistical models, and often there was not a "best" choice within an outcome domain. This issue was resolved by averaging preferred and acceptable effect sizes for a given design contrast within an outcome domain. This was done separately at posttest, follow-up, and long-term follow-up. Thus, the maximum possible number of aggregate effect sizes would be 1,590 (106 studies * 5 domains * 3 follow-ups), but not all studies reported in all five domains at all follow-ups. The effective number of effect sizes was 869, of which 479 were for treatment-control comparisons.

ANALYSIS

Three characteristic features of the data set that require some discussion are the nature of the comparisons made in the T/C and T/A contrasts, imputation of missing values, and reorganization the outcome domains into three categories—cognitive, school, and social/emotional. These topics are given separate treatment. However, the goals of the analysis are to identify features of programs that predict effectiveness and to understand their relative importance and combined influence. To understand relationships among the moderator impacts on the outcomes and estimate the combined effects, a multivariate approach is therefore necessary. Based on this information, the goal is to provide a deeper understanding of what types of programs and services are most effective.

DESCRIPTIVE ANALYSIS OF CONTRASTS

In this section, descriptive information is given for the two types of contrast: T/C and T/A. A total of 273 of 412 contrasts were obtained from studies conducted prior to 1970, and 38 were from studies conducted after 1990. A total of 71 contrasts were derived from randomized studies. Earlier studies were more likely to use random assignment to treatment conditions and were thus considered to have higher quality than later studies.

The first four data columns of Table 3 describe the instructional characteristics of the T/A contrasts for the treatment (T) and alternative treatment (A) groups. In both groups, the majority of the programs implemented formal curricula, followed by use of content- or performance- based standards. The largest percent of the programs used small-group approaches in both groups, whereas whole-group instruction was rarely provided. The primary pedagogical approach was direct instruction (DI) in 50% of treatment groups (T) and in 38% of alternative treatment groups (A). Direct instruction "refers to instruction involving mostly teacher-directed activities designed to teach information and develop skills" (Jacobs et al., 2004, p. 3–12). This approach is often contrasted with inquiry-based pedagogical approaches in which children take the lead in their learning by engaging with the environment and exploring ideas through hands-on experimentation.

Although many interventions focused on multiple aspects of development, the most common programmatic focus in the T/A contrasts was the development of general cognitive skills. Of the 149 T/A contrasts, 65% of the interventions for the experimental group (T) targeted general cognitive objectives, 66% targeted the reading/language objectives, and 51% targeted the social-emotional objectives. For the alternative treatment group (A), 68% of the interventions focused on general cognitive objectives, 52% focused on reading/language objectives, and 44% focused on social-emotional objectives. Note that outcome domains assessed in the studies are not synonymous, or entirely consistent, with the primary educational objectives.

Table 3. Instructional Characteristics of Programs Represented in the Treatment–Treatment Contrasts in the Extended Analysis.

Characteristics	T/A				T/C	
	T		A		T*	
	n	%	*n*	%	*n*	%
Curriculum						
Formal curriculum	118	79.2	90	60.4	194	74.3
Comprehensive curriculum	60	40.3	48	32.2	136	52.1
Standards-based curriculum	93	62.4	66	44.3	131	50.2
Primary Instructional Grouping						
Whole-group instruction	6	4.0	9	6.0	5	1.9
Small-group instruction	51	34.2	26	17.4	32	12.3
Individual instruction	6	4.0	24	16.1	13	5.0
Mixed	27	18.1	24	16.1	86	32.95
Primary Pedagogical Approach						
Direct instruction	50	33.6	38	26.8	20	7.7
Inquiry based	27	18.1	28	19.7	56	21.5
Mixed	16	10.7	19	13.4	59	22.6
Focus of the Intervention						
Reading/language	99	66.4	77	51.7	134	51.3
Mathematics	59	39.6	36	24.2	73	28
General cognitive	97	65.1	97	68.3	230	88.1
Social/emotional	76	51.0	62	43.7	147	56.3
Motor/physical health	25	16.8	34	23.9	79	30.3
Total Contrasts	149		149		263	

Note. Data are from 412 contrasts, and categories are not mutually exclusive.

*Control groups not coded for T/C contrasts.

MISSING VALUE IMPUTATION

Many variables had missing values, with the rate ranging from 1% to 58% per variable. The problems with this extent of missing data are (1) loss of efficiency, (2) complication in data handling and analysis, and (3) bias due to unknown systemic trends in the unobserved data. It is well known that mean substitution or pairwise deletion does not account for the variation that would be present if the variables are observed, resulting in downward bias in the estimation of variances and standard errors (Switzer, Roth, & Switzer, 1998). As noted by Pigott (2001), mean substitution and pairwise deletion methods "provide problematic estimates in almost all instances" (p. 380). In the present analysis, multiple imputation methods were used for handling incomplete data problems; SAS 9.1 (SAS Institute, 2008) multiple imputation procedures were used.

Multiple imputation is a procedure for analyzing data with missing values. According to theoretical and simulation studies, multiple imputation-based approaches produce less biased estimates of coefficients and standard errors, and more similar patterns to the original estimates in simulation studies than other methods (Noh, Kwak, & Han, 2004). The process generates several plausible values for each missing observation to obtain m (typically taken as 5–10) sets of imputed data. Each of the complete data sets is then analyzed by using standard procedures and the results combined for inference by incorporating appropriate variability within and across multiple imputations (Yuan, 2004). There are various methods for imputation depending on the type of missing data pattern. For this study, the data were assumed to be "missing at random," and consequently, this approach does not resolve biases arising for systematically missing values. However, the method would be expected to provide better estimates than mean substitution. The expectation-maximization (EM) algorithm was used for imputation of the missing values. This algorithm provides maximum likelihood estimates of the variance covariance matrix of the distribution of variables, which is in turn used to generate plausible values (Bilmes, 1998).

Because implemented multivariable normality is assumed, this method has been widely used for imputing missing values for binary variables. Because binary variables are treated like normal variables in the imputation steps, imputed values may be fractional. One strategy for imputing the binary data is to round up or down the imputed fraction to 1 or 0. However, it has recently been shown that such rounding can produce substantial bias (Ake, 2005; Allison 2005; Horton, Lipsitz, & Parzen, 2003), and it is generally recommended that the unrounded imputed values be used for analysis. In this analysis, therefore, imputed values for dichotomous variables were not rounded. Though individual values may be unrealistic, the goal of the imputation is to produce accurate estimates of the variance-covariance matrix.[2]

OUTCOME DOMAINS

In the current analysis, the five domains used for data collection were collapsed into three main domains based on the contextual similarities and considering the mean effect size differences between the original and combined domains. The intelligence and cognitive/reading achievement domains were combined into a single "cognitive" domain. The mean effect sizes of the two original domains were not significantly different, nor were the mean effect sizes different when controlling for other independent variables.[3] Combined effect sizes were available for 97 studies with 556 aggregate effect sizes. Instrument types included IQ measures, cognitive achievement tests (such as reading, writing, spelling, and verbal development), mathematics, and tests of school readiness.

The school progress domain was not collapsed. Data were available from 32 studies with 97 effect sizes. Instrument types included school grades, academic track, special education placement, high school completion, and college attendance. Social/emotional and antisocial outcome domains were combined into a single social-emotional domain. The mean effect sizes of these two domains were not significantly different. Data were available from 43 studies with 216 effect sizes. Instrument types included self esteem, school adjustment, educational aspirations, and aggressive or antisocial behaviors.

STATISTICAL ANALYSIS

The final database included the T/C and the T/A contrasts. These two subsets were analyzed separately. As explained previously, the T/C contrasts compared the performance of the children who received early intervention with the performance of the children who received no intervention or an unsystematic intervention. A T/A contrast compared an intervention with an alternative treatment. Whereas the T/C contrasts provide information to isolate the absolute effect of intervention, the T/A contrasts provide information to explore the relative impact of different intervention and implementation characteristics.

The T/C and T/A interventions tended to have different intervention focuses. Whereas 88.1% of the T/C contrasts involved interventions with a general cognitive focus, the corresponding figure was 65.1% for the T/A contrasts. The T/A contrasts were classified more often as having a reading/language focus (66% vs. 51%) or a mathematics focus (39.6%

vs. 28%). The T/C contrasts were classified more often as having a motor/physical focus (25% vs. 16.8%) or a social/emotional focus (56.3% vs. 51%). For these reasons, it is possible that the T/C contrasts may be more sensitive to cognitive-oriented interventions, whereas T/A contrasts may be more sensitive to content-oriented interventions.

STATISTICAL ANALYSIS

The approach to data analysis is facilitated by briefly considering a multilevel linear model. With this approach, two sources of random variation are distinguished: prediction errors within studies, and effects that vary randomly between studies. The model, accordingly, can be written:

$$ES_{ij} = \beta_{0i} + \beta_{1i}x_{1ij} + \ldots + \beta_{ki}x_{kij} + \varepsilon_{ij}, \tag{3}$$

where

$$\beta_{0i} = \beta_0 + \tau_i. \tag{4}$$

In this specification, an outcome variable of interest is denoted by ES_{ij} where the subscript i signifies study, and j signifies effect size within study. In the model given by Equations 3 and 4, the random components are assumed to be uncorrelated with each other and are typically defined as

$$\varepsilon_{ij} \sim \text{NID}(0, \sigma_\varepsilon^2) \text{ and} \tag{5}$$

$$\tau_j \sim \text{NID}(0, \sigma_\tau^2). \tag{6}$$

The use of ordinary least squares regression results would ignore the multilevel structure and may result in biased estimates of the fixed coefficients and biased inferential tests (Goldstein, 2003). Studies are typically weighted by the inverse of the (estimated) sampling variance of the effect size as given in Equation 2 (Raudenbush, 1994; Raudenbush & Bryk, 2002). However, in the current study, unweighted results are reported because of large sample-size discrepancies between different kinds of interventions. Although this can be expected to provide less optimal results, it seems safer at present than assuming that program effects are uncorrelated with study size.[4]

In the current study, the multilevel model as given in Equations 3–6 was used for this analysis, with imputation to compensate for missing data. This strategy was used to investigate the relationship between the program characteristics and effect size. Mixed modeling and imputation procedures were performed as implemented in SAS Version 9.1 (SAS Institute, 2008) using the procedures MI and MIANALYZE. An example of this code is given in Figure 3.

Figure 3. SAS code for mixed modeling with imputation

```
proc mi data=new simple nimpute=10 out=nieerimp seed=384756;
em MAXITER=500;
var {imputation variables};
mcmc initial=em (maxiter=300);
run;

proc mixed data=nieerimp;
class studyid;
model es = {model variables} / solution covb ddfm=bw;
random int / sub=studyid;
weight invar;
by _Imputation_;
ods output SolutionF=mixparms CovB=mixcovb;
run;

proc mianalyze parms=mixparms edf= {degrees of freedom}
CovB(effectvar=rowcol)=mixcovb;
modeleffects intercept {model variables};

run;
```

“Study” was used as the random variable indicator (“subject” in SAS), and effect sizes at different points in time were combined into a single analysis, with time as a quantitative design variable (coded 1, 2, and 3). Recall that *study* was defined as a research or evaluation project in which children enrolled in an early childhood intervention. A number of contrasts, which employ different types of interventions, are available for each study, but the number varies by outcome domain.

All moderators, for both design and program characteristics, were examined equivalently within the framework of the analytic model, including design quality. The relative effects of a number of moderator variables on the size of outcome were evaluated. The full set of moderator variables explored is given in Table 4.

Table 4. Moderator Variables Examined for Effect on Outcomes

Variable	Explanation
Time of testing	
1	End of treatment, children 3–5 years of age
2	Short-term follow-up, children 5–10 years of age
3	Long-term follow-up, children older than 10 years of age
Amount of treatment	
Days	Number of treatment days, mean = 280.56, range = 33–3120
Dose	Treatment hours per day, mean = 3.72, range = 1.04–9.75
Hours	Total treatment hours, mean = 1205, range = 99–21840 hours
Curriculum (coded as 1 = yes and 0 = no)	
Formal	Formal curriculum in preschool
Standards	Content or performance standards-based curriculum
Comprehensive	Comprehensive curriculum
Instructional group size (coded as 1 = yes and 0 = no)	
Individual	Individual instruction
Whole	Whole-class instruction
Small	Small-group instruction
Mixed	Mixed
Differentiated	Differentiated instruction
Individualized instruction (coded as 1 = yes and 0 = no)	
	Program had formal curriculum, class size < 10; the child/staff ratio < 5, or program used primarily small group or individual instruction.
Pedagogical approaches (coded as 1 = yes and 0 = no)	
Direct	Direct instruction: mostly teacher-directed activities designed to teach information and develop skills
Inquiry	Mostly hands-on instruction, student-directed learning w/ teacher as facilitator
Mixed	Mixed approaches: use of both direct and inquiry-based instruction
Instructional focus (coded as 1 = yes and 0 = no)	
Reading	Focused on reading
Math	Focused on math skill development
Cognitive	Focused on cognitive skill development
Noncog	Focused on emotional, behavioral, and physical health and motor development
Population served (coded as 1 = yes and 0 = no)	
Preschool	Preschool age only
Younger	Preschool and younger children
Older	Preschool and older children
Both	Preschool, younger, and older children
Additional service received (coded as 1 = yes and 0 = no)	
Services	Program provided services in addition to ECE
Program Target (coded as 1 = yes and 0 = no)	
Low$	Program targeted low-income families
Population characteristics	
PerLow$	Percentage of low-income families in the group
Age	Average age of children at the initiation of the intervention
Design characteristics (coded as 1 = yes and 0 = no)	
Design Quality	Design quality coded yes (1) if there was no attrition, attrition bias was tested or attrition was remedied, or there was baseline equivalence of the two groups; there was no evidence of lack of fidelity in program implementation; and the coder did not believe that the information was unfair or inadequate.
Equivalence	Yes (1) if pretest *ES* < .2 or for randomized or matched groups

Not all studies provided information on all moderator variables. In addition, moderator variables were not coded for comparison groups consisting of unsystematic programs or interventions.

RESULTS

The number of studies and effect sizes are given in Table 5, along with the unweighted average effect size for each type of contrast (T/C and T/A) under each outcome domain. Note that *contrast* is the unit of analysis for the three domains shown in Table 5.

Table 5. Overall Unweighted Effect Sizes (ES) by Domain

Domain	Cognitive		School		Social	
	T/C	T/A	T/C	T/A	T/C	T/A
# Studies	81	39	29	8	37	17
# *ES*	306	250	60	37	113	103
Mean *ES*	.231*	.067	.137*	.058	.156*	-.031

* $p < .01$.

Pooled across time of testing (end of treatment, short-term, and long-term), the unweighted mean effect sizes of the T/C comparisons were *ES* = .231(cognitive domain), *ES* = .137 (school domain), and *ES* =.156 (social domain). Average effects for each of the domains for T/C contrasts were significantly different from zero. For the T/A contrasts, none of the effects was significant. The T/C contrasts yielded larger effect sizes than the T/A contrasts, as would be expected, because both groups in the T/A studies attended programs. In both types of contrast, the largest effect was obtained in the cognitive domain.

In Table 6, estimated moderator effects are given in the three domains for the T/C and T/A contrasts. With respect to this table, the concept of *design consistency* is introduced and will be used in the remainder of this section.

An effect size is based on the average outcome in the treatment group minus the average outcome in the control group; this is the difference $\bar{X}_T - \bar{X}_C$ in the numerator of Equation 1. A beneficial intervention in the treatment group makes the difference *larger*, but a beneficial intervention in the alternative treatment group makes the difference *smaller* (because $\bar{X}_C$ is larger). Therefore, a positive regression coefficient for a moderator variable in the treatment is consistent with a negative coefficient for that moderator in the alternative treatment (moderators were coded separately for the two groups in question). In other words, if a moderator has a consistent effect, then the regression coefficients would have opposite signs for the two groups.

The requirement for a design-consistent effect included the condition of "opposite signs" for the regression coefficients and the condition that at least one of these coefficients was statistically significant. We did not explore a model that took this constraint into account when obtaining null probabilities; however, it seems plausible that the standard two-tail probability is overestimated in the case of design consistency, and future research may provide a clearer understanding of this issue. In any event, primary interpretations from Table 6 are based on this definition of design consistency. Note that moderator effects were not as apparent in the T/C contrasts as in the T/A contrasts.

Table 6. Moderator Effects for the Cognitive, School, and Social Domains

Moderators	Group	Cognitive Domain T/C	Cognitive Domain T/A	School Domain T/C	School Domain T/A	Social Domain T/C	Social Domain T/A
Intercept	—	.230[b]	.187[a]	.153[a]	.160	.335[b]	.056
Time	—	-.241[b]		-.080[a]			
Direct	T		.211[b]				
	Comp[c]		-.292[b]				
Individualized Instruction[d]	T		.161[a]				
Services	T	-.193[b]	-.471[b]		-.215[a]		
	Comp[c]		.233[b]			-.232[b]	-.289[b]
Design Quality		—		.275[b]			.266[b]

Note. Comparisons are defined as T/C = treatment/control, T/A = treatment/alternative treatment.
[a] $p < .05$. [b] $p < .01$. [c] Comparison group. [d] Variable coded for first treatment group only

Overall raw program effects are given by the intercept in Table 6. These intercept values are interpreted as the end-of-treatment average outcomes with the values of all moderator variables at zero (i.e., no directive or individualized instruction, no services, and low design quality). The moderator variable coefficients add or subtract from these baseline effects, and we give a number of profiles of total effect below after giving the results for individual moderator variables.

EFFECT OF TIME

The intervention effect in the cognitive domain decreased over time. The decrease is greater in the T/C than the T/A contrasts. In the T/C contrasts, the effect decreased by about 0.24 *ES* units per follow-up period. For the T/A contrasts, the decrease was small and not significant. This is an estimate of the *relative* loss over time given two treatments, and consequently, there is less of a gap to close in T/A than T/C contrasts. Significant change over time was not observed for the school or social domains.

EFFECT OF INSTRUCTIONAL CHARACTERISTICS

As mentioned previously, DI involves teachers explicitly instructing children in academic skills and procedures, usually through activities designed and led by the teacher. The DI approach contrasts with inquiry-based educational activities that involve mostly hands-on, student-directed learning. The effect of DI was design consistent in the cognitive domain for T/A contrasts, but the effect was not observed in the T/C contrasts. This result is in accord with that of Nelson (2005), who concluded that the "cognitive impacts during the preschool time period were greatest for those programs that had a direct teaching component in preschool" (p. 1). In addition, the effect of individualized instruction was positive in the T/A contrasts.

EFFECT OF ADDITIONAL SERVICES

As noted, some programs provided additional services to children and their families, such as health screening, nutrition, educational/teaching materials for home use, and home visits. Provision of additional services showed a strong and negative effect on the cognitive domain, and this effect was design consistent. A negative effect was also observed in the school domain (T/A), but it was not design consistent; a positive effect was observed in the social domain (T/C and T/A), but this effect was not design consistent.

In the cognitive domain, this result may signify an indirect effect of other variables with which services are confounded. For instance, receiving additional services correlated positively with the total number of days of intervention and negatively with the number of hours of instruction per day. For unknown reasons, those who received additional services had lower dose levels and longer durations of treatment. It is possible that programs added time spent on additional services to instructional time in arriving at total contact hours. Also, the additional service variable correlates negatively with DI, which has a positive impact on the cognitive outcomes. The children who received additional services tended to receive less direct instruction, and instruction in larger classes.

OTHER FINDINGS

Some previous studies (MacLeod & Nelson, 2000; Nelson, 2005) have reported a positive impact of longer duration of intervention on cognitive and social outcomes. In the current study, treatment duration did not have a significant effect. Likewise, no effects were observed for income or education. There may be two precipitating causes for this result. First, there was much missing information for these variables, and second, there was little sample variability: Almost all families could be described as having both low income and low education. Studies with high design quality yielded larger effect sizes (about .27 *ES*) in the cognitive domain (T/C) and social domain (T/A). High design quality was operationalized as: (1) there was no study attrition, (2) attrition bias was tested and remedied, or (3) the baseline equivalence of the two groups was established. High design quality was precluded if coders observed evidence of lack of fidelity in the implementation of the program, or indications that reported study information was unfair or inadequate. Finally, we note that other variables in Table 4 were examined and found not to have a significant effect on outcome.

EXPECTED OUTCOMES

Table 7 gives "scenarios" in which the total effects are calculated for different moderator combinations for the cognitive domain. This total effect is the projected average outcome for a set of interventions having as average moderator values those given in the table (rather than projected values for individual interventions). The following strategy was used for determining regression coefficients. First, coefficients for the T/C comparisons were used for the intercept, time, services, and design. (For the time effect, the T/C coefficient was used because this would describe the decrement over time in the absence of a systematic treatment.) For individualized instruction, the coefficient from the T/A analysis was used, and for direct instruction, the treatment and alternative group effects from the T/A analysis were averaged (after reversing signs for the alternative group coefficient). This representation of the results admittedly extrapolates from the observed data. However, it is important to project the kinds of effect sizes that might be encountered through intentional design rather than assuming the same constellation of moderator variables encountered in past early childhood programs.

Furthermore, the extrapolation is empirically-based, and its assumptions are explicit and public. What has happened in past programs is described in data rows 1, 4, and 7 of Table 7, where effect sizes are evaluated at the sample means of the program (moderator) variables. Interpretation of these effect sizes as predictions of outcomes for newly designed early education programs runs the risk of underestimating potential effects because such programs can be designed with other program characteristics (i.e., other values for the moderators).

Table 7. Outcome Profiles for the Cognitive Outcome Domain

	Variable (intercept = .230)						
Scenario	Time [a] -.241	DI 0.252	Individualized .161	Services -.193	Design .275	Estimated *ES* Linear	Estimated *ES* Nonlinear
1(mean)[b]	0	0.04	0.26	0.75	0.36	0.48	0.51
2	0	0.50	0.00	0.50	1.00	0.79	0.81
3	0	0.50	0.50	0.50	1.00	0.87	0.89
4 (mean)[b]	1	0.02	0.20	0.81	0.37	0.21	0.18
5	1	0.50	0.00	0.50	1.00	0.55	0.50
6	1	0.50	0.50	0.50	1.00	0.63	0.58
7 (mean)[b]	2	0.10	0.54	0.89	0.70	0.12	0.20
8	2	0.50	0.00	0.50	1.00	0.31	0.37
9	2	0.50	0.50	0.50	1.00	0.39	0.45

Note. The coefficient for DI is taken from the T/C contrasts in Table 6. The effects of small-group and individual instruction are averaged from the T/A contrasts.

[a] End of treatment = 0, short-term follow-up = 1, long-term follow-up = 2.
[b] Sample means for time period.

As seen in Table 7, assuming linearity of the time effect, the direct instruction and services variables have the most practically significant effect on estimated outcome. Evaluated at the sample means of the moderator variables, cognitive outcomes ranged from *ES* = .48 to .12. With a smaller amount of additional services and larger amounts of the other moderators, the impact of intervention on cognitive outcomes ranges from *ES* = .87 to .31 from the end of treatment to long-term follow-up (10+ years). Note that averaging across the follow-up times, (.48+.21+.12)/3 = .27, provides a rough approximation to the overall sample average for T/C contrasts (.23) given in Table 5 for the cognitive outcome domain. For the cognitive domain, the actual sample means for the T/C contrasts were .45, .16, and .23, respectively.

We also fit a model for cognitive outcomes in which the time variable was not constrained to be linear. For this model, global estimates (intercept + time effect) of effect size at Times 1–3 were *ES* = .50, .19, and .07, respectively. With these estimates, the decrease in effect size from Time 1 to 2 was about twice the drop from Time 2 to 3. Projections created with this assumption appear in the last column of Table 7. Under the linear assumption, the effect sizes in scenarios 2, 5, 8 have the pattern .79, .55, .31 over time, whereas the corresponding nonlinear pattern is .81, .50, .37. For scenarios 3, 6, 9, the linear pattern is .87, .63, .39, whereas the nonlinear pattern is .89, .58, .45. Thus, the time-dampening effect is somewhat lessened. However, under either the linear or the nonlinear assumption, the projected effect at long-term follow-up is educationally significant.

As noted, the services variable may combine a number of influences. It correlated positively with total days of intervention and instructional group size but negatively with hours/day of instruction and the provision of DI. It seems plausible that additional services provided did not have a direct impact on cognitive outcomes, but rather indicate that programs that focus more resources in the cognitive domain tended to be more successful for this type of outcome. The additional services, although they can be grossly categorized as aimed at children, parents, or both, were not defined in detail in the original publications. The two different types of instruction are combined in the third, sixth, and ninth rows of the table. There is no reason to believe that directive and individualized instruction cannot be combined in enhancing outcomes. However, individualization was derived (as noted above) rather than observed, and instructional design would benefit from an additional bridging study prior to policy recommendations.

It is indeed tempting to think of the results from the regression equations as causal effects, but we caution against this. These results are correlational for a number of reasons: Most of the original studies were not randomized, substantial interpretation was involved in coding studies because of chronic underreporting, and the studies comprise a historical record extending nearly 40 years into the past. In addition, the present understanding of treatments delivered almost two generations ago has most likely changed, and the meanings of some variables are no longer as crisp as they may have been in the original study contexts. It is perhaps best to think of the current meta-analyses as an effective summary of a complex research literature. Yet this set of comparative studies doubtlessly provides the most comprehensive evidence available for guiding rational policies in early education and should play an important, if not central, role. As Cronbach (1982) put it, “Many statements, only a few of them explicit, link formal reasoning to the real word; their plausibility determines the force of the argument” (p. 413).

DISCUSSION

The findings of this study span a wider range than previous meta-analyses, but they are consistent with those of earlier studies (e.g., Barnett, 1998; Gormley & Gayer, 2003; Nelson et al., 2003; Vandell & Wolfe, 2000). As Durlak (2003) observed, "consistent evidence obtained by different researchers surveying slightly different but overlapping outcome literatures confirms that preschool programs have statistically significant and practical long-term preventive impact." The current review, which covers 120 studies of cognitive outcomes carried out over 5 decades, provides even greater weight for the argument that preschool intervention programs provide a real and enduring benefit to children. It examines, in a greater level of detail than previous studies, the effects of program moderators. As Durlak found, the research is less clear regarding the specific program features that lead to optimal results.

The analysis reported here shows significant effect sizes in the cognitive domain for children who attend a preschool program prior to entering kindergarten. The intercept was moderate, and this can be interpreted as the predicted immediate posttreatment outcome in quasi-experiments for programs without a programmatic emphasis on DI, individualized instruction, and additional services. That is, the intercept is the predicted level of outcome with the values of the program moderators set to zero—or what could be described as the "bare bones" outcome. Although the largest effect sizes were observed for cognitive outcomes, positive results were also found for children's social skills and school progress. For the latter two categories, the intercepts for the T/C comparisons were statistically significant at conventional levels. Moreover, although the services indicator had a negative impact in the cognitive domain, it had a positive effect in both the school and social domains. Beyond this, policy makers must make more nuanced decisions regarding the intended population, duration and intensity of programming, type of instruction and curriculum, and structural characteristics such as class size and teacher qualifications. This expanded meta-analysis provides insights into two of these important implementation issues: instruction and range of services.

INSTRUCTION

In agreement with previous work (Nelson et al., 2003), it was found that the direct instruction component in preschool programs had an immediate effect on children's cognitive development in the T/A contrasts. Many early childhood educators might be concerned by this finding in light of the field's consensus that a developmentally appropriate approach (Bredekamp, 1987; Bredekamp & Copple, 1997) is not one in which children are drilled in basic concepts and have little opportunity to apply their knowledge in meaningful learning situations. However, the majority of these effect sizes were from studies conducted prior to 1983, when policy attention was directed toward determining the most effective curriculum model to ameliorate the effects of disadvantage (Goffin & Wilson, 2001). At that time, sufficient experimentation in curriculum design took place so that it was possible to find curricula that ranged along a continuum—from those that sought to convey basic skills through the presentation of concepts in small steps to large groups of children, to child-centered curricula associated with the traditional nursery school that used play as the core of instruction. Falling somewhere between these two extremes were curricula that used constructivist principles in which children engaged in in-depth inquiries with the guidance and support of teachers.

With developmentally appropriate practice becoming the conventional wisdom, there are fewer examples of curricula that used direct instruction as the main pedagogical method in the 1990s and beyond. In contrast to curriculum comparison studies conducted prior to 1983, more recent studies of naturally occurring variations in teaching practices (e.g., Marcon, 1992; Stipek et al., 1998) have found that children in developmentally appropriate settings outperform their counterparts in classrooms where DI is more the norm. Although most of these studies are not of high design quality, and therefore may confound the effects of the program with family and child characteristics, the findings of this body of research cloud the issue of determining which instructional type will more likely lead to improved student outcomes.

Complicating the issue further are the limited descriptions of teaching practices in the studies from which the T/A and T/C contrasts were drawn. Some of the contrasts used a specific curriculum model (e.g., Bereiter-Engelmann, DARCEE, Distar) that explicitly detail what is meant by direct instruction and the role of the teacher, but many of the contrasts examined are not as clear. As a consequence, the specific practices used by teachers in studies without an identifiable curriculum model are uncertain. Jacobs et al. (2004) defined *direct instruction* as instruction involving mostly teacher-directed activities designed to teach information and develop skills. Inquiry-based instruction, in contrast, involves mostly hands-on, student-directed learning, with the teacher acting as a facilitator. In the T/C and T/A contrasts, a number of programs were coded as direct instruction. These included Karnes Preschool, Direct Verbal, DARCEE, Distar, and, infrequently, two programs described as Adult Paraprofessionals and Teen Paraprofessionals. About

95% of the contrasts involving direct instruction were from studies completed prior to 1980, and 87% were completed prior to 1970. Given that more recent studies have found positive academic gains for children in programs in which teachers use more developmentally appropriate strategies, it is probably safer to conclude that the sum of this evidence provides support for teacher-directed or explicit instruction rather than DI per se as the primary method of teaching.

This meta-analysis also found that individuation (or "individualized" instruction) had a positive impact on cognitive and school outcomes in T/A contrasts. This finding is perhaps not surprising given that individual or small-group instruction is used widely in most early childhood curricula, no matter where they fall on the mentioned continuum. Smaller groups enable teachers to assess children's development and enact learning opportunities that help children engage with content and practice skills (Bredekamp & Copple, 1997). In other words, smaller groups and lower staff ratios provide more opportunity for teachers to match content to children's particular developmental levels so that they are able to learn various academic concepts. Similarly, Frede's (1998) analysis of the content of effective preschool programs found that by using small groups, children learn about classroom processes, such as sitting and paying attention to the teacher. Although it would seem to make sense that small-group instruction impacts children's cognitive development and helps them socialize into the culture of schooling, further clarification of what is taking place within the small groups in these contrasts would make it easier to discern exactly why this approach is effective (Graue, Clements, Reynolds, & Niles, 2004).

RANGE OF SERVICES

Another area of decision-making when implementing preschool programs concerns the range of services to be provided. Early childhood education has always had a commitment to development of the whole child. To this end, supporting families and providing a range of services to facilitate children's development is often considered an important component of effective preschool programs. Parent involvement is a central component of Chicago's Child Parent Centers, for example, and Head Start has always focused on children's health, nutrition, and educational needs. It was not surprising that in this study, Head Start accounted for a large proportion of the contrasts examined as providing additional services.

Despite the logic underpinning the provision of additional services, the results of this meta-analysis indicate that the children in the programs that provided these types of services did not perform as well as those who did not receive such services. The influence on cognitive outcomes may not be directly due to the extra services provided, but rather influences of other confounding variables. For example, children who received extra services tended to receive less direct instruction in larger groups and yet also received preschool for longer periods of time. As mentioned previously, it would seem to make sense that the provision of additional services could compete with instructional time. If a program is providing additional services but within the same time frame as other preschool programs whose sole focus is children's education, then the time available each day must be allotted differently so that teachers can provide these other services. Moreover, given that many of the contrasts with additional services involved Head Start, a program that offers health and education services to disadvantaged children, it is also likely that the performance of these children on cognitive measures was impacted by their personal circumstances.

Unfortunately, although some 310 contrasts employed at least one condition that involved additional services, the limited descriptions provided in many of the studies make it difficult to determine how the instructional time was used in these programs and with differing populations of students. More recent evaluations of preschool programs that offer additional services to targeted populations show positive outcomes. For example, a longitudinal evaluation of the Abbott preschool program (Frede, Jung, Barnett, Esposito Lamy, & Figueras, 2007) found that children who attended this program demonstrated substantial gains in language, literacy, and mathematics and that these gains were sustained through the kindergarten year.

The findings from this meta-analysis regarding additional services would suggest that policy makers should consider carefully not only what additional services, if any, they will provide but also how these services might be delivered in a way that does not dilute the intensity of children's preschool experience. Such decisions will require consideration of who will provide the additional services (teachers vs. others), for what proportion of the instructional day and week, and the main target of such services (children, families, or both).

DESIGN QUALITY

Although the findings of this meta-analysis provide some insight into the provision of preschool programs that will have an impact on children's development, one of the limitations of this analysis is the design quality of the studies examined. Larger effect sizes were associated with higher design quality (36% of T/C and 34% of T/A contrasts were coded as having high quality), and this result illustrates the risk of potentially underestimating program impact with weak research designs. In addition, this study also reveals the shortcomings of meta-analyses undertaken on studies conducted decades ago. The problem is that descriptive information, if not reported, is challenging to reconstruct—even if that information was well-known at the time. For example, a number of randomized studies are included in this study, but the details of the randomization procedures are obscure; in short, the fidelity of the randomized study is difficult to describe empirically. More important, variable labels suggest a constancy of treatment as though changes and innovations were not introduced to treatments such as DI over time. Finally, it was not possible to match intended outcomes with observed outcomes. This is to say that programs may have been advantaged or disadvantaged by the choice to average broadly across all outcomes in a particular domain.

LOOKING FORWARD

This study should be interpreted as a quantitative summary of the research from the period 1960–2000. This is one component of the validity argument regarding the efficacy of early childhood interventions. Yet, in examining the most contrasts to date, this meta-analysis demonstrates the need for randomized or carefully controlled experiments that detail the structural and process variables of differing preschool interventions. The findings raise many questions concerning teachers and staffing structure, instructional format, and the balance of additional services that will lead to short- and long-term improvements in children's learning and development, as well as academic and school success. When policy makers have access to a research base that documents clearly the relationships between child outcomes and instructional and programmatic characteristics, it might be possible once and for all to make policy implementation decisions with predictable outcomes.

Given the current state of research on the efficacy of early childhood interventions, there is both good and bad news. The good news is that a host of original and synthetic studies have found positive effects for a range of outcomes, and this pattern is clearest for outcomes relating to cognitive development. Moreover, many promising variables for program design have been identified and linked to outcomes, though little more can be said of the link other than it is positive. The bad news is that there is much less empirical information available for designing interventions at multiple levels with multiple components. Indeed, the extant literature is most accurately read as what has been effective in the past, and few, if any, of these studies can be read as design experiments. Thus, for example, it is not yet possible to combine an array of program elements in a way that would allow an estimate of program effect, and just as the magnitude of benefit remains somewhat murky, so does the cost.

Prospective controlled studies of early childhood program outcomes have much potential to remedy the uncertainties of quasi-experimental comparisons and, more important, to examine the efficacy of how program components are assembled and embedded in a multilevel context (e.g., school district, logistical and operational, community). Formal protocols could be useful to standardize program implementation and thus enhance the likelihood of replicability. A wealth of both qualitative and quantitative data could be collected not just for estimating programs effects but also for clarifying the precipitating mechanisms for those effects. Given that publicly funded preschool is expanding beyond targeted programs for disadvantaged students, new multisite randomized trials appear to have much potential benefit to advance our understanding about program efficacy and to ensure that all young children receive the quality early education they deserve.

Notes

1. See Cooper and Hedges (1994) for further methodological details on effect size computation. A description of the multilevel analysis approach to meta-analysis is provided by Raudenbush and Bryk (2002), and a readable introduction is given by de la Torre, Camilli, Vargas, and Vernon (2007).
2. Imputation allowed the investigation of a large number of moderator variables with different patterns of missing values. In the final models, very few missing values were imputed.
3. These two classes were combined for several reasons. First, the two measures had similar profiles of outcomes and were statistically indistinguishable. Second, at the preschool level, both kinds of measures are heavily driven by verbal knowledge. We recognize that there are important differences in the two constructs, but given the empirical similarities, the increased statistical power based on the combined sample must also be taken into account when judging whether it is prudent to report outcomes separately.

4 When analyses were performed using inverse variance weights, the same patterns of coefficients were observed, though the magnitude varied. It should be noted that studies of interventions employing individual or small-group instruction tended to have smaller sample sizes—sometimes much smaller—than studies without this instructional arrangement.

References

Ake, C. F. (2005). *Rounding after multiple imputation with non-binary categorical covariates.* SAS Focus session, SUGI 30. Retrieved January 18, 2008, from http://www2.sas.com/proceedings/sugi30/112-30.pdf

Allison, P. D. (2005). *Imputation of categorical variables with PROC MI.* SAS Focus session, SUGI 30. Retrieved January 18, 2008, from http://www2.sas.com/proceedings/sugi30/113-30.pdf

Anderson, L. M., Shinn, C., Fullilove M. T., Scrimshaw, S. C., Fielding, J. E., Normand, J., et al. (2003). The effectiveness of early childhood development programs. A systematic review. *American Journal Preventative Medicine, 24*(Suppl. 3), 32–46.

Barnett, W. S. (1996). *Lives in the balance: Age-27 benefit-cost analysis of the High/Scope Perry Preschool Program* (Monographs of the High/Scope Educational Research Foundation, 11). Ypsilanti, MI: High/Scope Press.

Barnett, W. S. (1998). Long-term effects on cognitive development and school success. In S. S. Boocock (Ed.), *Early care and education for children in poverty: Promises, programs, and long-term results* (pp. 11–44). Albany: State University of New York Press.

Barnett, W. S., Brown, K., & Shore, R. (2004). *The universal vs. targeted debate: Should the United States have preschool for all?* (NIEER Policy Brief, Issue 6). New Brunswick, NJ: National Institute for Early Education Research.

Bilmes, J. A. (1998). *A gentle tutorial of the EM algorithm and its application to parameter estimation for Gaussian mixture and hidden Markov Models.* Retrieved January 18, 2008, from the International Computer Science Institute Web site: http://ssli.ee.washington.edu/people/bilmes/mypapers/em.pdf

Bredekamp, S. (Ed.). (1987). *Developmentally appropriate practice in early childhood programs serving children from birth through age 8.* Washington, DC: National Association for the Education of Young Children.

Bredekamp, S., & Copple, C. (Eds.). (1997). *Developmentally appropriate practice in early childhood programs.* Washington, DC: National Association for the Education of Young Children.

Cooper, H., & Hedges, L. V. (1994). *The handbook of research synthesis.* New York: Russell Sage Foundation.

Cronbach, L. J. (1982). *Designing evaluations of educational and social programs.* San Francisco: Jossey-Bass.

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.

Durlak, J. A. (2003). The long-term impact of preschool prevention programs: A commentary [Editorial]. Retrieved June 18, 2007, from http://gateway.uk.ovid.com/gw2/ovidweb.cgi

Frede, E. C. (1998). Preschool program quality in programs for children in poverty. In W. S. Barnett & S. S. Boocock (Eds.), *Early care and education for children in poverty: Promises, practices, and long-term results* (pp. 77–98). Albany: State University of New York Press.

Frede, E., Jung, K., Barnett, W. S., Esposito Lamy, C., & Figueras, A. (2007). *The Abbott preschool program longitudinal effects study* (APPLES). New Brunswick, NJ: National Institute for Early Education Research.

Goffin, S. G., & Wilson, C. S. (2001). *Curriculum models and early childhood education: Appraising the relationship* (2nd ed.). Upper Saddle River, NJ: Prentice Hall.

Goldstein, H. (2003). *Multilevel statistical models* (3rd ed.). New York: Oxford University Press.

Gorey, K. M. (2001). Early childhood education: A meta-analytic affirmation of the short- and long-term benefits of educational opportunity. *School Psychology Quarterly, 16*(1), 9–30.

Gormley, W. T., Jr., & Gayer, T. (2003). *Promoting school readiness in Oklahoma: An evaluation of Tulsa's pre-K program* (CROCUS Working Paper No. 1). Washington, DC: Center for Research on Children in the U.S.

Gormley, W. T., Phillips, D., & Gayer, T. (2008). Preschool programs can boost school readiness. *Science, 320*, 1723–17-24.

Graue, E., Clements, M., Reynolds, A., & Niles, M. (2004). More than teacher directed or child initiated: Preschool curriculum type, parent involvement, and children's outcomes in the child-parent centers. *Education Policy Analysis Archives, 12*, 1–36.

Horton, N. J., Lipsitz, S. R., & Parzen, M. (2003). A potential for bias when rounding in multiple imputation. *American Statistician,* 57, 229–232.

Jacobs, R. T., Creps, C. L., & Boulay, B. (2004, July). *Meta-analysis of research and evaluation studies in early childhood education* (Final report to the National Institute of Early Education Research). Cambridge, MA: Abt Associates.

Karoly, L. A., Kilburn, M. R., & Cannon, J. S. (2005). E*arly childhood interventions: Proven results, future promises*. Retrieved June 14, 2006, from http://www.rand.org/pubs/monographs/2005/RAND_MG341.pdf

Layzer, J., Goodson, B. D., Bernstein, L., & Price, C. (2001). *National evaluation of family support programs final report. Volume A: The meta-analysis*. Cambridge, MA: Abt Associates.

MacLeod, J., & Nelson, G. (2000). Programs for the promotion of family wellness and the prevention of child maltreatment: A meta-analytic review. *Child Abuse and Neglect, 24,* 1127–1149.

Marcon, R. A. (1992). Differential effects of three preschool models on inner city 4-year-olds. *Early Childhood Research Quarterly, 7,* 517–530.

Nelson, G. (2005). Promoting the well-being of children and families: What is best practice? In J. Scott & H. Ward (Eds.), *Safeguarding and promoting the wellbeing of children, families and their communities* (pp. 184–196). London: Jessica Kingsley.

Nelson, G., & Westhues, A., & MacLeod, J. (2003). A meta-analysis of longitudinal research on preschool prevention programs for children. *Prevention and Treatment, 6*, 1–34.

NICHD Early Childcare Research Network Study. (2002). Early child care and children's development prior to school entry: Results from the NICHD Study of Early Child Care. *American Educational Research Journal, 39*, 133–164.

Noh, H., Kwak, M., & Han, I. (2004). Improving the prediction performance of customer behavior through multiple imputation. *Intelligent Data Analysis, 8*, 563–577.

Pigott, T. (2001). A review of methods for missing data. *Educational Research and Evaluation, 7*, 353–383.

Raudenbush, S. W. (1994). Random effects models. In H. Cooper & L.V. Hedges (Eds.), *The handbook of research synthesis* (pp. 301–322). New York: Russell Sage Foundation.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods.* Newbury Park, CA: Sage.

SAS Institute. (2008). *Multiple imputation for missing data.* Retrieved January 18, 2008, from http://support.sas.com/rnd/app/da/new/dami.html

Shadish, W. R., Robinson, L., & Lu, C. (1999). *ES: A computer program and manual for effect size calculation.* St. Paul, MN: Assessment Systems Corporation.

Stipek, D., Feiler, R., Byler, P., Ryan, R., Milburn, S., & Salmon, J. M. (1998). Good beginnings: What differences does the program make in preparing young children for school? *Journal of Applied Developmental Psychology, 19*, 41–66.

Switzer, F. S, III, Roth, P. L., & Switzer, D. M. (1998). Introduction to the feature on problematic data. *Organizational Research Methods, 6,* 279–281.

Vandell, D. L., & Wolfe, B. (2000). *Child care quality: Does it matter and does it need to be improved?* Madison: Institute for Research on Poverty, University of Wisconsin.

White, K. R. (1985). Efficacy of early intervention. *Journal of Special Education, 19,* 401–416.

White, K. R., Taylor, J. J., & Moss, V. D. (1992). Does research support claims about the benefits of involving parents in early intervention programs? *Review of Educational Research, 62,* 92–125.

Yuan, Y. C. (2004). *Multiple imputation for missing data: Concepts and new development.* Retrieved January 18, 2008, from http://www.ats.ucla.edu/stat/sas/library/multipleimputation.pdf

APPENDIX

SELECT STUDY BIBLIOGRAPHY

0001. Masse, L. N., & Barnett, W. S. (2002). *A benefit cost analysis of the Abecedarian early childhood intervention* (pp. 1–50). New Brunswick, NJ: National Institute for Early Education Research.

0002. Schweinhart, L. J. (2000). The High/Scope Perry Preschool Study: A case study in random assignment. *Evaluation and Research in Education, 14*(3&4), 136–147.

0003. Schweinhart, L. J., & Weikart, D. P. (1997). *Lasting differences: the High/Scope Preschool curriculum comparison study through age 23.* (Monographs of the High/Scope Educational Research Foundation). Ypsilanti, MI: High/Scope Press.

0004. Gray, S. W., Ramsey, B. K., & Klaus, R. A. (1983). The Early Training Project, 1962–1980. In Consortium for Longitudinal Studies (Eds.), *As the twig is bent . . . lasting effects of preschool programs* (pp. 33–70). Hillsdale, NJ: Erlbaum.

0006. Garber, H. L. (1988). *The Milwaukee Project: Preventing mental retardation in children at risk.* Washington, DC: American Association on Mental Retardation.

0007. Wasik, B. H., Ramey, C. T., Bryant, D. M., & Sparling, J. J. (1990). A longitudinal study of two early intervention strategies: Project CARE. *Child Development, 61,* 1682–1696.

0008. Spaulding, R. L. (1973, February). *A coping analysis schedule for educational settings (CASES).* Annual meeting of the American Educational Research Association, New Orleans, LA.

0009. Jordan, T. J., Grallo, R., Deutsch, M., & Deutsch, C. P. (1985). Long-term effects of early enrichment: A 20-year perspective on persistence and change. *American Journal of Community Psychology, 13*, 393–415.

0010. Engelmann, S., & Osborn, J. (1976). *Distar Language I: An instructional system. Teacher's guide* (2nd ed.). Chicago: Science Research Associates.

0011. Karnes, M. B., Shwedel, A. M., & Williams, M. B. (1983). A comparison of five

approaches for educating young children from low-income homes. In Consortium for Longitudinal Studies (Eds.), *As the twig is bent . . . lasting effects of preschool programs* (pp. 133–179). Hillsdale, NJ: Erlbaum.

0012. Blatt, B., & Garfunkel, F. (1969). *The educability of intelligence: Preschool intervention with disadvantaged children.* Washington, DC: Council for Exceptional Children Inc.

0013. Weiss, R. S. (1981). “INREAL intervention for language handicapped and bilingual children.” *Journal of the Division for Early Childhood, 4*, 40–51.

0014. Miller, L. B., & Bizzell, R. P. (1984). Long-term effects of four preschool programs: Ninth- and tenth-grade results. *Child Development, 55*, 1570–1587.

0015. Reynolds, A. J. (2001). *Success in early intervention: The Chicago Child-Parent Centers.* Lincoln: University of Nebraska Press.

0016. Lally, J. R., Mangione, P. L., & Honig, A. S. (1988). The Syracuse University Family Development Research Program: Long-range impact of an early intervention with low-income children and their families. In D. R. Powell (Ed.), *Parent education as early childhood intervention: Emerging directions in theory research and practice* (pp.79–104). Norwood, NJ: Ablex.

0018. Copple, C. E., Cline, M. G., & Smith, A. N. (1987). *Path to the future: Long-term effects of Head Start in the Philadelphia school district.* Washington, DC: U.S. Department of Health and Human Services.

0019. Hauser-Cram, P., Pierson, D. E., Walker, D. K., & Tivnan, T. (1991). *Early education in the public schools: Lessons from a comprehensive birth-to-kindergarten program.* San Francisco: Jossey-Bass.

0020. Northwest Regional Educational Laboratory. (2000). *An investment in children and families: Year 9 & 10 longitudinal study report.* Portland, OR: Child Family and Community Program NWREL, Washington State Community, Trade and Economic Development.

0021. Cornett, J. D., & Askins, B. E. (1978, March). *The long range effects of a “special” intervention program for low birth weight children: Some findings and methodologies.* Paper presented at the annual meeting of the American Educational Research Association, Toronto, Ontario, Canada.

0022. Barnett, W. S., & Camilli, G. (2000). Compensatory preschool education, cognitive development, and “race.” In J. Fish (Ed.), *Race and intelligence: Separating science from myth.* Mahwah, NJ: Lawrence Erlbaum Associates.

0023. Ellinwood, B. W. (1971). *Early School Admissions Program: 1969–70 evaluation (*ERIC Document Reproduction Service No. *ED 055675).* Baltimore: Baltimore City Public Schools.

0024. Smith, M. S., & Bissell, J. S. (1970). Report analysis: The impact of Head Start. *Harvard Educational Review, 40*, 51–104.

0025. Goodman, J. F., Cecil, H. S., & Barker, W. F. (1984). Early intervention with retarded children: Some encouraging results. *Developmental Medicine and Child Neurology, 26*, 45–55.

0026. Epstein, J. N. (1994). *Accelerating the literacy development of disadvantaged preschool children: An experimental evaluation of a Head Start emergent literacy curriculum.* Doctoral dissertation, State University of New York at Stony Brook. *Dissertation Abstracts International: Section B: The Sciences & Engineering, 55*(11-B): 5065.

0028. Peta, E. J. (1973). *The effectiveness of a total environment room on an early reading program for culturally different pre-school children.* Dissertation, Lehigh University, Bethlehem, PA.

0029. Abelson, W.D. (1974). Head Start graduates in school: Studies in New Haven, Connecticut. In S. Ryan (Ed.), *A report on longitudinal evaluations of preschool programs* (Vol. 1, pp.1–14). Washington, DC: Office of Child Development, US Department of Health, Education and Welfare.

0032. Gamse, B. C., Conger, D., Nelson, D., & McCarthy, M. (1996). *Follow-up study of families in the Even Start in-depth study.* Draft final report. Cambridge, MA: Abt Associates.

0033. Goodstein, H. A. (1971). The use of a structured curriculum with black preschool disadvantaged children. *Journal of Negro Education, 40*, 330–336.

0034. Adkins, D., & Herman, H. (1970). *Hawaii Head Start evaluation: 1968–69* (Final report, FS 17.638 [ERIC Document Reproduction Service No. ED 042 511]). Washington, DC: US Department of Health and Education.

0036. Hebbeler, K. (1985). An old and a new question on the effects of early education for children from low income families. *Educational Evaluation and Policy Analysis, 7*, 207–216.

0038. Bittner, M., Rockwell, M., & Matthews, C. (1968). *An evaluation of the preschool readiness centers' program in East St. Louis, July 1, 1967–June 30, 1968* (Final Report, ERIC Document Reproduction Service No. ED 023 472). East St. Louis, MO: Southern Illinois University, Center for the Study of Crime, Delinquency, and Corrections.

0039. Scott, R. (1974). Research and early childhood: The Home Start Project. *Child Welfare, 53*(2), 112–119.

0040. Edwards, J., & Stern, C. (1969). *A comparison of three intervention programs with disadvantaged preschool children: Final Report 1968–1969.* Los Angeles: University of California Head Start Research and Evaluation Center. (ERIC Document Reproduction Service No. ED 041616)

0041. Nedler, S., & Sebera, P. (1971). Intervention strategies for Spanish-speaking preschool children. *Child Development, 42*, 259–267.

0042. Wolff, M., & Stein, A. (1967). *Long-range effect of pre-schooling on reading achievement.* New York: Yeshiva University, Ferkauf Graduate School of Education.

0043. Alpern, G. D. (1966). The failure of a nursery school enrichment program for culturally disadvantaged children. *American Journal of Orthopsychiatry, 36*, 244–245.

0045. McAfee, O. (1972). An integrated approach to early childhood education. In J. C. Stanley (Ed.), *Preschool programs for the disadvantaged: Five experimental approaches to early childhood education* (pp. 67–91). Baltimore: Johns Hopkins University Press.

0046. Kohlberg, L. (1968). Montessori with the culturally disadvantaged: A cognitive-developmental interpretation and some research findings. In R. D. Hess & R. M. Bear (Eds.), *Early education: Current theory, research, and action* (pp. 105–118). Chicago: Aldine.

0048. Hodes, M. R. (1966). *An assessment and comparison of selected characteristics among culturally disadvantaged kindergarten children who attended Project Head Start (summer program 1965); culturally disadvantaged kindergarten children who did not attend Project Head Start; and kindergarten children who were not culturally disadvantaged.* Glassboro, NJ: Glassboro State College. (ERIC Document Reproduction Service No. ED 014 330)

0049. Herzog, E., Newcomb, C. H., & Cisin, I. H. (1974). Double deprivation: The less they have, the less they learn. In S. Ryan (Ed.), *A report on longitudinal evaluations of preschool programs* (Vol. 1, pp.69–93). Washington, DC: Office of Child Development, U.S. Department of Health, Education and Welfare.

0050. Plant, W. T., & Southern, M. L. (1972). The intellectual and achievement effects of preschool cognitive stimulation of poverty Mexican-American children. *Genetic Psychology Monographs, 86*, 141–173.

0054. Engelmann, S. (1970). The effectiveness of direct instruction on IQ performance achievement in reading and arithmetic. In J. Hellmuth (Ed.), *Disadvantaged child: Compensatory education: A national debate* (Vol. 3, pp. 339–361). New York: Brunner/Mazel.

0055. Woolman, M. (1983). The micro-social learning environment: A strategy for accelerating learning. In Consortium for Longitudinal Studies (Eds.), *As the twig is bent . . . lasting effects of preschool programs* (pp. 265–297). Hillsdale, NJ: Erlbaum.

0057. Quay, L. C., Kaufman-McMurrain, M., Minore, D. A., Cook, L., & Steele, D. C. (1996). *The longitudinal evaluation of Georgia's Prekindergarten Program: Results for the third year.* American Educational Research Association. New York: Georgia State University.

0058. Clements, D. H. (1983, April). *Training effects on the development and generalization of Piagetian logical operations and counting strategies.* Paper presented at the biennial meeting of the Society for Research in Child Development, Detroit, MI.

0059. Webb, R. A. (1974). *The second-year evaluation of the style-oriented cognitive curriculum in the I.V.Y. (Involving the Very Young) program of the Baltimore City Public Schools and the evaluation of the Baltimore City Day Care Center Training Program.* (ERIC Document Reproduction Service No. ED 130 769)

0060. Carlson-Kest, E. (1969, November). *Programmed versus incidental teaching in the prekindergarten.* Paper presented at the annual meeting of the National Association for the Education of Young Children, Salt Lake City, UT.

0061. State Education Department of the University of the State of New York. (1982). *Evaluation of the New York State experimental prekindergarten program: Final report.* Albany: New

York State Education Department. (ERIC Document Reproduction Service No. ED 219 123)

0064. Pietrangelo, D. J. (1999). Outcomes of an enhanced literacy curriculum on the emergent literacy skills of Head Start preschoolers. Doctoral dissertation. School of Education, Division of School Psychology, State University of New York at Albany. *Dissertation Abstracts International Section A: Humanities & Social Sciences, 60*(4-A): 1014.

0066. Wolff, M., & Stein, A. (1967). Head Start six months later. *Phi Delta Kappan, 48,* 349–350.

0067. Harding, J. (1966). *A comparative study of various project Head Start programs.* Ithaca: State University of New York. (ERIC Document Reproduction Service No. ED 019 987)

0068. Holmes, D., & Holmes, M. B. (1966). *An evaluation of differences among different classes of Head Start participants: Final report.* New York: Associated YM-HWHAS of Greater New York. (ERIC Document Reproduction Service No. ED 015 012)

0069. Capobianco, R. J. (1967). A pilot project for culturally disadvantaged preschool children. *Journal of Special Education, 1*, 191–194.

0070. Lee, V. E., Brooks-Gunn, J., Schnur, E., & Liaw, F. R. (1990). Are Head Start effects sustained? A longitudinal follow-up comparison of disadvantaged children attending Head Start, no preschool, and other preschool programs. *Child Development 61,* 495–507.

0071. Abbott-Shim, M., Lambert, R., & McCarty, F. (2003). A comparison of school readiness outcomes for children randomly assigned to a Head Start program and the program's wait list. *Journal of Education for Students Placed at Risk, 8*(2), 191–214.

0101. Braun, S. J., & Caldwell, B. M. (1973). Emotional adjustment of children in day care who enrolled prior to or after the age of three. *Early Child Development and Care, 2*, 13–21.

0124. Pinkelton, N. B. H. (1976). *A comparison of referred Head Start, non-referred Head Start and non-Head Start groups of primary school children on achievement, language processing, and classroom behavior.* Doctoral dissertation, University of Cincinnati, Cincinnati, OH.

0130. Marcon, R. A. (1992). Differential effects of three preschool models on inner-city 4-year-olds. *Early Childhood Research Quarterly, 7,* 517–530.

0133. Clark, C. M. (1979). *Effects of the project Head Start and Title I preschool programs on vocabulary and reading achievement measured at the kindergarten and fourth grade levels.* Doctoral dissertation, Wayne State University, Detroit, MI.

0140. Stipek, D. J., Feiler, R., Byler, P., Ryan, R., Milburn, S., & Salmon, J. (1998). Good beginnings: What difference does the program make in preparing young children for school? *Journal of Applied Developmental Psychology, 19,* 41–66.

0141. Cawley, J. F., Burrow, W. H., & Goodstein, H. A. (1970). Performance of Head Start and non-Head Start participants at first grade. *Journal of Negro Education, 39*(2), 124–131.

0148. Hemmeter, M. L., Wilson, S. M., Townley, K. F., Gonzalez, L., Epstein, A., & Hines, H. (1997). *Third party evaluation of the Kentucky Education Reform Act preschool programs.* Lexington: University of Kentucky, College of Education and College of Human Environmental Sciences.

0165. Evans, E. (1985). Longitudinal follow-up assessment of differential preschool experience for low income minority group children. *Journal of Educational Research, 78*(4), 197–202.

0168. Robison, H. F. (1968). *Data Analysis, 1967-68. Cue Project—CHILD—Curriculum to Heighten Intellectual and Language Development: Disadvantaged prekindergarten children, Central Harlem, New York City, 1-2. Office of Education (DHEW).* New York: Center for Urban Education.

0170. Hyman, I. A., & Kliman, D. S. (1967). First grade readiness of children who have had summer Head Start programs. *Training School Bulletin, 63*, 163–167.

0172. Steglich, W. G., & Cartwright, W. J. (1965). *Report of the effectiveness of Project Head Start, Lubbock, Texas. Parts I, II, and appendices.* Lubbock: Texas Technological College. (ERIC Document Reproduction Service No. ED 019 131)

0174. Wexley, K., Guidubaldi, J., & Kehle, T. (1974). An evaluation of Montessori and day care programs for disadvantaged children. *Journal of Educational Research, 68*(3), 95-99.

0175. Smith, M. P. (1968). Intellectual differences in five-year-old underprivileged girls and boys with and without pre-kindergarten school experience. *Journal of Educational Research, 61*, 348–350.

0177. Holmes, D., & Holmes, M. D. (1965). *Evaluation of two associated YM-YWHA Head Start programs.* New York: Associated YM-YWHAS of Greater New York. (ERIC Document Reproduction Service No. ED 014 318)

0178. Mosley, B. B., & Plue, W. V. (1980). *A comparative study of four curriculum programs for disadvantaged preschool children.* Hattiesburg: University of Southern Mississippi.

0180. Minnick, K. F. (1991). *Evaluation of the Family Development Program: Technical report.* Albuquerque, NM: Family Development Program.

0185. Stodolsky, S. S., & Karlson, A. L. (1972). Differential outcomes of a Montessori curriculum. *Elementary School Journal, 72*, 419–433.

0187. Larsen, J. M., & Robinson, C. C. (1989). Later effects of preschool on low-risk children. *Early Childhood Research Quarterly, 4*, 133–144.

0190. Horton, K. B., McConnell, F., & Smith, B. R. (1969). Language development and cultural disadvantagement. *Exceptional Children, 35*, 597–606.

0191. Friedman, M. I., Lackey, G. H., Mandeville, G. K., & Statler, C. R. (1970). *An investigation of the relative effectiveness of selected curriculum variables in the language development of Head Start children.* Columbia: Evaluation and Research Center for Project Head Start, University of South Carolina. (ERIC Document Reproduction Service No. ED 046 497)

0192. Krider, M. A., & Petsche, M. (1967). *An evaluation of Head Start pre-school enrichment programs as they affect the intellectual ability, the social adjustment, and the achievement level of five-year-old children enrolled in Lincoln, Nebraska.* Lincoln: University of Nebraska. (ERIC Document Reproduction Service No. ED 015 011)

0193. Chalkey, M. A., & Leik, R. K. (1997). The impact of escalating family stress on the effectiveness of Head Start intervention. *National Head Start Association Research Quarterly, 1*(1), 157–162.

0195. Horowitz, F. D., & Rosenfeld, H. (1966). *Comparative studies of a group of Head Start and a group of non-Head Start preschool children.* Washington, DC: University of Kansas Project Head Start Research and Evaluation. (ERIC Document Reproduction Service No. ED 015 013)

0196. Team, S. S. E. (1999). *A six-county study of the effects of Smart Start child care on kindergarten entry skills.* Chapel Hill, NC: Frank Porter Graham Child Development Center. (ERIC Document Reproduction Service No. ED 433 154)

0198. Di Lorenzo, L. T. (1968). Effects of year-long prekindergarten programs on intelligence and language of educationally disadvantaged children. *Journal of Experimental Education, 36*, 36–39.

0199. Sontag, M., Sella, A. P., & Thorndike, R. L. (1969). The effect of Head Start training on the cognitive growth of disadvantaged children. *The Journal of Educational Research, 62*, 387–389.

0200. Stern, C. (1969). *The effectiveness of a standard language readiness program as a function of teacher differences.* 17. Office of Economic Opportunity. Washington, DC. University of California, Los Angeles. (ERIC Document Reproduction Service No. ED 039932)

0201. St Pierre, R., Ricciuti, A., Tao, F., Creps, C., Swartz, J., Lee, W., & Parsad, A. (2003). *Third national Even Start evaluation: Program impacts and implications for improvement.* Washington, DC: U.S. Department of Education, Planning & Evaluation Service.

0202. Burchinal, M. A., Lee, M., & Ramey, C. T. (1989). Type of day-care and preschool intellectual development in disadvantaged children. *Child Development, 60*, 128–137.

0204,0216. Seefeldt, C. (1977, April). *Montessori and responsive environment models: A longitudinal study of two preschool programs, phase two.* Paper presented at the annual meeting of the American Educational Research Association, New York, NY.

0205. Fleege, U. H., Black, M., & Rackausas, H. (1967). *Montessori preschool education: Final report.* (ERIC Document Reproduction Service No. ED 017 320)

0208. Hubbard, J., & Zarate, L. T. (1967). *An exploratory study of oral language development among culturally different children.* Austin: University of Texas Child Development Evaluation and Research Center. (ERIC Document Reproduction Service No. ED 019 120)

0209. Morris, B., & Morris, G. L. (1966). *Evaluation of changes occurring in children who participated in project Head Start*. Kearney, NE: Kearney State College. (ERIC Document Reproduction Service No. ED 017 316)

0211. Cataldo, C. Z. (1978). A follow-up study of early intervention, University of New York-Buffalo, 1977. *Dissertation Abstracts International, 39*(2-A), 657.

0212. Vance, B. J. (1967). *The effect of preschool group experience on various language and social skills in disadvantaged children. Final report.* (ERIC Document Reproduction Service No. ED 019 989)

0213. Adkins, D. C. (1969). *Preschool Mathematics Curriculum Project. Final report.* (ERIC Document Reproduction Service No. ED 038 168)

0215. Sprigle, H. (1974). Learning to learn. In S. Ryan (Ed.), *A report on longitudinal evaluations of preschool programs* (Vol. 1, pp. 109–124). Washington, DC: Office of Child Development, U.S. Department of Health, Education and Welfare.

0217. Ametjian, A. (1965). *The effects of a preschool program upon the intellectual development and social competency of lower class children.* Doctoral dissertation, Stanford University, Palo Alto, CA.

0218. Howard, J. L., & Plant, W.T. (1967). Psychometric evaluation of an operant Head Start program. *Journal of Genetic Psychology, 111*, 281–288.

0224. Perry, D.G. (1999). *A study to determine the effects of pre-kindergarten on kindergarten readiness and achievement in mathematics.* Master's thesis. Master of arts degree program, Salem-Teikyo University, Salem, WV.

0225. Schmitt, D. R. (1994, September). *Longitudinal study of a bilingual program for four year olds.* Paper presented at the annual meeting of the Association of Louisiana Evaluators, New Orleans, LA.

0227. Weisberg, P. (1988). Direct instruction in the preschool. *Education and Treatment of Children, 11*, 349–363.

0228. Anderson, B. (1994). *The effect of preschool education on academic achievement of at risk children.* (ERIC Document Reproduction Service No. ED 382 318)

0229. Munoz, M .A. (2001). *The critical years of education for at-risk students: The impact of early childhood programs on student learning.* Louisville, KY: Jefferson County Public Schools. (ERIC Document Reproduction Service No. ED 456 913)

0230. Garces, E., Thomas, D., & Currie, J. (2000). *Longer term effects of Head Start.* Santa Monica, CA: RAND.

0232. Xiang, Z., & Schweinhart, L. J. (2002). *Effects five years later: The Michigan School Readiness Program evaluation through age 10.* Ypsilanti, MI: High/Scope Educational Research Foundation.

0234. Larson, D.E. (1972). *Stability of gains in intellectual functioning among White children who attended a preschool program in rural Minnesota: Final report.* Mankato, MN: Mankato State College, Office of Education (DHEW). (ERIC Document Reproduction Service No. ED 066 227)

0237. Porter, P. J., Leodas, C., Godley, R. A., & Budroff, M. (1965). *Evaluation of Head Start educational program in Cambridge, Massachusetts: Final Report.* Cambridge, MA: Harvard University. (ERIC Document Reproduction Service No. ED 013 668)

0239. Epstein, A. S. (1993). *Training for quality: improving early childhood programs through systematic in-service training* (Monographs of the High/Scope Educational Research Foundation, 9). Ypslanti, MI: High/Scope Educational Research Foundation.

0240. Hartford City Board of Education. (1973). *Child development—Head Start program.* Hartford, CT. (ERIC Document Reproduction Service No. ED 086 365)

0241. McNamara, J. R. (1968). *Evaluation of the effects of Head Start experience in the area of self-concept, social skills, and language skills. Pre-publication draft.* Miami, FL: Dade County Board of Public Instruction. (ERIC Document Reproduction Service No. ED 028 832)

0243. Sandoval-Martinez, S. (1982). Findings from the Head Start Bilingual Curriculum Development Effort. *NABE: The Journal for the National Association for Bilingual Education, 7*(1), 1–12.

0249. Tamminen, A. W., Weatherman, R. F., & McKain, C. W. (1967). *An evaluation of a preschool training program for culturally deprived children. Final report.* U.S. Department of Health, Education, and Welfare. Duluth: University of Minnesota. (ERIC Document Reproduction Service No. ED 019 135)

0250, 0281. Barnett, W. S., Frede, E. C., Mobasher, H., & Mohr, P. (1987). The efficacy of public preschool programs and the relationship of program quality to efficacy. *Education Evaluation and Policy Analysis, 10*(1), 37–49.

0252. Stipek, D. J., Feiler, R., Daniels, D., & Milburn, S. (1995). Effects of different instructional approaches on young children's achievement and motivation. *Child Development, 66*, 209–223.

0254. Banta, T. J. (1968). *The Sands School Project: First-year results.* Cincinnati, OH: University of Cincinnati. (ERIC Document Reproduction Service No. ED 054 870)

0255. Allerhand, M. (1967). Effectiveness of parents of Head Start children as administrators of psychological tests. *Journal of Consulting Psychology, 31*, 286–290.

0258. Rentfrow, R. K. (1972). *Intensive evaluation of Head Start implementation in the Tucson Early Education model.* Office of Child Development (DHEW). Tucson: University of Arizona, Arizona Center for Educational Research and Development. (ERIC Document Reproduction Service No. ED 071 778)

0259. Alexanian, S. (1967). *Report D-I, Language Project: The effects of a teacher developed preschool language training program on first grade reading achievement.* Boston: Boston University, Head Start Evaluation and Research Center. (ERIC Document Reproduction Service No. ED 022 563)

0262. Dunlap, J. M., & Coffman, A. O. (1970). *The effects of assessment and personalized programming on subsequent intellectual development of prekindergarten and kindergarten children: Final report.* Office of Education (DHEW). University City, MO: University City Schools District. (ERIC Document Reproduction Service No. ED 045 198)

0265. Vondrak, M. (1996). *The effect of preschool education on math achievement.* (ERIC Document Reproduction Service No. ED 399 017)

0266. Hulan, J. R. (1972). Head Start Program and early school achievement. *Elementary School Journal, 73,* 91–94.

0269. Kasten, W. C., & Clarke, B. K. (1989). *Reading/writing readiness for preschool and kindergarten children: A whole language approach.* Sanibel: Florida Educational Research and Development Council, Inc. (ERIC Document Reproduction Service No. ED 312 041)

0273. Sevigny, K. S. (1987). *Thirteen years after preschool—Is there a difference?* Detroit, MI: Detroit Public Schools, Office of Instructional Improvement. (ERIC Document Reproduction Service No. ED 299 287)

0278. Bowlin, F. S., & Clawson, K. (1991, November). *The effects of preschool on the achievement of first, second, third, and fourth grade reading and math students.* Paper presented at the annual meeting of the Mid-South Educational Research Association, Lexington, KY.

0279. Nummedal, S. G., & Stern, C. (1971, February). *Head Start graduates: One year later.* Paper presented at the annual meeting of the American Educational Research Association, New York, NY.

0199, 0280. Thorndike, R. L. (1966). *Head Start Evaluation and Research Center, Teachers College, Columbia University. Annual report (1st), September 1966-August 1967.* New York: Columbia University Teachers College. (ERIC Document Reproduction Service No. ED 020 781)

0284. Cline, M. G., & Dickey, M. (1968). *An evaluation and follow-up study of summer 1966 Head Start children in Washington, DC.* Washington, DC: Howard University. (ERIC Document Reproduction Service No. ED 020 794)

GREGORY CAMILLI is Professor, University of Colorado at Boulder, School of Education. His current research interests include meta-analysis, educational effectiveness, differential item functioning, and affirmative action in law school admission. His publications include *Summarizing Item Difficulty Variation With Parcel Scores* (Camilli, Prowker, Dossey, Lindquist, Chiu, Vargas, & de la Torre, 2008); *Illustration of a Multilevel Model for Meta-Analysis* (de la Torre, Camilli, Vargas, & Vernon, 2007); and

Handbook of Complementary Methods in Education Research (Green, Camilli, & Elmore, 2006).

SADAKO VARGAS is Research Associate in the Graduate School of Education at Rutgers, The State University of New Jersey. Her research interests meta-analysis and the effectiveness of occupational therapy. Her publications include *A Meta-Analysis of Research on Sensory Integration Therapy*(Vargas & Camilli, 1999); *The Origin of the National Reading Panel: A Response to "Effects of Systematic Phonics Instruction Are Practically Significant"* (Camilli, Kim, & Vargas, forthcoming); and *Teaching Children to Read: The Fragile Link Between Science and Federal Education Policy* (Camilli, Vargas, & Yurecko, 2003).

SHARON RYAN is Associate Professor of Early Childhood and Elementary Education in the Graduate School of Education at Rutgers, The State University of New Jersey. Her research focuses on early childhood teacher education, curriculum, and policy. Her publications include *Creating an Effective System of Teacher Preparation and Professional Development: Conversations with Stakeholders* (Lobman & Ryan, 2008) and the newly published report, *Partnering for Preschool: A Study of Center Directors in New Jersey's Mixed Delivery Abbott Programs* (Whitebook, Ryan, Kipnis, & Sakai, 2008).

W. STEVEN BARNETT is Board of Governors Professor and Director of the National Institute for Early Education Research (NIEER) at Rutgers University. His research includes studies of the economics of early care and education, including costs and benefits, the long-term effects of preschool programs on children's learning and development, and the distribution of educational opportunities. His publications include *The State of Preschool 2007: State Preschool Yearbook* (Barnett, Hustedt, Friedman, Boyd, & Ainsworth, 2007); *Boundaries With Early Childhood Education: The Influence of Early Childhood Policies on Elementary and Secondary Education* (Barnett & Ackerman, 2007); and *Early Childhood Program Design and Economic Returns: Comparative Benefit-Cost Analysis of the Abecedarian Program and Policy Implications* (Barnett & Masse, 2007).

2010. â€œMeta-Analysis of the Effects of Early Education Interventions on Cognitive and Social Development.â€ Teachers College Record 112.3: 579-620. Cunha, Flavio, James J. Heckman, and Susanne M. Schennach. 2010. "Estimating the Technology of Cognitive and Noncognitive Skill Formation," Econometrica 78.3:883-931. Durlak, Joseph A.. Roger P. Weissberg, Allison B. Dymnicki, Rebecca D. Taylor, and Kriston B. Schellinger. 2011. "The Impact of Enhancing Students' Social and Emotional Learning: A Meta-Analysis of School-Based Universal Interventions." Child Development 82.1: A Meta-Analysis of Research Examining the Relationship between Mindfulness and Posttraumatic Stress Symptoms Kevyn Ziemann, 2019 Advisor: Patricia Veach and Thomas Hummel. A Word, a Shadow, a Breath: A Phenomenological Investigation of Therapists' Perceptions of the Stigma Experienced by Women Residing in Ireland Who Have had Abortions Meredith Anne Martyr, 2018 Advisor: Patricia Veach.Â A study of peer-nominated exemplars of social justice commitment in counseling and psychology Adam D. Sumner, 2013 Advisor: Michael P. Goh.Â Development and validation of the master psychotherapist characteristics inventory Fengqin Lian, 2012 Advisors: Thomas J. Hummel and Kay Herting Wahl.