

**Evaluating early care and education programs:  
A review of research methods and findings**

Carol Horton  
Erikson Institute<sup>1</sup>

Report prepared for the  
National Early Childhood Accountability Task Force  
Pew Charitable Trusts  
June 2006<sup>2</sup>

---

<sup>1</sup> Contact information: Carol Horton, Associate Research Scientist, Herr Research Center for Children and Social Policy, Erikson Institute, 420 N. Wabash Avenue, Chicago, Illinois, 60611; 312/893-7196 (phone), 312/755-0928 (fax); [chorton@erikson.edu](mailto:chorton@erikson.edu) (e-mail).

<sup>2</sup> An earlier version of this report was prepared for the Chicago Program Evaluation Project (C-PEP) with support from the Chicago Public Schools and Chicago Department of Children and Youth Services. Thanks to John Love and Sam Meisels for reviewing both this report and its C-PEP precursor and providing helpful commentary. Thanks also to Robert Pianta for his review and comments on the C-PEP report. All statements made in this report are the responsibility of the author alone. All views expressed are likewise those of the author only and do not represent those of Erikson Institute or any other party.

## **Table of Contents**

Executive summary .....	i
Introduction .....	1
Program evaluation and accountability .....	2
Overview of ECE program research .....	4
Notable research initiatives .....	7
<i>Experimental designs</i> .....	7
The Head Start Impact Study .....	8
<i>Quasi-experimental designs</i> .....	12
Chicago Child-Parent Centers .....	13
Oklahoma Universal Pre-K Program .....	16
NIEER Five-State Pre-Kindergarten Study .....	19
<i>Non-experimental designs</i> .....	21
Family and Child Experiences Survey (FACES) .....	21
Cost, Quality, and Outcomes Study .....	25
The NICHD Study of Early Child Care .....	28
Early Childhood Longitudinal Study – Birth Cohort .....	31
Early Childhood Longitudinal Study – Kindergarten Cohort .....	32
Georgia Early Childhood Study .....	37
NCEDL Multi-State Pre-K Study and SWEEP .....	39
Overview of research questions and methods .....	43
Synthetic interpretation of research findings .....	45
Conclusion: Implications for ECE program evaluation and accountability .....	51

## Evaluating early care and education programs: A review of research methods and findings

### Executive Summary

During the past 10-15 years, research on the relationship between participation in early care and education (ECE) programs and child development has grown substantially. Rigorous research designs, formerly the provenance of studies of small model programs, have been used increasingly to examine widely available, large-scale programs, most notably Head Start, child care, and state pre-k. At the same time, research on different ECE program types has come to focus on a common set of issues, most notably the significance of the relationships among program structure, teacher characteristics, classroom quality, individual child and family characteristics, and child outcomes, including both cognitive and social-emotional development.

This body of research includes both ECE program evaluations and other ECE program studies that, while not evaluations per se, have generated (or are in the process of generating) important findings regarding the relationship between ECE program participation and children's development. This report presents the research designs and primary findings of several recently completed or ongoing studies of Head Start, child care, state pre-k, and other widely available ECE program types. Specifically, these include the following:

- The *Head Start Impact Study*, a six-year impact evaluation designed to determine the effects of program participation on children's school readiness. This study is one of the few major ECE program evaluations to employ an experimental design, which randomly assigns study subjects to participate in treatment and control groups.
- A longitudinal study of the *Chicago Child-Parent Centers (CPC)*, an evaluation of the *Oklahoma Universal Pre-K Program*, and the *NIEER Five-State Pre-Kindergarten Study*. These studies employ quasi-experimental methods (i.e., a matched comparison group or regression-discontinuity design) to determine the effects of participation in a particular ECE program or a set of programs.
- Seven non-experimental studies that do not employ a non-treatment comparison group. This includes the *Family and Child Experiences Survey (FACES)*, which is tracking three nationally representative cohorts of Head Start children through first grade and the *Cost, Quality, and Outcomes Study*, which examined the relationship between child care cost, classroom quality, and child outcomes from preschool through second grade. Three longitudinal studies, the *NICHD Study of Early Child Care, Early Childhood Longitudinal Study – Birth Cohort (ECLS – B)*, and *Early Childhood Longitudinal Study – Kindergarten Cohort (ECLS-K)* are examining the relationships among a wide range of early child experiences and child outcomes, including the association between ECE program participation and children's cognitive and social-emotional development. Other studies of pre-k include an

evaluation of the *Georgia Early Childhood Program*, which compares outcomes among children who attended state pre-k, Head Start, private preschool, or no preschool as four-year-olds and the *NCEDL Multi-State Pre-K Study* and *Study of State-Wide Early Education Programs (SWEEP)*, which are analyzing how important variations in state pre-k programs relate to child outcomes at the end of pre-k and kindergarten.

These research initiatives have collected (or are in the process of collecting) a wide range of data in order to answer their primary research questions concerning the relationship between ECE program participation and children's development. The most important and frequently collected sets of data include (1) assessments of children's cognitive development using individually administered standardized, normed instruments; (2) assessments of children's social-emotional development using teacher-administered instruments; (3) child and family demographic information (e.g., race/ethnicity, gender, home language, family income); (4) teacher characteristics (e.g., race, gender, education and training, teaching experience); (5) assessments of classroom quality using externally administered, standardized instruments; and (6) ECE program characteristics (e.g., duration, auspice, staffing). In all cases, sampling techniques are employed to reduce a larger universe of children and classrooms to a smaller number of representative cases, as well as to protect children and teachers against such inappropriate uses of research data as individualized evaluations used for high-stakes purposes.

Each of the 11 featured studies found that ECE program participation is associated with improvements in children's level of cognitive skills. Several found that children from economically disadvantaged families reaped comparatively larger cognitive gains than their more affluent peers. Two studies also found significant positive relationships between classroom quality and cognitive outcomes. However, the estimated size of these associations or effects varied significantly, both for the typical child and specific subgroups. Further, the fact that different studies employed significantly different research methods means that in some cases research findings are not comparable.

Several studies found that ECE program participation continues to be associated with small cognitive gains through the early elementary school years, and that this relationship is larger and longer-lasting for children from economically disadvantaged homes. In addition, the CPC evaluation, which is the only study of a large-scale ECE program that tracks children from preschool to adulthood, found that the "cognitive advantage" produced by program participation interacted with the effects of parent involvement, the quality of the post-preschool learning environment, and avoidance of school mobility to produce long-term effects including increased rates of high school graduation and decreased rates of grade retention, special educational placement, and juvenile arrests.

Findings on the short- and long-term relationships between ECE program participation and children's social-emotional outcomes varied significantly. Some studies found small-to-moderate positive associations or effects; other found negative relationships of a similar magnitude. Outcomes tended to vary significantly among subgroups and by number of hours spent in center-based care. Although some consistent

findings were reported across studies, others seem to contradict one another. Overall, it appears that the relationship between ECE program participation and children's social-emotional development is significantly more complex, varied, and potentially troubling than is true with regard to their cognitive development.

The eleven studies presented in this report illustrate the complexities involved in assessing the relationship between ECE program participation and children's development. Such research requires a multi-dimensional process designed to examine the relationship among program characteristics and quality (including teacher-child interactions), on the one hand, and child and family characteristics and child outcomes, on the other. Although it is tempting to simplify this process in an attempt to save time and money, to do so would be at best, wasteful and at worst, potentially harmful. Particularly as the numbers of children enrolled in ECE programs continues to grow at an unprecedented rate, it is critical to keep processes of program assessment firmly tethered to the complex realities of young children's lives.

## **Evaluating early care and education programs: A review of research methods and findings**

### **Introduction**

This report presents a detailed review of the most notable recent or ongoing research studies that examine the relationship between early care and education (ECE) program participation and child development. These research projects include both ECE program evaluations and other longitudinal studies that assess how participation in these programs affects children's lives. Although not strictly program evaluations per se, these studies are included because of their high quality and prominence in the early childhood field.

This literature review is designed to answer the following questions:

- (1) What is the basic purpose and structure of ECE program evaluation? How can it best inform issues of ECE program accountability?
- (2) What is the current state of research on primary ECE program types, including Head Start, child care, and state-funded pre-k?
- (3) What are the most important recent or ongoing research projects that examine the relationship between ECE program participation and child development? What is the basic structure of their individual research designs? What are their primary research findings with regard to preschool-age children?
- (4) What lessons do these studies offer to policy makers and early childhood professionals concerned with ECE program evaluation and accountability?

It should be noted that this report is not intended to provide a comprehensive survey of the research literature that examines the relationship between ECE program participation and child development. Rather than providing a broad and more general overview of what would be a huge and diverse array of studies, it focuses on eleven particularly important recent or ongoing projects. This narrower, but more in-depth examination of leading research in the field allows for a detailed comparison of research methods and findings across a diverse set of high-quality projects. It does not, however, represent the full array of research methods and findings in the field.<sup>3</sup>

Section 1 provides a brief introduction to ECE program evaluation, including its purpose, primary components, and relationship to program accountability issues. Section 2 provides a brief introduction to the state of research on primary ECE program types,

---

<sup>3</sup> For more general overviews of the relevant research literature, see Barnett, W. Steve. (1998). Long-term outcomes of early childhood programs on cognitive and school outcomes. *The Future of Children* 5, no. 3. Available: [www.futureofchildren.org/usr\\_doc/vol5no3ART2.pdf](http://www.futureofchildren.org/usr_doc/vol5no3ART2.pdf); Karoly, Lynn A., Greenwood, Peter W., Everingham, Susan S., Hoube, Jill, Kilburn, M. Rebecca, Rydell, C. Peter, Sanders, Matthew R., and Chiesa, James R. (1998). *Investing in Our Children: What We Know and Don't Know About the Costs and Benefits of Early Childhood Interventions*. Washington, DC: Rand Corporation. Retrieved April 26, 2005 from <http://www.rand.org/publications/MR/MR898/>; and Shonkoff, Jack P. and Phillips, Deborah A., eds. (2000), *From neurons to neighborhoods: The science of early childhood development* (Washington, DC: National Academy Press), Chap. 11.

including Head Start, child care, and state-funded pre-k. Section 3 presents detailed information on ten particularly important experimental, quasi-experimental, and non-experimental studies. Section 4 presents a synthetic overview of the primary research questions and methods represented by this literature. Section 5 provides a comparative synthesis of primary research findings, focusing in particular on short- and long-term cognitive and social-emotional outcomes and related issues of classroom quality. The concluding section discusses the implications of this research for the early childhood field and for issues of ECE program evaluation and accountability in particular.

### **(1) Program evaluation and program accountability**

The contemporary political and cultural climate has unfortunately caused policy-relevant discussions of program evaluation and accountability to become increasingly divorced from one another. In fact, the primary purposes of ECE program evaluation and accountability are properly understood to be the same: that is, to assess whether a given program is achieving its stated goals.<sup>4</sup> (A program accountability assessment might also be comparatively more concerned with whether a program is operating in a fiscally responsible, cost-effective manner.) Collecting and analyzing the data necessary to determine whether a given ECE program is accomplishing its objectives is a necessarily complex enterprise, involving careful analysis of the program at issue, multiple assessments of children's development, assessment of program quality, and information on relevant child, family, teacher, and program characteristics.

Any high-quality ECE program is by definition committed to supporting the healthy development of children in a multi-dimensional way, including their cognitive, social-emotional, and physical development. However, different programs have different primary commitments. Some, for example, have a pronounced pre-academic emphasis, while others focus more intensively on children's basic needs or family support services. At the same time, programs serve a wide variety of children and families with a wide array of preferences and needs. Different programs also have different sets of resources and constraints that affect what they do and how they are able to do it. Consequently, any meaningful assessment of how well a program is functioning should consider its stated objectives, who it is serving, how it is structured, and how these different components function together both in theory and practice.

Accomplishing this task requires adhering to the basic standards of ECE program evaluation research. Specifically, it is necessary to examine the following:

- (a) Whether a program is being implemented as intended,
- (b) The level and range of quality in typical classroom operations,
- (c) Variations in program structure that may affect child outcomes,
- (d) Differences in the characteristics of children and families that may affect how they experience the program and what they gain from it, and

---

<sup>4</sup> Meisels, Samuel J. (March 2006). Accountability in early childhood: No easy answers. *Erikson Institute Occasional Paper #6*, 7-16. Available: <http://www.erikson.edu/research.asp?file=publications8series#2006> (June 2006).

- (e) The extent to which children are achieving desired outcomes.

In short, it is necessary to examine how different program characteristics – as manifest in typical day-to-day operations – promote selected outcomes for different types of children. Only this type of multi-faceted assessment is capable of providing administrators and policy makers with meaningful information that can be used to improve program quality and child outcomes, and effectively serve the needs of young children and families.

Such an in-depth examination of program functioning stands in stark contrast to the one-dimensional approach to educational accountability exemplified by the *No Child Left Behind Act*, which evaluates schools on the basis of children's aggregate standardized test scores. Bracketing the question of its advisability for older students, early childhood experts agree that such high-stakes testing is inappropriate for children who are less than eight years of age.<sup>5</sup> This is true for numerous reasons. These include the widely varying and rapidly shifting rate of typical development among young children, the difficulty that they experience in trying to sit still and follow directions, and the low predictive value of standardized tests that is typical for children of this age.<sup>6</sup>

Despite these issues, it remains true that standardized, normed instruments may play a useful role in ECE program evaluation if they are properly employed. Correct usage entails:

- 1) Administering tests to only a sample of children,
- 2) Using an instrument of proven reliability and validity that is sensitive to individual differences within the age range(s) targeted,
- 3) Administering tests in an appropriate environment by well-trained examiners,
- 4) Considering test scores in combination with other appropriately used child assessment methods, including observational assessment,
- 5) Analyzing the relationship between test scores and differences among program types, classroom experiences, individual children, and their families and communities, and
- 6) Using test data exclusively for program evaluation and/or other forms of program assessment (as contrasted to evaluating individual children, teachers, or classrooms).<sup>7</sup>

<sup>5</sup> Bowman, B. T., Donovan, M. S., & Burns, M. S. (Eds.) (2001). *Eager to learn: Educating our preschoolers*. Washington, DC: National Academies Press.

<sup>6</sup> Meisels, Accountability in early childhood, 8-11. For an alternative perspective, see Zill, Nicholas. (Feb. 2006). Appropriate, reliable, valid: The use of direct child assessment in the evaluation of preschool programs. Paper prepared for the National Early Childhood Accountability Task Force, Pew Charitable Trusts (typescript).

<sup>7</sup> National Association for the Education of Young Children and the National Association of Early Childhood Specialists in State Departments of Education (1990), *Guidelines for appropriate curriculum content and assessment in programs serving children ages 3 through 8* (adopted November 1990), retrieved February 23, 2001, from [ericps.crc.uiuc.edu/naecs/position/currcont.html](http://ericps.crc.uiuc.edu/naecs/position/currcont.html) (pp. 22-23 of 31); Epstein, A. S., Schweinhart, L. J., DeBruin-Parecki, A., & Robin, K. B. (July 2004), Preschool assessment: A guide to developing a balanced approach, *NIEER Preschool Policy Matters* issue 7, 9; Meisels, S. J. & Atkins-Burnett, S. (2004), Evaluating early childhood assessments: A differential analysis, in K. McCartney & D. Phillips, eds., (2006), *Handbook of early childhood development* (Oxford: Blackwell), 533-549.

In short, both ECE program evaluation and accountability assessment should adhere to the same basic set of procedures that have been determined to be developmentally appropriate and scientifically credible by relevant research in the early childhood field. Determining whether an ECE program is achieving its stated goals requires a multi-dimensional process designed to examine the relationship among program characteristics and quality, on the one hand, and child and family characteristics and child outcomes, on the other. Although standardized child assessment instruments may play a critical role in this process, they cannot in and of themselves provide the information necessary to achieve a meaningful understanding of a given ECE program or its impact on the young children who participate in it. Reducing program evaluation or accountability to the isolated and one-dimensional metric of aggregate test scores may adversely affect the children who participate in these programs by subjecting them to inappropriately designed research methods and the consequences of inherently misleading research findings.

## **(2) Overview of ECE program research**

Given the relationship between ECE program evaluation and accountability, it is particularly helpful to consider the state of program evaluation research in the context of the various large-scale, widely implemented programs that are currently estimated to enroll over five million three- to five-year-old children annually.<sup>8</sup> The literature on such programs is most usefully simplified to the four primary categories of research on center-based Head Start, child care, state funded pre-k, and other. This section provides a brief introduction to the evolution of research on these primary program types in order to contextualize the detailed project descriptions that follow.

It should be noted that this review does not include analyses of such model programs as the Abecedarian Project or the High/Scope Perry Preschool Project. Such studies have traditionally dominated both the ECE program evaluation field and early childhood policy discourse more broadly.<sup>9</sup> They are, however, relatively uninformative with regard to large-scale, widely available programs, which are quite different from a typical model program.<sup>10</sup>

---

<sup>8</sup> Figure extrapolated from Iruka, I.U. and Carver, P.R. (2006). *Initial results from the 2005 NHES Early Childhood Program Participation Survey* (NCES 2006-075). U.S. Dept. of Education. Washington, DC: National Center for Education Statistics, Table 1, p. 7.

<sup>9</sup> Reynolds's (2000) synthesis of existing reviews of the ECE program evaluation literature concludes that "the vast majority of empirical evidence" on program effects "comes from model demonstration programs rather than large-scale, government-initiated programs like Head Start." Reynolds, Arthur J. (2000). *Success in early intervention: The Chicago Child-Parent Centers*. Lincoln, NE: University of Nebraska Press, 2000, 9.

<sup>10</sup> Compared to large-scale programs, model programs are typically of higher quality, more expensive to operate, employ larger and better trained staff, and are rarely implemented in inner-city communities. With regard to research, it is significantly easier for model programs to use random assignment to create treatment and non-treatment groups. Reynolds, *Success in early intervention*, 11; Barnett, Long-term effects of early childhood programs, 28.

**Head Start.** Established in 1965, Head Start is the oldest, and remains the largest publicly-funded ECE program in the nation. As such, there have been hundreds, if not thousands of studies conducted of it, including many program evaluations.<sup>11</sup> In the late 1990s, however, significant controversy erupted regarding how to best understand this body of research. An important 1997 report produced by the U.S. General Accounting Office (GAO) argued that only 22 studies were sufficiently rigorous to meet their standards,<sup>12</sup> and that even these were insufficient to determine the impact of the national program on the children and families that it serves. The U.S. Department of Health and Human Services (HHS) strongly disagreed with these findings, arguing that over 70 studies collectively provided “clear evidence of the positive impact of Head Start services.”<sup>13</sup> However, the 1998 Head Start reauthorization directed HHS to appoint an expert panel to offer recommendations for a study or studies to provide a national analysis of the impact of the Head Start program.<sup>14</sup> This led to development of the Head Start Impact Study, which is currently ongoing and represents one of the most extensive evaluations of the national Head Start program ever undertaken. A second major Head Start study, the Family and Child Experiences Survey (FACES), is also in progress. (Detailed information on both of these studies is provided in Section 3, below.)

**Child Care.** Research on child care programs dates back to the 1970s. Initially, this literature centered on issues of maternal attachment and social development, particularly as concerned the issue of whether child care is harmful to children. Beginning in the 1980s, this focus broadened to include the more complex question of how child care interacts with family factors to impact child development. In the mid-1980s, this research agenda expanded still further to focus on the relationship between child care quality, family characteristics, and child outcomes. At the same time, this literature’s traditional emphasis on social-emotional development was expanded to include issues of cognitive development and later school achievement. Conversely, studies of Head Start and other preschool programs began to incorporate more comprehensive assessments of classroom quality and devote greater attention to children’s social-emotional development.<sup>15</sup> These developments have significantly narrowed the gap between the research methods of and variables studied by the preschool literature, which was traditionally more academically

---

<sup>11</sup> For a review of the research literature on Head Start, see Love, John M. et al. (2006). *Head Start: What do we know about its effectiveness? What do we need to know?* Forthcoming in McCartney, Kathleen & Phillips, Deborah, eds. (2006). *Handbook of early childhood development*. Blackwell.

<sup>12</sup> Criteria for inclusion in the literature review were: (1) Head Start participation had occurred in 1976 or later, in order to provide a reasonably comparable experience with the program as it existed in the late 1990s; (2) outcomes of Head Start participants were compared to those of children who did not attend preschool, attended a different preschool, or with test norms; and (3) tests of statistical significance were performed on these differences, except in cases in which outcomes were measured using normed instruments. U.S. General Accounting Office. (2000). *Head Start: Research provides little information on impact of current program*. (GAO/HEHS publication no. 00-83). Washington, DC: Author, Appendix I.

<sup>13</sup> *Ibid.*, Appendix IV.

<sup>14</sup> U.S. DHHS. (Oct. 1999). *Evaluating Head Start: A recommended framework for studying the impact of the Head Start program*. Washington, DC: Author. Available: <http://www.acf.hhs.gov/programs/hsb/research/hsreac/1999report/index.htm> (June 2006).

<sup>15</sup> Love, John M., Schochet, Peter Z., Meckstroth, Alicia L. (1996, May). *Are they in any real danger? What research does – and doesn’t – tell us about child care quality and children’s well-being*. Princeton, NJ: Mathematica Policy Research, Inc., 13-18.

and cognitively focused, and child care studies, which were more concerned with issues of program quality, maternal attachment, and social-emotional development.

Currently, there are three ongoing longitudinal studies that are in the process of providing the most comprehensive information available to date regarding the relationship between child care and children's development. These are (1) the National Institute of Child Health and Human Development (NICHD) Study of Early Child Care, (2) the National Center for Education Statistics (NCES) Early Childhood Longitudinal Study – Birth Cohort (ECLS-B), and the NCES Early Childhood Longitudinal Study – Kindergarten Cohort (ECLS-K). These studies are highlighted below, along with the Cost, Quality, and Outcomes Study, which is widely regarded as one of the best examinations of the relationship between child care quality and child outcomes to date.<sup>16</sup>

**State-Funded Pre-K.** In comparison with Head Start and child care, research on state-funded pre-kindergarten programs is relatively limited. Although Gilliam and Zigler's recent (2004) review of the pre-k evaluation literature identifies a total of 67 reports on 20 programs in 18 states, only 14 are deemed sufficiently rigorous to qualify as true program evaluations. The majority of these studies, however, contain serious methodological flaws. For example, only a few used psychometrically valid instruments to assess children, examined program quality and its association with child outcomes, or analyzed the significance of variable program structures.<sup>17</sup>

Two important studies are underway, however, that promise to contribute significantly to the relatively scanty knowledge base provided by the existing state-funded pre-kindergarten literature.<sup>18</sup> Both are being conducted by the National Center for Early Development and Learning (NCEDL), a national early childhood research project based at the Frank Porter Graham Child Development Institute at the University of North Carolina, Chapel Hill, and supported by the U.S. Department of Education's Institute for Educational Sciences. In addition, evaluations of particular state programs – such as the universal pre-k programs in Georgia and Oklahoma – have received substantial attention

---

<sup>16</sup> Love, Are they in any real danger?, 41. Other particularly notable child care studies include Howes, C., Smith, E., and Galinsky, E. (1996). *The Florida Child Care Quality Improvement Study: 1996 report*. New York: Families and Work Institute; Whitebook, Marcy, Carollee Howes, and Deborah Phillips. *Who cares? Child care teachers and the quality of care in America. Final report: National child care staffing study*. Berkeley, CA: Child Care Employee Project, 1989; and Ruopp, Richard R. et al: *Children at the center: Final results of the National Day Care Study*. Cambridge, MA: Abt Books, 1979.

<sup>17</sup> Gilliam, Walter S. and Zigler, Edward F. (2004). State efforts to evaluate the effects of pre-kindergarten: 1977 to 2003. Manuscript. Retrieved April 25, 2005 from <http://nieer.org/resources/research/StateEfforts.pdf>, 7-8, 16, 21-22, 38; Gilliam, Walter S. and Zigler, Edward F. (2001). A critical meta-analysis of all impact evaluations of state-funded preschool from 1977 to 1998: Implications for policy, service delivery, and program evaluation. *Early Childhood Research Quarterly* 15, 441-473.

<sup>18</sup> Another study currently in progress, the National Pre-Kindergarten Study (NPS), is not included in this review because it is a purely descriptive study of implementation and classroom practices and will not include any child outcome measures. However, it does represent a critical new source of information on these programs. See Gilliam, Walter S., and Marchesseault, Crista M. (March 2005). From capitols to classrooms, policies to practice: State-funded pre-kindergarten at the classroom level. New Haven, CT: Yale University Child Study Center.

in the field. Each of these pre-k studies is described below, in conjunction with the recently completed NIEER Five-State Pre-Kindergarten Study.

**Other.** The ongoing evaluation of the Chicago Child-Parent Centers (CPCs) of the mid-to-late 1980s being conducted by Arthur Reynolds and his colleagues is in a category of its own. This work is frequently cited in conjunction with the Abecedarian and Perry Preschool projects, as all three employ rigorous research designs, follow study group children into adulthood, and have reported relatively impressive outcomes. Unlike Abecedarian and Perry Preschool, however, the CPC is a large-scale, publicly funded program located in inner-city neighborhoods.<sup>19</sup> Reynolds's work on the CPCs is widely regarded as the most well-developed, theoretically and methodologically rigorous, longitudinal examination of a large-scale ECE program in existence, and is consequently included below.

### **(3) Notable research initiatives**

It is useful to categorize the featured studies according to their primary methodological structure. They may accordingly be divided into the three basic categories of research that employs (1) experimental, (2) quasi-experimental, or (3) non-experimental designs. The following subsections briefly describe each of these methods and provide detailed information on each of the eleven highlighted studies.

#### **(3.1) Experimental designs**

Experimental designs randomly assign individuals participating in a given study to either a treatment group or a control group. This method is commonly (if somewhat controversially) considered to represent the “gold standard” of evaluation research because it eliminates the problem of selection bias which occurs when there are systematic differences between those who are included in a study sample and those who are not.<sup>20</sup> This method is widely considered to represent the best means of determining the impact of a particular intervention (such as participation in an ECE program) because outcome data on participants may be compared with that of a statistically equivalent group of non-participants.

It should be noted, however, that experimental designs are extremely difficult to employ when evaluating large-scale social and/or educational interventions such as ECE programs. Researchers are necessarily – and appropriately – constrained by the need to respect the ethical and practical issues involved in attempting to assign subjects to treatment and control groups, as the resulting experience may have a significant impact on their lives. Negotiating these complex issues while setting up and tracking two distinct

---

<sup>19</sup> This program is still active. As of April 2005, there were 48 CPC classrooms in 15 Chicago locations serving 1,383 children. At a glance fact sheet (April 2005). Available: Chicago Public Schools Office of Early Childhood Education.

<sup>20</sup> For a recent critique of the common tendency to accord experimental designs this preeminent status, see Chatterji, Madhabi (2005), Evidence on ‘what works’: An argument for extended-term mixed-method (ETMM) evaluation designs. *Educational researcher* 34, no. 5, 14-24.

study groups in the context of a large-scale research project also entails a sizeable increase in research costs.<sup>21</sup>

### **(3.1) Project description: Experimental design**

#### THE HEAD START IMPACT STUDY

Summary: The Head Start Impact Study, a six-year impact evaluation designed to determine the effect of program participation on children's school readiness, is one of the few major ECE program evaluations to employ an experimental design.<sup>22</sup> The study population includes 4,667 three- and four-year-old children divided into treatment and control groups. Study data are collected using child assessments (both externally and teacher-administered), parent interviews, teacher/staff surveys and interviews, classroom observations, and administrative records. First-year findings include small-to-moderate positive impacts for both 3- and 4-year-old children on standardized assessments of pre-reading, pre-writing, and vocabulary.<sup>23</sup>

Duration: 2000 – 2006

Sponsoring agency: Administration for Children and Families, U.S. Department of Health and Human Services.

Contractors: Westat, Inc. (In collaboration with the Urban Institute, American Institutes for Research, and Decision Information Resources.)

#### Primary research questions:

- (1) What is the impact of the national Head Start program on participants' school readiness?
- (2) Under which conditions and for which children does the program works best?

---

<sup>21</sup> Some researchers believe that the experimental method has inherent limitations when used to evaluate ECE programs, as control group children may participate in alternative programs that are similar to those experienced by the treatment group. For example, in the Head Start Impact Study, children denied access to the Head Start program of their choice (in order to create a control group) were free to enroll in other ECE programs – including other Head Start programs – if their parents arranged it. The likely variability of the resultant control group experience, as well as its potential overlap with that of the treatment group, arguably muddies the waters when it comes to comparing outcomes between the two. Further, the experience of the control group in the larger ECE program universe is necessarily affected by the fact that the program that they were initially denied access to exists. The experience of children denied access to the Head Start program of their choice, in other words, is unlikely to represent the experience that children would have if Head Start did not exist at all.

<sup>22</sup> In recent years, the National Early Head Start Research and Evaluation Project represents the only other major ECE program evaluation to employ an experimental design. For a recent report on study findings, see Love, John M., et al. (2005). The effectiveness of Early Head Start for 3-year-old children and their parents: Lessons for policy and programs. *Developmental psychology* 41, no. 6, 885-901.

<sup>23</sup> The extent to which program characteristics including quality, teacher educational level, full- versus half-day programs, etc. produce variations in program impacts will be examined in future project reports. U.S. DHHS. (June 2005). *Head Start Impact Study: First year findings*. Washington, DC: Author, iii.

Notable characteristics: This study stands out in the field due to its use of an experimental design. Specifically, a sample of newly entering 3- and 4-year-old Head Start applicants was randomly selected either to be enrolled in the program (the treatment group) or to have their application declined (the control group). Children denied access to Head Start could be enrolled in other ECE programs – including other Head Start programs – if their parents arranged it. This method allows researchers to assess the impact of Head Start by comparing outcome data on participating children with a statistically equivalent set of children who did not take part in the program.<sup>24</sup>

Study sample: A sample of 4,667 children divided into treatment and control groups will be followed from the time that they are three or four years old through the conclusion of first grade.<sup>25</sup>

Data collection: As listed on Table 1, basic categories of data include:

- (1) *Child assessments.*<sup>26</sup> The assessment battery focuses primarily on tasks that relate to the acquisition of reading skills. It is drawn from 19 standardized instruments, takes 35-45 minutes per child, and is administered by trained assessors in the child's normal daily setting.<sup>27</sup> Assessments of children's social-emotional development are made using parent/primary caregiver and teacher/caregiver reports.
- (2) *Parent/Primary Caregiver Interviews.* These interviews collect information on (a) parental beliefs and attitudes toward their child's learning, and satisfaction with her ECE experience; (b) family household and demographic information, including parent-child relationships and characteristics of the child's home environment; (c) ratings of child behavior problems, social skills, and competencies; (d) perceptions of their child's accomplishments; (e) perceptions of their relationship with the child; and (f) child and family receipt of a variety of comprehensive services.
- (3) *Teacher and Staff Surveys and Interviews.* In center-based programs, the center director is interviewed, and the child's teacher is administered a survey. For children in family child care, a single provider interview is conducted. In both cases, data are collected on teacher/care provider demographic information; education, experience,

---

<sup>24</sup> For a discussion of the statistical methods used to determine equivalence, see U.S. DHHS. (Sept. 2003). *Building futures: The Head Start Impact Study interim report*. Washington, DC: Author, 15-18.

<sup>25</sup> Three- and four-years-olds were placed in separate groups, each with its own treatment and control. The former group will be studied through two years of Head Start participation; the latter for one year. Of the 4,667 total study children, 2,559 are in the 3-year-old group and 2,108 in the four-year-old group. A total of 383 randomly selected Head Start centers, administered by 84 different grantees/delegate agencies in 23 different states, are participating. *Ibid.*, 8, 17.

<sup>26</sup> For non-English speaking children, interviewers determine the appropriate language to use immediately before administering the assessment battery. Spanish-speaking children were assessed primarily in Spanish in the fall of their preschool year, along with two subtests in English. The following spring, they were assessed primarily in English, along with two Spanish language measures. For children who could not be assessed in English or Spanish, a bilingual interviewer or interpreter was used for the fall of their preschool year. The following spring, they were assessed in English. *Ibid.*, 25.

<sup>27</sup> The creation of test scores and scales is discussed in U.S. DHHS, *Head Start Impact Study: First year findings*, 4-7 – 4-8.

and training with young children; types of services provided; and literacy-enhancing activities. Teacher and care providers are also asked to rate study children on a variety of cognitive and social-emotional skills. Information on the program, such as child recruitment and enrollment, staffing, professional development, parent involvement, curriculum, child assessment, and kindergarten transition are also collected.

- (4) *Observed Quality of Care Settings.* Trained observers visit children’s primary non-parental care setting in the spring of their preschool year to directly assess quality. Three standardized instruments are used: the Early Childhood Environment Rating Scale, Revised (ECERS-R),<sup>28</sup> the Classroom Observation of Teacher-Directed Activities Checklist, and the Arnett Scale of Teacher/Provider Behavior. To provide consistency across different types of settings, a five-question observational instrument is additionally employed to measure overall safety, basic hygiene standards, availability of educational materials, and overall positive and negative interaction between provider and child.
- (5) *Administrative Records.* At the kindergarten and first grade levels, the study relies primarily on existing indicators of school quality culled from the U.S. Department of Education’s National Center for Education Statistics (NCES) datasets. These include such factors as student free and reduced-price lunch status, student race and ethnicity, average pupil/teacher ratios, teacher and staff characteristics, measures of average student achievement by subject area, and attendance and dropout statistics.<sup>29</sup>

Data collection schedule: Data collection began in Fall 2002 and continues through Spring 2006.<sup>30</sup> Comparable data are being collected for both Head Start and non-Head Start children.<sup>31</sup> Table 1 presents the data collection schedule for the study’s four-year-old cohort.

**Table 1: Head Start Impact Study: Data Collection Schedule (4-year-old cohort only)**

	Year 1 (preschool)		Year 2 (K)		Year 3 (Grade 1)	
	Fall	Spring	Fall	Spring	Fall	Spring
Child assessment	√	√		√		√
Parent/primary caregiver interview	√	√	√	√	√	√
Staff interviews		√		√		√
Observed quality of care settings		√				
Administrative records				√		√

Source: US DHHS, Building Futures, 20.

Available reports: First year findings of the Head Start Impact Study were released in June 2005. This report summarizes key findings for the three- and four-year-old cohorts in the areas of child cognitive development, social-emotional development, and health, as

<sup>28</sup> As appropriate, the Family Day Care Rating Scale (FDCRS) was used in place of the ECERS-R.

<sup>29</sup> U.S. DHHS, Building futures, 25-30.

<sup>30</sup> ACF recently announced a competition for selecting instruments to be used for a third grade follow-up study. While it is uncertain whether this research will be implemented, the possibility is under consideration.

<sup>31</sup> U.S. DHHS. *Head Start Impact Study: First year findings*, v.

well as parenting practices. It also analyzes whether there are differential effects due to children's race or ethnicity, home language, gender, or special needs status; as well as parental depression, parental marital status, and age of mother at first birth. Future reports will examine additional areas of possible impact, explore possible variation in impact by community and program characteristics (e.g., classroom quality, teacher education level, full-day versus part-day program), and follow children through the end of first grade.<sup>32</sup>

Primary research findings: Table 2 lists statistically significant effect sizes found for each of the 14 externally-administered cognitive assessment instruments used in the study.<sup>33</sup>

**Table 2. Head Start Impact Study: Effect sizes for externally-administered child cognitive assessments**

	Age 3	Age 4
<b><i>Language development &amp; literacy</i></b>		
WJ-R Letter Word Identification	0.24	0.22
WJ-R Dictation	–	–
WJ-R Oral Comprehension	–	–
WJ-III Spelling	–	0.16
Comprehensive Test of Phonological and Print Processing (CTOPPP) - Print Awareness subtest	–	–
CTOPP Word Elision subtest	–	–
Letter Naming Task	0.19	0.24
Peabody Picture Vocabulary Test - III	0.12	–
Color Naming	0.1	–
Story and Print Concepts	–	–
<b><i>Mathematics</i></b>		
WJ-R Applied Problems	–	–
Color Names and Counting	–	–
<b><i>Perceptual Motor</i></b>		
McCarthy Draw-a-Design	0.13	–

Source: U.S. DHHS, *Head Start Impact Study: First year findings*, Chap. 5.

For three-year-olds, small-to-moderate effects were found on four measures of language development and literacy, as well as on one measure of perceptual motor skills. For four-year-olds, similar effects were found on three measures of language development and literacy.<sup>34</sup> No statistically significant effects were found for early math assessments administered to either age group.<sup>35</sup>

<sup>32</sup> U.S. DHHS, *Head Start Impact Study: First year findings*; U.S. DHHS (n.d.), “Building futures: Head Start Impact Study Frequently Asked Questions,” downloaded July 19, 2005 from [http://www.acf.hhs.gov/programs/opre/hs/impact\\_study/](http://www.acf.hhs.gov/programs/opre/hs/impact_study/), 9-10.

<sup>33</sup> Sixty percent of three-year-olds and 59% of four-year-olds in the control group participated in some alternative type of ECE program (with 42% and 48% were in center-based programs, respectively). This means that the effect sizes listed in Table 2 represent the difference between being admitted to Head Start and a variety of alternative experiences, a significant proportion of which may have included something similar to Head Start. U.S. DHHS, *Head Start Impact Study: First year findings*, ix.

<sup>34</sup> These findings translate into a sizeable reduction in the gap between the mean scores of Head Start children and national norms on some measures. For example, three-year-old Head Start participants scored four points below the national norm on the W-J III Letter-Word Identification subtest in the spring of their

Preliminary analyses of subgroup differences found fewer cognitive effects for Spanish-speaking three-year-olds and none for Spanish-speaking four-year-olds. Although particularly positive impacts were found in several domains for three-year-old African American and Hispanic children, fewer were found for four-year-old African Americans and none for Hispanics in this age group.<sup>36</sup>

Based on parent reports, access to Head Start had a small positive impact on three-year-old's problem behaviors, one of three social-emotional constructs measured. Specifically, total behavior problems and hyperactive behavior decreased by .13 and .18 respectively. No statistically significant impacts were found for four-year-olds on these measures.<sup>37</sup>

### **(3.2) Quasi-experimental designs**

Quasi-experimental designs are similar to experimental designs in that they utilize treatment and comparison groups. They are different, however, in that while experimental designs randomly assign individuals to each group, quasi-experimental designs do not. Instead, researchers create a comparison group by identifying a set of individuals who are willing to participate in the study and who share relevant characteristics with those in the treatment group. Although generally considered to be less scientifically rigorous than experimental designs, quasi-experimental designs provide an important means of comparing outcomes between treatment and non-treatment groups in cases where practical and/or ethical considerations prevent individuals from being randomly assigned to receive or not receive services.

Quasi-experimental designs may create comparison groups in a variety of ways. For example, a sample of children enrolled in a given ECE program might be compared to a matched sample of children who remained on the waiting list for the program and shared important characteristics with the study sample (e.g., family socio-economic status).<sup>38</sup> Another method, termed a regression-discontinuity design, was employed in the Oklahoma Universal Pre-K and NIEER Five-State Pre-K studies discussed below. Using a specialized set of statistical techniques, this method compares outcomes of children who completed a year of preschool with those of children who are just entering the program.<sup>39</sup> In all cases, the purpose of the quasi-experimental method is to provide

---

preschool year. However, children in the comparable treatment group scored 7.6 points below this norm. Alternately stated, access to Head Start reduced the mean score gap between qualifying Head Start children and the national average by 47%. (The equivalent gap reduction for four-year-old was 45%.) Ibid., xi.

<sup>35</sup> Effect sizes of < 0.2 are defined as small, 0.2 – 0.5 as moderate, and > 0.5 as large. Ibid., ii, fn 1.

<sup>36</sup> Ibid., xviii.

<sup>37</sup> Ibid., xiv. Teacher reports on children's social-emotional development will be analyzed in future reports.

<sup>38</sup> The purpose of drawing from a program waiting list is to minimize selection bias by having the comparison group include only those children whose parents went through the procedures necessary to attempt to enroll them in the program – just like the parents of those in the treatment group. Children whose parents did not attempt to enroll them in the program are more likely to have different individual characteristics that will significantly affect outcomes, thereby producing selection bias.

<sup>39</sup> This technique uses statistical methods to compare cohorts of children whose birthdays fall within the specified range for preschool admittance for two successive years. Using this technique requires that ECE

researchers with a means of determining the effect of ECE program participation on a representative sample of children.

### **(3.2a) Project descriptions: Quasi-experimental designs**

#### CHICAGO CHILD-PARENT CENTERS

Summary: The longitudinal evaluation of the Chicago Child-Parent Centers (CPCs) has been ongoing for almost two decades. The study employs a quasi-experimental design, which compares children who participated in the CPCs in the mid-1980s with a statistically equivalent group that graduated from CPS kindergartens in 1986 but were not involved in the CPC program.<sup>40</sup> The primary purpose of the study is to determine the short- and long-term effects of CPC participation on child/youth and family outcomes.<sup>41</sup> Of the original sample of over 1,500 children, more than 1,100 remained in the study by the time they reached early adolescence. All of these children were non-white and attended CPS kindergartens located in high-poverty neighborhoods. Among this study's many important findings is that CPC program participation was significantly associated with improved reading and math achievement scores and decreased rates of cumulative grade retention and special education placement for study group children at ages 14 – 15.

Duration: Data collection began in 1986 and is ongoing.

Principal investigator: Arthur J. Reynolds, Institute of Child Development, University of Minnesota.

#### Primary research questions:

- 1) Controlling for child gender and risk status, did CPC participation improve child/youth and family outcomes?
- 2) Did participation through second or third grade improve youth and family outcomes relative to participation solely in preschool through first grade? (Additional control used for kindergarten achievement.)

---

programs have a strict cut-off date with regard to age eligibility. For applied discussions of the regression-discontinuity method, see Gormley, William T., Jr., et al. (2005). The effects of universal pre-k on cognitive development. *Developmental psychology* 41, no. 6, 876-878; and Barnett et al, The effects of state pre-kindergarten programs on young children's school readiness in five states, 7-9.

<sup>40</sup> The CPC program is a center-based early intervention that provides comprehensive education and family support services to economically disadvantaged children and their parents from preschool to early elementary school (ages 3 – 9). Established in 1967 with Title I funds, it is the second oldest federally funded ECE program in the U.S. (after Head Start), and the oldest extended early intervention program. To be eligible, children must reside in a school neighborhood that is receiving Title I funds and must not be enrolled in another preschool program. Parents or guardians must agree to participate in the program at least one-half day a week (although many in fact fall short of this requirement). The program conducts extensive community outreach in an effort to recruit the children most in need of program services who are residing in the neighborhoods served. Screening interviews are also conducted, and children determined to be the most high-need are given admittance priority. Reynolds, *Success in early intervention*, 23-24.

<sup>41</sup> The most recent published analyses track study participants through age 22. See, for example, Ou, S. and Reynolds, A. J. (2006). Early childhood intervention and educational attainment: Age 22 findings from the Chicago Longitudinal Study. *Journal of education for students placed at risk* 11, no. 2, 175-198.

- 3) Did some groups of children benefit from program participation more than others? (Sub-groups were defined by gender, educational status of parent/guardian, risk status, school poverty during preschool and kindergarten, parent participation in school, and instructional emphasis in preschool.)
- 4) Which individual, family, and school-related factors affect participants' long-term outcomes? Or, are program effects most powerfully explained by processes of enhanced cognitive development, family support, social adjustment, individual motivation, or school support?<sup>42</sup>

Notable characteristics: Reynolds argues that it is preferable to use an aggregate risk index, rather than separate variables (such as parents' educational attainment) to assess whether program effects vary for children who are more or less at-risk for school failure. Cumulative or multiple risks have been found to be substantially associated with developmental functioning. Consequently, Reynolds holds, a risk index is a better measure of meaningful subgroup differences than are separate indicator variables. Specifically, Reynolds constructs a risk index from eight dichotomously coded variables: 1) parent/guardian does not have a high school degree, 2) child is eligible for a free lunch subsidy, 3) family has more than four children, 4) child attends a school in which 60% or more of the students are low-income, 5) parent-guardian is not employed, 6) child lives in a single-parent family, 7) missing data on family background (education or lunch; missing data coded as risk),<sup>43</sup> and 8) minority status. (There was no variability on this indicator as 100% of the study group is either black or Hispanic.)<sup>44</sup>

Study sample: Data for Reynolds's work is derived from the Chicago Longitudinal Study (CLS), a prospective longitudinal study that investigated the academic development of 1,539 children who graduated from CPS kindergartens in 1986. This study group included all of the 1,150 children enrolled in the 20 extant CPCs with preschool and kindergarten programs who began preschool in fall 1983 and graduated from kindergarten in Spring 1986. In the overall study group, some children began the CPC program in preschool, others in kindergarten, and others in the early primary grades. Consequently, Reynolds is able to examine the relationship between levels of participation in the program and later outcomes. Of the original set of 1,539 children, over 1,100 remained in the primary study sample by the time they reached early adolescence. Ninety-three percent of these study participants are African-American and the remaining 7% are Hispanic.<sup>45</sup>

The CLS additionally created a comparison group of 389 children who had similarly graduated from CPS kindergartens in 1986, but had no involvement with the CPC program at any time. In order to maximize comparability, children in this group attended schools participating in the Chicago Effective Schools Project, which, like the CPCs, was designed to serve high-risk children in high-poverty neighborhoods. In addition, the

---

<sup>42</sup> Reynolds, *Success in early intervention*, 65.

<sup>43</sup> Missing data were coded as a risk variable because research indicated that children with missing data were more likely to be economically disadvantaged. *Ibid.*, 228 n. 2.

<sup>44</sup> *Ibid.*, 71-72.

<sup>45</sup> *Ibid.*, 51-52, 64, 71.

possibility of selection bias due to the quasi-experimental design of the study was extensively investigated by project researchers. Their conclusion was that selection bias was small and did not affect estimates of program impact.<sup>46</sup>

Data collection: Because the CLS was designed in the mid-1980s, its child and program assessment instruments do not conform to the most highly regarded practices in the field today. (For example, kindergarten achievement was measured using group-administered subtests of the Iowa Test of Basic Skills, which is not a method or an instrument recommended by child assessment experts.)<sup>47</sup> Consequently, this report will not detail the full set of measures utilized in this research.

Data collection schedule: The data collection schedule of this longitudinal study, which has been ongoing for almost two decades, is variable and complex. For an overview, see Reynolds (2000), Chapter 3.<sup>48</sup>

Reports available: A book written by principal investigator Arthur Reynolds, *Success in early intervention: The Chicago Child-Parent Centers*, was published in 2000. Reynolds, as well as his collaborators and students, have also produced (and continue to produce) numerous reports on multiple issues pertaining to the effects of CPC program participation. For a list of project publications, see <http://www.waisman.wisc.edu/cls/PUBLICATION.HTM>.

Primary research findings: By age 15, youth who had participated in the CPC program had a five-month gain in both reading and math achievement over the comparison group. By age 18, 14% of CPC participants had received special education services, as contrasted to 25% of the comparison group. At the same age, 23% of CPC participants repeated a grade, compared to 38% of the comparison group. By age 22, 65% of program participants completed high school, as compared with 54% of the comparison group. Participation was also associated with a 37% reduction in juvenile arrests, with 16.4% of the treatment and 25.9% of the comparison group experiencing one or more arrests.<sup>49</sup>

It is important to note that these findings represent the average effect of CPC participation regardless of the length of time individual children were involved in the program, which could range from 1 – 7 years (ages 3 – 9, or preschool – 3<sup>rd</sup> grade).<sup>50</sup> However, 4 – 6 years of participation yielded significantly higher reading and math

---

<sup>46</sup> Ibid., 52, 62.

<sup>47</sup> Ibid., 73.

<sup>48</sup> Ibid.

<sup>49</sup> *Newsletter of the Chicago Longitudinal Study* 1 (Aug. 2000) & 2 (June 2002); <http://www.waisman.wisc.edu/cls/index.html>.

<sup>50</sup> The effects of CPC participation did not vary significantly by gender, risk status, level of parent education, or parent participation in school. The strongest evidence for differential effects was that children who attended programs in relatively high-poverty neighborhoods (i.e., over 60% low-income) benefited more than children living in lower poverty neighborhoods (ibid., 171, 179). It is important to note, however, that the study employed a relatively homogenous sample, as the CPC program enrolled what were determined to be the most high-risk children living in the low-income neighborhoods where the program operated.

achievement, and lower rates of grade retention, special education placement, and juvenile arrests than less extensive or no participation.<sup>51</sup> Overall, duration of program participation significantly associated with all indicators of educational and social competence across all methods of data analysis. Consequently, Reynolds (2000) emphasizes that “the dosage-response relationship was the most robust finding of the study.”<sup>52</sup>

The CPC study identified three sets of factors that best explain why participating children experienced long-term program effects. These include (1) the cognitive boost provided by the preschool experience; (2) the quality of the post-preschool learning environment and avoidance of school mobility; and (3) parent involvement in school. Notably, early cognitive development did not directly predict the major effects of CPC participation. Rather, it was part of a larger process that interacted with later school and familial experiences to place children on a path toward improved educational attainment and pro-social outcomes.<sup>53</sup>

#### THE OKLAHOMA UNIVERSAL PRE-K PROGRAM

Summary: Two closely related studies of the state pre-k program in Tulsa, Oklahoma were conducted by Georgetown University researchers during the 2001-02 and 2002-03 academic years. The primary difference between the two projects is that while the former used a locally-developed instrument to assess child outcomes, the latter used subtests of the nationally-normed Woodcock-Johnson Achievement Test. Both studies received a relatively high degree of recognition in the ECE program evaluation field due to their use of the regression-discontinuity method. The 2002-03 study found moderate-to-large positive effects on the Letter-Word Identification, Spelling, and Applied Problems subtests for the full study sample, as well as for subgroups of white and black, and full-price and free lunch-eligible children.

Duration: An initial study examined the 2001 – 2002 academic year; a second study the 2002 -2003 academic year.

Principal investigators: William T. Gormley, Jr., Ted Gayer, Deborah Phillips, and Brittany Dawson, Georgetown University, Washington, D.C.

Primary research questions (2002-03 study):

- (1) Does participation in Tulsa’s state pre-k program improve children’s school readiness as assessed by subtests of the Woodcock-Johnson Achievement Test?
- (2) Do child outcomes vary by children’s race/ethnicity or family income?

<sup>51</sup> *Newsletter of the Chicago Longitudinal Study* 1 (Aug. 2000), 7; Reynolds, *Success in early intervention*, 172. Even after 5 – 6 years of participation, however, the school achievement of CPS students remained below the national average for 14-15 year olds. Reynolds, *Success in early intervention*, 182.

<sup>52</sup> Reynolds, *Success in early intervention*, 172-173.

<sup>53</sup> Reynolds, Arthur J., Ou, Suh-Ruu, and Topitzes, James W. (2004). Path of effects of early childhood intervention on educational attainment and delinquency: A confirmatory analysis of the Chicago Child-Parent Centers. *Child development* 75, no. 5, 1299-1328.

- (3) Do child outcomes vary according to program duration (full- or half-day) and children's racial/ethnic status?

Notable characteristics: Both the 2001-02 and the 2002-03 studies were widely noted in the field due to their use of the regression-discontinuity method. This method was possible to use in this case because (1) the Oklahoma pre-k program employs a strict birthday eligibility criterion for entering children and (2) the Tulsa school district administers the same assessment instrument to pre-k and kindergarten children at the same time at the beginning of the school year. The regression-discontinuity method assumes that entering kindergarten children who just made the cut-off date for pre-k the previous year will be statistically similar to children entering pre-k who just missed the previous year's cut-off date, except for the fact that the former group will have completed a year of pre-k whereas the latter will not yet have done so.<sup>54</sup> Consequently, samples of entering pre-k and kindergarten children are considered to represent treatment and comparison groups, which are used to gauge the effects of pre-k participation on child outcomes.

Study sample (2002-03 study): 1,567 pre-k students (85% of the total enrolled in the Tulsa pre-k program) and 1,461 entering kindergarteners (84.5% of the total).

Data collection (2002-03 study): The study analyzes three key sets of data:

- (1) Subtests of the Woodcock-Johnson Achievement Test (letter-word identification, spelling, and applied problems),
- (2) Family and child characteristics, including full- or reduced-price lunch status, race/ethnicity, gender, and maternal education, and
- (3) Program duration (full- or half-day).

Data collection schedule: Child assessments were administered to pre-k and kindergarten children at the beginning of the school year. The data collection schedule for program and child and family characteristics is not specified, but presumably occurred at approximately the same time.

Reports available: The most recent report on the 2002-03 study was published in *Developmental psychology* in 2005.<sup>55</sup> Previous reports produced by Georgetown University's Center for Research on Children in the U.S. (CROCUS) are available on the center website at <http://www.crocus.georgetown.edu/papers.html>.

Primary research findings: Table 3 presents child outcome findings for the three subtests of the Woodcock-Johnson Achievement Test–III employed in the study: Letter-Word Identification, Spelling, and Applied Problems. These findings include statistically

---

<sup>54</sup> Although the study abstract states that it is comparing “‘young’ kindergarten children who just completed pre-k to ‘old’ pre-k children just beginning pre-k,” most of the analyses were based on data that included children born within a full year of the relevant cut-off date. Precisely how the regression-discontinuity method may be adjusted to account for this large age disparity is unclear. The report is also silent on the question of how the experience of summer break may have affected outcomes for entering kindergartners. Gormley et al, The effects of universal pre-k on cognitive development, 872, 878.

<sup>55</sup> Gormley et al, The effects of universal pre-k on cognitive development.

significant regression coefficients, standard errors, and effect sizes for the study sample, as well as for selected subgroups with  $N > 870$ . As indicated below, moderate-to-large statistically significant effects were found for the full study sample, as well as for subgroups of white and black, and full-price and free lunch eligible children.<sup>56</sup>

**Table 3. Oklahoma Pre-K Study: Statistically significant regression coefficients and standard errors for Woodcock-Johnson subtests ( $N > 870$ )**

	Letter-word ID	Spelling	Applied problems
All: Regression coefficient & SE	2.999 (.501)	1.857 (.324)	1.939 (.506)
Effect (SD for control group)	0.79	0.64	0.38
White: RC & SE	3.022 (.886)	2.067 (.516)	–
Effect	0.76	0.72	–
Black: RC & SE	2.911 (.810)	1.469 (.545)	1.682* (.759)
Effect	0.74	0.52	0.38
Full-price lunch: RC & SE	2.687 (.927)	1.590 (.532)	1.543** (.887)
Effect	0.63	0.54	0.29
Free lunch: RC & SE	2.791 (.659)	1.750 (.446)	2.097 (.681)
Effect	0.81	0.65	0.45

Notes:  $p < .01$  for all cells except as indicated by \* ( $p < .05$ ) or \*\* ( $p < .10$ ).

Source: Gormley et al, The effects of universal pre-k on cognitive development, 880-881.

As Gormley et al. point out, these effect sizes exceed those reported for other state-funded pre-k programs, which range from .23 - .53; for pre-k programs generally, which range from .10 - .13; and for high-quality child care programs, which seldom exceed .10. Notably, the study's reported effect sizes for the Letter-Word Identification subtest, which range from 0.63 – 0.81, are on a par with reported effects of the Abecedarian project (.73 and .79 for four- and five-year-olds, respectively) and the Perry Preschool program (0.60).<sup>57</sup>

#### NIEER FIVE STATE PRE-KINDERGARTEN STUDY

**Summary:** This study used a regression-discontinuity design to estimate the effects of state-funded pre-kindergarten programs on the academic skills of entering kindergartners in Michigan, New Jersey, Oklahoma, South Carolina, and West Virginia. Receptive vocabulary, early literacy, and math skills were assessed in a sample of 5,071 children. Pre-k programs were found to have a statistically significant and meaningful impact on children's early language, literacy and math development, with some evidence of an

<sup>56</sup> By extrapolating projected test scores of the comparison group, Gormley et al estimated that pre-k participation increased children's Letter-Word scores by 52.95%, Spelling by 26.42%, and Applied Problems by 17.94%. Alternatively, they claim that pre-k participation produced an average gain of 7-8 months in Letter-Word Recognition, 6-7 months in Spelling, and 4 months in Applied Problems. Ibid., 878, 882.

<sup>57</sup> Ibid., 880-881.

enhanced program effect for print awareness skills with children from low-income families.<sup>58</sup>

Duration: 2004 – 2005

Sponsoring agency: National Institute for Early Education Research (NIEER), in partnership with state government and public universities in five selected states.

Principal investigators: W. Steven Barnett, Cynthia Lamy, and Kwanghee Jung.

Primary research question: The primary research question addressed in this study is whether participation in state-funded preschool programs positively affects children's receptive vocabulary, early literacy, and math skills.

Notable characteristics: The primary report produced for this study notes that it found effect sizes that were at least 2 – 3 times larger for the PPVT and W-J III Applied Problems subtest than the ones identified by the Head Start Impact Study. It is important to note, however, that this comparison is made without controlling for child and family characteristics.<sup>59</sup> This represents a critical omission, as it is well known that Head Start typically serves a population that is significantly lower-income and higher-risk than those of most state-funded pre-kindergarten programs. Further, it is questionable whether the effect sizes produced by the experimental design of the Head Start Impact study and the regression-discontinuity method used in this case should be considered comparable.<sup>60</sup> This is particularly true given that the Five State Pre-K Study, similar to the previously discussed Oklahoma pre-k evaluation, appears to have constructed treatment and comparison groups using children born within a full year of one another.<sup>61</sup>

Data collection: The research design was structured to include the same number of pre-k and kindergarten classrooms for each school district included in the study. An initial random sample of 1,937 classrooms (approximately half pre-k and half kindergarten) in five states was reduced to 1,320 due to difficulties in accessing the full sample.

---

<sup>58</sup> Barnett, W. Steven, Lamy, Cynthia, and Jung, Kwanghee. (2005). The effects of state pre-kindergarten programs on young children's school readiness in five states. Rutgers, NJ: National Institute for Early Education Research, Rutgers University, [nieer.org/resources/research/multistate/fullreport.pdf](http://nieer.org/resources/research/multistate/fullreport.pdf), 2.

<sup>59</sup> Data were collected on the following child characteristics: gender, race, free lunch status, and whether the home language is English. However, data on free lunch status could not be obtained for almost 20% of the sample. Using these data, researchers found family income to have a small effect on child outcomes in two states. Barnett et al, The effects of state pre-kindergarten programs, 20. However, it is questionable that sufficient data were collected to allow comparison to Head Start children, given the discrepancy in applicable family income eligibility guidelines (under the federal poverty line for Head Start; up to 130% of it for free lunch status), as well as the relatively high number of child risk factors that are generally associated with the Head Start population.

<sup>60</sup> The Georgia Early Childhood Study, however, found that children from working poor or very low-income families had stronger cognitive outcomes if they attended state pre-k, as opposed to Head Start or private preschool. Henry, Gary T. et al. (2005). The Georgia Early Childhood Study, 2001 – 2004: Final report. Atlanta, GA: Andrew Young School of Policy Studies, Georgia State University, vii – viii, 80.

<sup>61</sup> Barnett et al, The effects of state pre-kindergarten programs, 5.

Approximately four children were randomly sampled per classroom for a total of 5,278 (2,728 pre-k and 2,550 kindergarten). This sample was racially and ethnically diverse at 47% white, 15% African American, 21% Hispanic, 3% Native American, and 2% Asian. Study group children were individually assessed in either English or Spanish based on teacher recommendations. Domains assessed included receptive vocabulary, early math skills, phonological awareness, and print awareness.<sup>62</sup>

Data collection schedule: Child assessments were conducted in the fall of the 2004 – 2005 school year.

Reports available: The final five-state report, as well as reports on each of the five states individually, are available on the NIEER website, <http://nieer.org/docs/?DocID=129>.

Primary research findings: The largest pre-k effect was found for print awareness, which increased an average of 16.64% across the five states. Findings were statistically significant in all states, although they exhibited significant variation. The average estimated effect size was 0.64. This represents a 39% increase in mean scores and 85% more growth over the course of the school year.<sup>63</sup>

Vocabulary scores of children in the treatment group increased 3.96 raw score points across the five states. Findings were statistically significant in each state and did not exhibit much variation. The average estimated effect size was 0.26. This represents an 8% increase in mean scores and 31% more growth over the course of the school year than projected for the comparison group. Put in different terms, participation in state pre-k typically translated into an additional three months worth of progress on this measure.<sup>64</sup>

Math skills, as measured by W-J III Applied Problems subtest, increased by 1.41 raw score points, which represents 35% of the relevant standard deviation. The average estimated effect size across the five states was 0.28. Pre-k participation did not have significant effects on children's phonological awareness in any of the five states.<sup>65</sup>

### **(3.3) Non-experimental designs**

In contrast to both the experimental and quasi-experimental methods, non-experimental designs do not employ a control or comparison group. Consequently, they are not able to determine the effect of a given treatment on study subjects with the same degree of precision as an experimental or quasi-experimental design. However, non-experimental designs are well-equipped to determine the statistical significance of important relationships, such as the association between ECE program participation and child outcomes. They may also use statistical methods to control for important factors

---

<sup>62</sup> Instruments used included the Peabody Picture Vocabulary Test, 3<sup>rd</sup> edition; Test de Vocabulario en Imagenes Peabody; Woodcock Johnson Tests of Achievement, 3<sup>rd</sup> edition, Applied Problems subtest; Bateria Woodcock-Munoz Prueba 25, Problemas Aplicados; Preschool Comprehensive Test of Phonological and Print Processing, Blending and Print Awareness subtests. Ibid., 6-7.

<sup>63</sup> Ibid., 11, 13.

<sup>64</sup> Ibid., 9, 13.

<sup>65</sup> Ibid., 10, 12, 13.

that may affect outcomes, such as child risk status or program quality. This technique allows researchers to determine, for example, whether outcomes differ systematically for children enrolled in ECE programs with significantly varying levels of classroom quality.

Non-experimental studies may feature contemporaneous, pre-post, or longitudinal designs.<sup>66</sup> Contemporaneous designs examine the relationship between selected factors, such as child care quality and child outcomes, at a given point in time. Pre-post designs examine whether selected outcomes experienced a significant change before and after a particular intervention; for example, letter recognition at the beginning and end of the preschool year. Longitudinal designs examine the long-term relationship between particular variables, such as participation in a preschool program and educational attainment. In all cases, well-designed non-experimental studies additionally control for key factors that could have a significant independent effect on the outcomes under study.

### **(3.3a) Project descriptions: Non-experimental designs**

#### THE FAMILY AND CHILD EXPERIENCES SURVEY (FACES)

Summary: FACES is a non-experimental study designed to provide information on program quality, child outcomes, and school readiness that can be used for national decision-making and program quality improvement.<sup>67</sup> Launched in 1997, the study includes three nationally representative cohorts of children with an average sample size of over 2,900 each. Data are collected using child assessments (both direct and teacher-administered), parent interviews, teacher and staff interviews, and classroom observations. Although further data collection and analysis are ongoing, particularly notable findings to date include the differential levels of cognitive and social-emotional gains experienced by children whose baseline scores were in the highest and lowest quartiles of their cohort at program entry. By the spring of the program year, children in the lowest quartile posted substantial gains on early literary, math, and social skills, as well as reductions in problem behaviors. The cognitive outcomes of those in the highest quartile, however, declined slightly in comparison to national norms; scores on social skills and problem behavior measures were also somewhat lower for these children.

Duration: 1996 – 2010

Sponsoring agency: Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

---

<sup>66</sup> This discussion is based on Love et al (1996), which additionally discusses recommended statistical techniques for these different types of non-experimental studies. Love et al, Are they in any real danger?, 37-45.

<sup>67</sup> Zill, Nicholas et al. (May 2003). Head Start FACES 2000: A whole-child perspective on program performance. Washington, DC: U.S. DHHS, 1. FACES is as a key part of the Head Start Program Performance Measures Initiative, which established a new system of outcome-based accountability for the national program.

Contractors: (1) Westat, Inc.; Xtria, LLC; and the CDM Group, Inc. (through 2005); (2) Mathematica Policy Research, Inc., Educational Testing Service, and Research Support Services (2005-2010).

Primary research questions:

- (1) What is the overall quality of Head Start classrooms? Are particular program or teacher characteristics associated with quality?
- (2) What skills and knowledge do children have when they enter Head Start? Do they make significant gains in Head Start and during kindergarten?
- (3) Do children show significant gains in social skills and declines in problem behavior during the Head Start year?
- (4) Are particular program or teacher characteristics associated with improved child outcomes?

Notable characteristics. The primary difference between the Head Start Impact Study and FACES is that FACES uses a non-experimental design, which tracks the progress of the three cohorts of newly entering three- and four-year-old Head Start children through kindergarten (and, for the first two cohorts, first grade) using pre- and post-tests. Because of this difference, FACES is classified as an *outcome evaluation*, which does not have the capacity to identify causal relationships between program participation and child outcomes. Although FACES can demonstrate the extent to which child outcomes changed during the course of the study, it cannot disentangle the effects of program participation, child development, and family and community influences with the same level of precision as an experimentally-designed impact study.

Study sample: FACES is designed to be a nationally representative study and to date includes three cohorts of study children: (1) the initial sample of 3,200 children in 40 Head Start programs established in 1997, (2) a sample of 2,800 children in 43 programs launched in 2000, and (3) a sample of 2,800 children in 60 programs established in 2003.<sup>68</sup>

Data collection. Four primary sets of data are collected: direct child assessments, parent/primary caregiver interviews, teacher and staff interviews, and classroom observations. While the primary types of data being collected are in many ways similar to the Head Start Impact Study, there are several important differences. First, FACES collects more extensive data on children's social-emotional (S-E) development. In contrast to the Head Start Impact Study, which relies exclusively on teacher and parent reports for S-E data, FACES additionally uses two instruments that are administered by trained outside observers: the Assessment Behavior Scale and the Peer Play Observation Scale.<sup>69</sup> Second, unlike the Head Start Impact Study, FACES does not use school administrative data to assess the quality of kindergarten and first grade classrooms. Instead, kindergarten and first grade teachers are administered questionnaires. In

<sup>68</sup> [www.acf.hhs.gov/programs/opre/hs/faces/faces\\_overview.html](http://www.acf.hhs.gov/programs/opre/hs/faces/faces_overview.html) (July 2005).

<sup>69</sup> Both instruments were developed for this study, as were several other cognitive measures used. For further information on child and program assessment instruments employed, see [www.acf.hhs.gov/programs/opre/hs/faces/instruments](http://www.acf.hhs.gov/programs/opre/hs/faces/instruments) and Zill et al, *Head Start FACES 2000*, 139-149.

addition, for the first two cohorts of study children, direct classroom observations were conducted in the spring of their first grade year.<sup>70</sup>

Data collection schedule: In general, key data are collected in the fall and spring of the preschool year, as well as in the spring of kindergarten and first grade. Data collection schedules for FACES and the Head Start Impact Study have two primary differences. First, FACES interviews teachers and observes classrooms in both the fall and spring of the study children's preschool year, whereas the Head Start Impact Study does this exclusively in the spring. Second, FACES interviews parents in the spring of children's kindergarten and first grade years, while the Head Start Impact Study does this in both the fall and the spring.

Available reports: The full set of available reports is available at <http://www.acf.hhs.gov/programs/opre/hs/faces/index.html>. The most recent major report, the *Head Start Performance Measures Center: Family and Child Experiences Survey (FACES 2000) Technical Report*, was issued in February 2006. This report presents findings on (1) cognitive gains of Head Start children and their achievement in kindergarten, (2) the relationship between cognitive gains and program, classroom, and child characteristics, (3) changes in social skills and problem behavior in Head Start, (4) the relationship between social behavior and program, classroom, and child characteristics, (5) the social-emotional and cognitive development of children with disabilities, (6) the level of quality in Head Start classrooms and factors that help explain variations in quality, (7) the relationship between curricula and family, program, and classroom characteristics, (8) longitudinal changes in family structure and other household characteristics of Head Start children, (9) family risk factors, parental involvement, and Head Start's protective role, and (10) the predictive validity of the FACES 2000 battery.<sup>71</sup>

Primary research findings: The notable findings to date are the differential cognitive and social-emotional outcomes experienced by children whose baseline scores were in the highest and lowest quartiles of their cohort at program entry. Although most children enter Head Start with early literacy and math skills well below national averages, there is considerable diversity in skill levels. At program entry, the highest-scoring quartile in the study sample was at or above the national average (50<sup>th</sup> percentile) in early language and

---

<sup>70</sup> The "data collection" section of the U.S. HHS Administration for Children and Families (ACF) FACES website states that no first-grade follow-up is planned for the 2003 cohort, but does not explain why. There is also a discrepancy between the 2003 FACES report, which states that the 2000 cohort will be followed through kindergarten and this webpage, which states that it will be followed through first grade. Similarly, a 2001 FACES document reports that all children will be followed through first grade. See Tarullo, Louisa B., and McKey, Ruth Hubbell (April 2001), Design and implications of the Head Start Family and Child Experiences Survey (FACES), downloaded June 2005 from <http://www.acf.hhs.gov/programs/opre/hs/faces/index.html>.

<sup>71</sup> Zill, Nicholas et al. (Feb. 2006). *Head Start Performance Measures Center: Face and Child Experiences Survey (FACES 2000) Technical Report*, 12-13. Downloaded 5/9/06 from [www.acf.hhs.gov/programs/opre/hs/faces/index.html](http://www.acf.hhs.gov/programs/opre/hs/faces/index.html).

number skills. The lowest quartile, however, ranked in the lowest 2% of all preschoolers in these areas nationwide.<sup>72</sup>

On average, children made gains toward national averages during the Head Start year, especially with regard to vocabulary knowledge and early writing skills. Those entering with lower levels of knowledge and skill showed the largest gains. However, they remained considerably behind national averages at the end of the program year. While those entering with higher levels of skills posted higher scores in the spring than the fall, their standing in terms of national norms declined slightly for seven of the nine measures used.

For example, while the vocabulary scores of the overall study sample experienced a mean gain of 3.8 points on the PPVT-III, scores of the bottom quartile rose 8.4 points (more than half a standard deviation), while those of the top quartile declined .5 points relative to national norms. Overall, the cognitive skill gains experienced by Head Start children from Fall to Spring are considered to be “relatively modest,” although “educationally meaningful.”<sup>73</sup>

After one year of Head Start, teachers reported significant improvements in three of four social behavior scales (cooperative classroom behavior, hyperactive behavior, and withdrawn behavior). Parents, however, did not report significant gains until children had experienced two years of the program. Children who demonstrated the lowest levels of social skills and the highest levels of problem behaviors at program entry showed the greatest gains, including significant improvements in all reported social behaviors.

Although the gap between these children and those in the highest quartile narrowed significantly while in Head Start, it remained substantial at the time that they graduated from the program. Further, the scores of children who entered Head Start with high social skills and minimal problem behavior declined slightly during the course of the program. For example, the mean “cooperative behavior” score for the top quartile went from 20.47 at program entry to 20.00 at graduation. The scores of those in the bottom quartile, however, rose from 8.72 to 15.40.<sup>74</sup>

---

#### THE COST, QUALITY, AND OUTCOMES STUDY

**Summary:** The Cost, Quality, and Outcomes Study was a five-year project that examined the relationship between cost and quality in full-time, center-based ECE programs, as well as how these factors affect children’s development from preschool through second grade. The study sample included 826 preschoolers enrolled in 151 child care centers in four states. Data collected included child assessments (direct and teacher-administered), parent and teacher reports, and classroom observations. Although the Cost, Quality, and Outcomes Study is widely regarded as an exceptionally high-quality research project, the

---

<sup>72</sup> Ibid., E-2.

<sup>73</sup> Ibid., E-2, 2-7 – 2-8.

<sup>74</sup> Ibid., 3-6

fact that its study sample was largely white and middle-income limits its applicability to predominantly minority, non-English speaking, and/or high-poverty communities. Findings indicate that preschool quality (considered independently of K – 2 classroom quality) had a modest positive association with children’s receptive language, early math skills, cognitive/attention skills, sociability, and problem behaviors through kindergarten; and with math skills and problem behaviors through second grade.

Duration: 1993-97

Principal investigators: Carolee Howes, Richard Clifford, Ellen S. Peisner-Feinberg, Mary L. Culkin, and Sharon Lynn Kagan.

Primary research questions:

- (1) Does ECE program quality remain associated with child outcomes through the early elementary years?
- (2) Were these relationships different for children at greater risk?
- (3) To what extent did children’s ECE program experiences versus their early elementary school environment predict developmental outcomes?
- (4) To what extent did the peer climate in children’s ECE classrooms predict their peer relationships in second grade?<sup>75</sup>

Notable characteristics: Researchers initially conducted a comprehensive examination of the costs and quality of center-based ECE programs in four states: California, Colorado, Connecticut, and North Carolina. In the first phase of the study, detailed information about operating costs and structural and dynamic classroom quality was collected from 401 randomly selected ECE centers, about half of which were for-profit and half nonprofit.

Study sample: In the second, longitudinal phase of the study, 826 preschoolers in their next-to-last year of child care were recruited from 183 of the classes in 151 of the previously studied centers. Only children whose primary home language was English were included in this sample.<sup>76</sup> Similar to the NICHD Study of Early Child Care, the demographic profile of the children involved in the Cost, Quality, and Outcomes Study was significantly different from that of children enrolled in Head Start. In addition to including only English speaking children, the study sample from preschool through second grade was 70-78% white. The median family income of the study children ranged from \$47,753 to \$52,381 annually – far above the 2005 Head Start Family Income Guidelines of \$16,090 for a family of four.<sup>77</sup> Further, mothers of children in the study had typically attended at least two years of college. Researchers did, however, conduct

---

<sup>75</sup> Brooks-Gunn et al, *Early Child Development in the 21<sup>st</sup> Century*, 204; Peisner-Feinberg, E.S., et al. (2000). The children of the cost, quality, and outcomes study go to school: Technical report. (Chapel Hill: University of North Carolina at Chapel Hill, Frank Porter Graham Child Development Center), 3.

<sup>76</sup> Peisner-Feinberg et al., The children of the cost, quality, and outcomes study go to school: Technical report, 2.

<sup>77</sup> Head Start family income guidelines for 2005. Typescript in possession of author received from Karen Carradine, Manager, Head Start Program, Chicago Public Schools, Spring 2005.

differential analyses of children within their sample who had characteristics indicating that they might be at-risk of school failure.<sup>78</sup>

Data collection: Data were collected from five different sources: (1) individual child assessments, (2) teacher ratings of children, (3) parent reports of child and family characteristics, (4) teacher reports of beliefs and practices, and (5) classroom observations.<sup>79</sup>

**Table 4. Cost, Quality, and Outcomes Study: Data Collection Schedule (1993-1997)**

	Preschool		K	1st G	2nd G
	Fall	Spring	Spring	Spring	Spring
<b>Child</b>					
Indv. Assessment - Cognitive Skills	√	√	√		√
Indv. Assessment - S-E Skills	√	√	√		√
Teacher Ratings - S-Behv. Skills	√	√	√		√
Parent Reports - Child Care & Child Health	√	√	√	√	√
School Records (absences, refs, placemts)			√		√
<b>Parent-Family Environment</b>					
Child & Family Demographic Information	√	√	√	√	√
Parental Beliefs/Practices	√	√	√	√	√
<b>Family - School Links</b>					
Parent Involvement (parent & teacher reports)	√	√	√	√	√
Transition/Selection		√	√	√	√
<b>Classroom</b>					
Observed Practices Quality	√	√	√		√
Teacher Reported Relationship Quality	√	√	√		√
Teacher Demographics	√	√	√	√	√
Teacher Beliefs			√	√	√
Structural Child Care/School Characteristics	√	√	√	√	√
Financial	√	√			

Source: Peiser-Feinberg et al., *The children of the cost, quality, and outcomes study go to school: Technical report*, 55-56

Data collection schedule. Table 4 details the data collection schedule for the 5-year study, which was conducted during 1993-97.<sup>80</sup> As in the case of the Head Start Impact Study and FACES, the basic pattern was to collect data (1) during both the fall and spring of children's preschool year and (2) in the spring of the kindergarten and early elementary school years, with a more complete set of data being collected in the final year of the study.

Available reports: The study's major report, *The children of the cost, quality, and outcomes study go to school*, is available at <http://www.fpg.unc.edu/~ncedl/pages/cq.cfm>.

<sup>78</sup> Ibid., 6, xiii; Peiser-Feinberg, E. S. et al (2001). The relation of preschool child-care quality to children's cognitive and social developmental trajectories through second grade. *Child Development* 72, 1534-1553, 1550.

<sup>79</sup> Ibid., 4.

<sup>80</sup> Ibid., 55-56.

Papers on teachers' background, parents as child care consumers, the relationship between structural features of child care and dynamic quality, and the relationship between child care experience and child development have also been published.<sup>81</sup>

Primary research findings: This study presented two sets of findings. The first examined the influence of preschool quality on children's cognitive and social-behavioral skills through second grade, without controlling for the quality of their K – 2 classroom experiences. The second set of findings examined the same set of relationships, but additionally controlled for later school quality.

Considered independently of K – 2 classroom quality, preschool quality had a modest but significant positive association with children's receptive language, early math skills, cognitive/attention skills, problem behaviors, and sociability through kindergarten; and on their math skills, cognitive/attention skills, and problem behaviors through second grade. Letter-word recognition was the only skill measured that had no significant relationship to preschool quality. In general, positive associations were stronger for children whose mothers had relatively lower levels of education.<sup>82</sup>

After controlling for the quality of children's K – 2 classroom experiences, the relationship between preschool quality and children's math skills remained significant through second grade. In addition, children in preschool classrooms with higher levels of peer interaction during play had significantly better relationships with their peers in second grade.<sup>83</sup> Children of mothers with fewer years of education also continued to show a significant reduction in problem behaviors. In contrast, children in preschool classrooms with high levels of problem behaviors and low levels of teacher-child closeness<sup>84</sup> demonstrated higher levels of aggression and disruptive behavior in second grade.<sup>85</sup>

#### NICHD STUDY OF EARLY CHILD CARE

Summary: The primary purpose of the NICHD Study of Early Child Care is to examine how variations in non-maternal care are related to children's social-emotional adjustment, cognitive and linguistic development, school achievement, and physical growth and

<sup>81</sup> A citation list is provided in Brooks-Gunn et al, *Early child development in the 21<sup>st</sup> century*, 212-213.

<sup>82</sup> Maternal education was measured on a continuum from 10 – 20 years. Peisner-Feinberg et al., *The children of the cost, quality, and outcomes study go to school: Technical report*, 27 – 31.

<sup>83</sup> The 2<sup>nd</sup> grade peer relationship measure was adjusted for individual behavior before considering the effect of child care quality in order to control as much as possible for individual child characteristics. *Ibid.*, 34.

<sup>84</sup> Teacher-child closeness was measured using the Student-Teacher Relationship Scale (STRS). Items on this measure were summed into three factors representing different aspects of the teacher-child relationship (as assessed by teacher self-reports): closeness, conflict, and overdependency. *Ibid.*, 5.

<sup>85</sup> *Ibid.*, 32-34.

health.<sup>86</sup> An initial study sample of 1,364 children born in 1991 has been tracked through their seventh year in school. (The final phase of the study will follow over 1,000 members of the original sample through age 15.) Extensive data have been collected, including direct observation of children's home, child care, and school environments; and multiple measures of child development, family dynamics, caregiver and teacher characteristics, and classroom quality. The study's most striking findings to date with regard to preschool-aged children are that higher levels of cumulative time spent in child care are modestly associated with negative social behaviors and that higher levels of classroom quality are modestly associated with improved pre-academic and language skills.

Duration: The full study, which extends the analysis of early child care to include youth development, was initiated in 1989 and consists of four phases. Phase I tracked the study sample of 1,364 children from birth (in 1991) to age 3. Phase II encompasses the preschool portion of the study, following the 1,226 children who continued to participate in the research from age 3 through their second year in school (1995-2000). Phase III followed the 1,100 children remaining in the study sample through their seventh year in school (2000-2005). Phase IV, which is currently in progress, will follow over 1,000 of the original children through age 15.<sup>87</sup>

Sponsoring agency: National Institute of Child Health and Development (NICHD).

Principal investigators: The NICHD Early Child Care Research Network is a collaborative team of 30 researchers. A full list of members is available at <http://secc.rti.org/investigators.cfm#researchers>.

Primary research questions: Research questions most pertinent to the concerns of the Pew Early Childhood Accountability Project include:

- (1) Is ECE program quality associated with the psychological or health development of children?
- (2) Are past experiences in child care predictive of later psychological or health outcomes?
- (3) Is the relationship between child care and children's development different for children who are members of racial/ethnic minorities and/or live in economically disadvantaged households?
- (4) Is the average number of hours that children spend in child care associated with their psychological development or physical health?
- (5) What is the relationship between the aspects of care that are possible to regulate and the quality of care that children receive?<sup>88</sup>

---

<sup>86</sup> Non-maternal care includes both center-based care and relative or non-relative care in the child's home or in someone else's home.

<sup>87</sup> NICHD Study of Early Child Care and Youth Development website, "Study Summary" <http://secc.rti.org/summary.cfm> (July 21, 2005).

<sup>88</sup> Adapted from NICHD abstract available on the National Center for Children in Poverty's Research Forum website, [http://www.researchforum.org/project\\_general\\_185.html](http://www.researchforum.org/project_general_185.html).

Notable characteristics: The NICHD represents the most sophisticated examination of the relationship between child care and child development ever conducted. However, its applicability to high poverty, non-English speaking, and/or predominantly minority communities is reduced.

Study sample: The initial sample of 1,364 children was large enough to permit a reasonably precise estimation of effect sizes. The study children were selected from 10 sites distributed across major regions of the country and included urban, suburban, and rural areas. Children were selected based on their mother's employment plans at the time of their birth in order to establish a sample in which 60% of mothers would be working full-time (more than 30 hours per week), 20% part-time (10-30 hours per week), and 20% staying at home (not working outside of the home for more than 10 hours per week). Ethnic minority, single-parent, and low-education families constituted at least 10 percent of the sample at every site. However, mothers under age 18 were excluded from sample, as were non-English speaking mothers, and those who lived in neighborhoods deemed by police to be unsafe for visitation.<sup>89</sup> This method produced a sample that was 76% white, 13% black, and 6% Hispanic. Consequently, the applicability of the study to non-English speaking and/or predominantly minority communities is somewhat limited.

Data collection: Phases I and II of the study, which tracked children from birth through first grade, included extensive direct observation of their home, child care, and school experiences; and multiple measures of social-emotional development, cognition, language, achievement, and physical growth and health. A total of 75 different assessment instruments, including standardized tests, interview schedules, questionnaires, and other rating scales for use with children, families, caregivers, teachers, and classrooms were used during Phase II, the preschool portion of the study.<sup>90</sup>

Data collection schedule: Face-to-face child assessments were conducted at 1, 6, 15, 24, 36, and 54 months, and in first grade. In addition, families were contacted via telephone every three months during Phase I (birth to 36 months) and every four months during Phase II (37 months to the end of first grade).

Available reports: Data from Phases I-III have been analyzed by the NICHD research team, which has produced nearly 120 scientific publications on the project, a high proportion of which have appeared in peer-reviewed professional journals.<sup>91</sup> Detailed abstracts of almost half of these publications can be found on the project website, <http://secc.rti.org/>. A recent book edited by the NICHD Early Child Care Research Network, *Child care and child development: Results from the NICHD study of early child*

---

<sup>89</sup> Brooks-Gunn, Jeanne, Fuligni, Allison Sidle, and Berlin, Lisa J. (2003). *Early Child Development in the 21<sup>st</sup> Century: Profiles of Current Research Initiatives*. (New York: Teachers College Press), 187-188; NICHD Study of Early Child Care and Youth Development website, "Study Summary" <http://secc.rti.org/summary.cfm> (July 2005).

<sup>90</sup> Brooks-Gunn et al, *Early Child Development in the 21<sup>st</sup> Century*, 187-191. A full list of all 75 instruments, along with detailed descriptions of each, is available on the NICHD project website, <http://secc.rti.org/>.

<sup>91</sup> NICHD Early Child Care Research Network, ed. (2005). *Child care and child development: Results from the NICHD study of early child care and youth development*. (New York: Guilford), xi.

*care and youth development* (New York: Guilford, 2005) provides a useful synthesis of this research.

Primary research findings: To date, most published findings are from Phase I of the study, which covered children ages 0 – 36 months. These findings “make it clear that child care experiences cannot be adequately assessed without reference to children’s experiences in their families.” Although the type and quality of early child care have a clear impact on children’s cognitive and social-emotional outcomes, family factors are consistently better predictors of these outcomes than early child care experiences alone. In addition, the effects of early child care are highly mediated by children’s home environment.<sup>92</sup>

With regard to preschool-age children in particular, the study’s most striking findings to date concern the cumulative effect of time spent in ECE programs, and the relationship between program quality and cognitive outcomes.<sup>93</sup>

By the time study group children reached 4 –5 years of age, researchers found a modest, but statistically significant relationship between the cumulative time spent in child care and negative social behaviors. Specifically, “the more time children spend in any variety of non-maternal care arrangements across the first 4½ years of life, the more externalizing problems and conflict with adults they manifest at 54 months of age and in kindergarten, as reported by mothers, caregivers, and/or teachers.” These effects persist after controlling for the quality, type, and instability of child care,<sup>94</sup> as well as for maternal sensitivity and other family background factors.<sup>95</sup>

Although child care quality did not override the modest negative behavioral effects of accumulation of a high number of hours in child care, quality was found to have a significant effect on children’s pre-academic and language skills. Children whose child care was in the highest tercile of quality obtained higher scores on relevant assessments than children whose child care was in the bottom tercile. After controlling for the study’s full set of child and family covariates, the adjusted mean scores for children in higher quality care were 2.2 points higher on pre-academic skills and 2.3 points higher on language skills, with effects sizes of .24 and .15 respectively. Using a more limited set of covariates increased the mean score differences to 3.8 points for pre-academic skills and 4.8 points for language skills, with effect sizes of .39 and .29 respectively.<sup>96</sup>

---

<sup>92</sup> Ibid., 36.

<sup>93</sup> Ibid., 432-433.

<sup>94</sup> Other studies have produced different findings. Love et al (2003), for example, reviewed three different studies that examined the relationship between early child care and child outcomes using child samples with significantly different characteristics than those of the NICHD. (Specifically, these studies included children from more high-risk, high-poverty families, as well as classrooms with more variation in terms of low to high quality.) These studies suggest that quality may be an important moderator of the effects of time spent in child care, at least for children from higher-risk families. Love, John M., et al. (2003). Child care quality matters: How conclusions may vary with context. *Child development* 74, no. 4, 1021-1033.

<sup>95</sup> NICHD Early Child Care Research Network, ed. (2005). *Child care and child development*, 315.

<sup>96</sup> In comparison, the adjusted mean scores for children assessed to be receiving high-quality parenting were 8.3 points higher on pre-academic skills and 13.7 points higher on language skills than for children in the low-quality parenting group, with effect sizes of .88 and .87, respectively. Ibid., 389.

By the end of third grade, having participated in higher quality child care continued to have a small positive effect on cognitive outcomes as measured by standardized tests of math, reading achievement, and memory. (Effect sizes ranged from 0.07 – 0.09.)<sup>97</sup> However, spending over 30 hours per week in child care also had a small association with poorer work habits and lower social skills at this time (effect sizes ranging from 0.01 – 0.12). The relationships among amount of care and externalizing behaviors and teacher-child conflict decreased during the primary grades and were not significant by Grade 3.<sup>98</sup>

#### EARLY CHILDHOOD LONGITUDINAL STUDY – BIRTH COHORT (ECLS-B)

Summary: The ECLS-B is a longitudinal study of a nationally representative sample of 14,000 children who were born in 2001 and will be tracked through first grade. The study collects data regarding children’s homes, communities, health care, and non-parental care arrangements, including participation in ECE programs. Once the children are in school, it will collect data on their schools, classrooms, and teachers. Assessments of children’s physical, social-emotional, cognitive, and language development are being conducted throughout the course of the project.<sup>99</sup>

Duration: 2001 – 2007

Sponsoring agencies: National Center for Education Statistics (NCES), Institute of Education Sciences, U.S. Department of Education, in collaboration with seven institutes from the National Institutes of Health (NIH), U.S. Department of Health and Human Services.

Principal investigators: NCES is working collaboratively on this study with 10 federal and non-federal health, education, and human services agencies and organizations.<sup>100</sup>

Research questions: The study is designed to answer a wide variety of research questions. Those that are most pertinent to the relationship between ECE program participation and child development include:

- (1) How do different characteristics of ECE programs interact with child and family variables to affect children’s social and cognitive development?
- (2) What effect does participation in different types of ECE programs have on children’s development and achievement?
- (3) Do child outcomes vary by race and ethnicity, socio-economic status, and other characteristics of children and their families?

<sup>97</sup> Specific instruments used included four subtests of the Woodcock-Johnson Psycho-Educational Battery–Revised: letter-word identification, applied problems, memory for sentences, and picture vocabulary.

<sup>98</sup> NICHD Early Child Care Research Network. (2005). Early child care and children’s development in the primary grades: Follow-up results from the NICHD Study of Early Child Care. *American Educational Research Journal* 43, 537-570.

<sup>99</sup> <http://nces.ed.gov/ecls/Birth.asp> (June 2006).

<sup>100</sup> ECLS-B project summary, available at <http://nces.ed.gov/pubsearch/getpubcats.asp?sid=024>.

Notable characteristics: The ECLS-B is a companion study of the ECLS-K (Early Childhood Longitudinal Study – Kindergarten Cohort). (See ECLS-K project description, below.)

Study sample: The study is following a nationally representative cohort of 14,000 children born in 2001 from birth through first grade.<sup>101</sup> The parents of 10,688 of these children participated in the first wave of the study when these children were approximately nine months old.<sup>102</sup>

Data collection: Information on the preschool assessment battery, which was implemented in Fall 2005, is not yet publicly available.

Data collection schedule: Data collection began in 2001 when the study children were nine months old. The next major wave occurred when they were two years old. Additional waves of data collection are planned for when they will be approximately four years old (preschool), entering kindergarten, and in first grade.

Available reports: A full list of publications is available at <http://nces.ed.gov/pubsearch/getpubcats.asp?sid=024>.

Primary research findings: Given that ECLS-B preschool data are still in the process of being collected, no findings are yet available.

#### EARLY CHILDHOOD LONGITUDINAL STUDY – KINDERGARTEN COHORT (ECLS-K)

Summary: The ECLS-K is a longitudinal study of a nationally representative sample of approximately 22,000 children who attended kindergarten during 1998-99. The study sample is being tracked through 2011, when most participants will be completing twelfth grade. Data on children’s cognitive, social, emotional, and physical development are collected from study participants and their families, teachers, and schools. In addition, data on children’s home, school, and classroom environments, home educational practices, classroom curriculum, and teacher qualifications are obtained. The two primary purposes of the study are (1) to provide descriptive information on children’s status at school entry, transitions through critical grade levels, and progression through twelfth grade, and (2) to establish a data set that enables researchers to study how a wide range of family, school, community, and individual variables affect success in school.<sup>103</sup>

<sup>101</sup> <http://nces.ed.gov/ecls/Birth.asp> (June 2006).

<sup>102</sup> Bethel, James et al. (2005). *Early Childhood Longitudinal Study, Birth Cohort (ECLS-B), Sampling*. Vol. 2 of the *ECLS-B methodology report for the 9-month data collection, 2001-02* (NCDS 2005-147). U.S. Department of Education. Washington, DC: National Center for Education Statistics, 1. Available: <http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2005147> (June 2006). For demographic information on the nine-month study sample, see “ECLS-B 9-month sample and population sizes,” available at <http://nces.ed.gov/ecls/Birth.asp> (June 2006).

<sup>103</sup> ECLS-K project summary and study information, <http://nces.ed.gov/ecls/Kindergarten.asp> (June 2006).

Although the ECLS-K does not include detailed information on study children's preschool years, participating parents were questioned about their children's prior ECE program experiences. Researchers have used these data, in conjunction with kindergarten and later child outcome data, to estimate the effects of ECE program participation. These studies found that participation in non-Head Start center-based programs is associated with modest increases in pre-reading and math skills and small increases in negative behaviors.

Duration: 1998 - 2011

Sponsoring agencies: National Center for Education Statistics (NCES); Administration on Children, Youth, and Families (ACYF), U.S. Department of Health and Human Services (HHS); Economic Research Service (ERS), U.S. Department of Agriculture (USDA); Office of Special Education Programs (OSEP), Office of English Language Acquisition (OELA), and Policy and Programs Studies Services, U.S. Department of Education.<sup>104</sup>

Principal investigators: NCES developed the study research design and is collecting the core project data in conjunction with a number of investigators and subcontractors. Several of the above federal agencies contribute to instrument content, specialized forms of data collection, and sampling operations.<sup>105</sup>

Research questions: Primary questions addressed by the study include:

- (1) What is the developmental status of children at kindergarten entry? What are school expectations regarding entering children's skills, behaviors, and attributes? How well do children with different backgrounds and life experiences fare in the kindergarten environment?
- (2) How do child, family, and school factors interact to affect children's transitions from kindergarten to first grade, from elementary to middle school, and from middle to high school?
- (3) To what extent do schools and classrooms successfully address the needs of all children, including those with special needs?
- (4) When do children begin to experience problems with their school work? What are the circumstances surrounding those difficulties? How long do these problems last? How do children's families, schools, and teachers respond to them?
- (5) What roles do parents and families play in preparing for and supporting their children's education? How do families, schools, and communities interact to support children's education?<sup>106</sup>

Notable characteristics: Although the ECLS-K is a companion study of the ECLS-B, it was initiated several years before the latter project commenced.

<sup>104</sup> <http://nces.ed.gov/ecls/KinderSponsoringAgencies.asp> (June 2006).

<sup>105</sup> Ibid.

<sup>106</sup> ECLS-K project summary, <http://nces.ed.gov/ecls/Kindergarten.asp>, 2-3 (June 2006).

Study sample: The study sample includes approximately 22,000 children who were enrolled in close to 1,000 kindergarten programs during the 1998-99 school year. These children were selected from both public and private kindergartens, which offered both full- and part-day programs, as well as children with limited English proficiency or special needs. The sample includes children of different racial/ethnic and socio-economic backgrounds. Through its weighted design, it allows subgroup analyses of children who attended kindergarten in public and private schools; who are classifiable as black, white, Hispanic, or Asian; and whose families are of differing socio-economic status.<sup>107</sup>

Data collection: Basic categories of data are as follows:

- (1) *Child assessments.* Child assessments measure both cognitive outcomes (e.g., general knowledge, literary, and quantitative skills), social-emotional development (i.e., social skills, problem behaviors, approaches to learning), and physical and psychomotor development (e.g., general health, nutrition, and physical activity). All cognitive assessments are made using an untimed, one-on-one CAPI method.<sup>108</sup> Beginning in third grade, study children reported their own perceptions of their abilities and achievements, as well as their interest in and enjoyment of all school subjects.
- (2) *Parent/guardian reports.* Particularly during the first years of the study, parents or guardian reports represent a primary source of information on study group children. Data on children's development at school entry and their experiences with family members and others are collected from parents or guardians using CATI<sup>109</sup> (or CAPI if they do not have a phone).
- (3) *Teacher reports.* Self-administered teacher questionnaires are used to collect data on teachers' backgrounds, teaching practices, and experience, as well as the classroom setting for the sampled children that they teach. In addition, teachers evaluate each sample child on a number of cognitive and non-cognitive dimensions.
- (4) *Other school reports.* School administrators, principals, and headmasters complete self-administered questionnaires that provide information on the physical, organizational, and fiscal characteristics of their schools. Special attention is paid to the school's learning environment and programs, including its instructional philosophy and expectations for students.<sup>110</sup>

Data collection schedule: Data on the kindergarten cohort were collected in the beginning and near the end of the 1998-99 school year. These data included child assessment and parent and teacher reports in both the fall and the spring, as well as school reports in the spring only. The subsequent school year, when most study children moved into first grade, data were collected from a 30% subsample of the cohort in the fall and from the full study sample in the spring. These data included child assessments and parent reports in the fall and spring, as well as teacher and school reports in spring only. Additional

---

<sup>107</sup> Ibid., 5, 9.

<sup>108</sup> Computer Assisted Personal Interviewing (CAPI).

<sup>109</sup> Computer Assisted Telephone Interviewing (CATI).

<sup>110</sup> Ibid., 6, 8.

child, parent, teacher, and school data were (or will be) conducted in the spring of 2002 (third grade), 2004 (fifth grade), 2007 (eighth grade), 2009 (tenth grade), and 2011 (twelfth grade).<sup>111</sup>

Available reports: A full list of publications is available at <http://nces.ed.gov/pubsearch/getpubcats.asp?sid=024>.

Primary research findings: This analysis reports findings from three studies of ECLS-K data produced by Magnuson, Meyers, Ruhm, and Waldfogel (2004a, 2004b, 2005) as well as a report authored by Loeb et al. (2005).<sup>112</sup> In each case, researchers used parental reports of children's preschool experience to sort study participants into four primary groups, according to whether they (1) attended Head Start, (2) participated in a non-Head Start center-based program (Magnuson et al. (2004b) also subdivided this group into pre-k, preschool, or child care), (3) received other non-parental care, or (4) were primarily in parental care in the year prior to kindergarten entry. Although the resulting analyses are limited in that they do not include child outcomes from the preschool year or assessments of preschool quality (as the ECLS-K did not begin data collection until children were enrolled in kindergarten), they are important because they are based on the largest nationally representative sample of young children available to date.

Loeb et al. (2005) found that participating in a non-Head Start center-based program that operates for at least one-half day was associated with an approximately 0.10 standard deviation (SD) advantage in the typical child's pre-reading skills and math skills.<sup>113</sup> Magnuson et al. (2004a) found a slightly larger, but still modest 0.15 SD increase (equivalent to an effect size of 0.15).<sup>114</sup> Examining pre-k programs in particular, Magnuson et al. (2004b) found that participation predicts a 1.20 higher reading score and a 0.95 higher math score, which corresponds to effect sizes of 0.12 and 0.10 respectively. This represents about one more question answered correctly and would move the median child from the 50<sup>th</sup> to 55<sup>th</sup> percentile in reading and from 50<sup>th</sup> to 54<sup>th</sup> percentile in math.<sup>115</sup> None of the studies found significant cognitive effects for Head Start participation after adding appropriate demographic controls.<sup>116</sup>

For the typical child, 60% of the cognitive gains associated with participation in non-Head Start center-based programs have faded out by the spring of first grade, leaving a

---

<sup>111</sup> Ibid., 7.

<sup>112</sup> Magnuson, Katherine A., Meyers, Marcia K., Ruhm, Christopher J., and Waldfogel, Jane. (2004a). Inequality in preschool education and school readiness. *American educational research journal* 41, no. 1, 115-157; Magnuson, Katherine A., Ruhm, Christopher, Waldfogel, Jane. (2004b). Does prekindergarten improve school preparation and performance? Downloaded May 10, 2005 from [www.russellsage.org/publications/workingpapers/prek/document](http://www.russellsage.org/publications/workingpapers/prek/document); Magnuson, Katherine, Meyers, Marcia K., Ruhm, Christopher J., Waldfogel, Jane. (Fall 2005). Inequality in children's school readiness and public funding. *Focus* 24, no. 1, 12-18 (note: this article is based on both ECLS-K and Current Population Survey (CPS) data); Loeb et al., How much is too much? The influence of preschool centers on children's development nationwide.

<sup>113</sup> Loeb et al., How much is too much?, 14.

<sup>114</sup> Magnuson et al., Inequality in preschool education and school readiness, 142.

<sup>115</sup> Magnuson et al., Does pre-kindergarten improve school preparation and performance?, 17.

<sup>116</sup> See for example Magnuson et al., Inequality in preschool education and school readiness, 135.

small, but statistically significant effect that indicates close to one additional question answered correctly on relevant assessments.<sup>117</sup> For children who attended pre-kindergarten, 70-80% of associated cognitive gains have faded by this point, equivalent to a small, but statistically significant effect size of 0.03 for reading and math.<sup>118</sup>

Cognitive, and particularly pre-reading gains were larger and longer-lasting for children from economically disadvantaged homes. For example, Magnuson et al. (2004a) found an effect size of approximately 0.30 for pre-reading skills associated with preschool or pre-k attendance for this population.<sup>119</sup> Similarly, Magnuson et al. (2005) found that pre-kindergarten attendance raised the level of kindergarten reading skills exhibited by the average child living in poverty or with a parent with relatively low levels of formal education from the 33<sup>rd</sup> to the 44<sup>th</sup> percentile. Further, reading effects remained large and math effects remained significant for children in families receiving TANF support through the spring of first grade.<sup>120</sup> Alternately, Loeb et al. (2005) found that Hispanic children in center-based programs experienced a 0.23 SD increase in reading scores, which was almost three times the effect size for white children.<sup>121</sup>

Each of the three studies that examined behavioral issues found that ECE program participation had small, but statistically significant negative effects.<sup>122</sup> For example, Magnuson et al. (2004b) found that pre-kindergarten attendance is associated with increased externalizing behavior (effect size of approximately 0.11) and increased negative associations with self-control (effect size of -0.07).<sup>123</sup> Stated alternatively, pre-k is predicted to raise children from the median to the 54<sup>th</sup> percentile of externalizing behavior, and lower them to the 47<sup>th</sup> percentile of self-control.<sup>124</sup>

Loeb et al. (2005) found that the negative behavioral effects of ECE program participation increased in tandem with the number of hours per week in care.<sup>125</sup> For example, while 15 – 30 hours per week increased negative behavioral outcomes by -0.10 SD, 30 or more hours per week increased this to -0.25 SD. Notably, these effects varied

<sup>117</sup> Magnuson et al., *Inequality in preschool education and school readiness*, 135.

<sup>118</sup> Magnuson et al., *Does pre-kindergarten improve school preparation and performance?*, 26.

<sup>119</sup> Disadvantaged was operationalized as living in poverty, having a mother with low levels of education, being in a single-parent family, and/or having a mother who did not speak English. Magnuson et al., *Inequality in preschool education and school readiness*, 139, 142.

<sup>120</sup> Temporary Assistance for Needy Families (TANF). Magnuson et al., *Inequality in children's school readiness and public funding*, 17.

<sup>121</sup> Loeb et al., *How much is too much?*, 12

<sup>122</sup> Magnuson et al. (2004a) only examine cognitive effects.

<sup>123</sup> Similar results were found for Magnuson et al.'s examination of ECLS-K and CPS data. However they note that "in this study the absolute levels of externalizing behavior were typically quite low and the levels of self-control quite high, even for children who attended pre-k." Magnuson et al., *Inequality in children's school readiness and public funding*, 17-18.

<sup>124</sup> These figures represent the effect sizes found after including the maximum numbers of controls for family, neighborhood, and state conditions. Magnuson et al., *Does pre-kindergarten improve school preparation and performance?*, 17.

<sup>125</sup> Researchers used teacher reports of study children's kindergarten-level social-behavioral skills and problems to create a composite score that combines measures of self-control, interpersonal skills, and externalizing behavior. The behavior score is standardized with a mean of zero and standard deviation of one so reported coefficients may be interpreted as effect sizes. Loeb et al., *How much is too much?*, 7.

by subgroup. Middle-class and affluent children in care for 30 or more hours per week exhibited the largest negative outcomes (-0.28 and -0.29 SD respectively). Low-income children in center-based care showed less negative results (-0.12 SD), which were not statistically significant when broken down into dosage components. Hispanic children, in contrast, demonstrated no statistically significant effects at all.<sup>126</sup>

It is particularly noteworthy that participation in both Head Start and other center-based programs is associated with a 2% reduction in the rate of kindergarten retention. Given that approximately 7.5% of the study sample was retained overall, this translates into a 27% rate of reduction.<sup>127</sup>

### GEORGIA EARLY CHILDHOOD STUDY

Summary: The primary purpose of this study was to compare child outcomes from the beginning of the preschool year to the end of first grade among children who had attended the state pre-k program, Head Start, private preschools, or no preschool as four-year-olds. Data collected included child assessments (externally- and teacher-administered), teacher and parent surveys, and classroom observations. The most notable finding was that all four groups of children in the study sample gained relative to national norms from the beginning of their preschool year to end of first grade on measures of receptive language, letter-word recognition, expressive language, and problem-solving.

Duration: 2001 – 2005

Sponsoring agencies: Georgia Office of School Readiness; National Institute for Early Education Research (NIEER).

Principal investigator: Gary T. Henry, Andrew Young School of Policy Studies, Georgia State University.

Primary research questions:

- (1) How much do individual child and family characteristics influence the development of four-year-olds?
- (2) What are the characteristics of the families of resilient children – that is, children who beat the odds of poverty and other risk factors?
- (3) Does the development of four-year-olds differ by program?
- (4) What is the effect of Georgia Pre-K on children in poverty and minorities?<sup>128</sup>

Notable characteristics: The final report of this study is unusual in that it includes a detailed discussion of qualitative findings regarding the apparently significant differences between families of academically “resilient” and “non-resilient” low-income children,

<sup>126</sup> Ibid., 13, 23, 26.

<sup>127</sup> This association persists even after adding large set of controls for child and family demographic characteristics. Magnuson et al., *Inequality in preschool education and school readiness*, 135-136.

<sup>128</sup> Henry et al, *The Georgia Early Childhood Study, 2001 – 2004: Final report*, iii – iv.

with resiliency measured by children's comparative cognitive outcome scores. In-depth interviews were conducted with parents or guardians of 36 children living in poverty, half of whom had academic outcomes that met or exceeded national norms (resilient) and half of whom scored substantially below that level (non-resilient).<sup>129</sup> Findings suggested that while these two sets of families shared many similar characteristics, parents of resilient children demonstrated a significantly stronger (1) understanding of the role of the educational system in American society, (2) conscious commitment to guiding their children's social and academic development, and (3) sense of efficacy regarding the likely consequences of their parenting decisions. While not surprising, these findings are illuminating in that they provide insight into the significant, and academically consequential diversity of low-income families, which are typically lumped together into a single high-risk category.<sup>130</sup>

Study sample: The original study sample included 353 children enrolled in state pre-k, 134 in Head Start, and 143 in private preschool (for a total of 630 children). In the second year of the study, an additional sample of 225 children who did not attend preschool was added.<sup>131</sup> Not surprisingly, these groups varied considerably in terms of race and socio-economic status. For example, 57% of children in the Head Start sample were black, as compared to 26% of the private preschool sample. However, the researchers also developed matched sub-samples of children from the pre-k and Head Start groups who had equivalent scores on a risk index that included race, parent education, family income level, parents' living arrangement, parents' preschool participation, and families' TANF, food stamp, or Medicaid eligibility in order to obtain a better understanding of the differential effects of these programs.<sup>132</sup>

Data collection: Study measures included:

- (1) direct assessments of children's language and literacy, and math and academic skills;
- (2) performance ratings by preschool and kindergarten teachers,
- (3) surveys of teacher attitudes and practices,
- (4) surveys of parent attitudes and school involvement,
- (5) observations and quality ratings of classroom activities, and
- (6) child and family demographic data including gender, race, family economic risk status, mother's level of education, and whether child had lived continuously with both parents since birth.<sup>133</sup>

Data collection schedule: Direct assessments of children were conducted at the beginning and end of the preschool, the beginning of kindergarten, and the end of first grade.<sup>134</sup>

---

<sup>129</sup> Henry et al., *The Georgia Early Childhood Study, 2001 – 2004*, iv, 41.

<sup>130</sup> *Ibid.*, Chap. 3.

<sup>131</sup> A total of 466 (74%) of the original sample remained in the study through the end of first grade; as well as 204 (91%) of the additional non-preschool sample. *Ibid.*, iv.

<sup>132</sup> Henry, Gary T. et al. (August 2003). *Report of the findings from the Early Childhood Study: 2001-02*. Atlanta, GA: Andrew Young School of Policy Studies, Georgia State University, 2, 11-12, 86.

<sup>133</sup> Henry, et al, *An evaluation of the implementation of Georgia' pre-k program*, 2, 9, 14-16.

<sup>134</sup> *Ibid.*

Reports available: The final report for this study was released in late 2005. Two previous reports, which detail the findings of the Early Childhood Study for 2001-02 and 2002-03, are also available. All are posted on the Andrew Young School of Public Policy website at <http://aysps.gsu.edu/epg/index.htm>.

Primary research findings.<sup>135</sup> Children who participated in the Georgia Pre-K program gained relative to national norms from beginning of preschool to end of first grade on measures of receptive language (mean scores increased from 92.9 to 98.0), letter-word recognition (102.7 vs. 111.1), expressive language (90.7 vs. 98.8), and problem-solving (96.9 vs. 109.3). However, the same general pattern of gains was found for the entire study sample, including children who did not attend preschool at all. (The single exception to this rule was some slight deviations in two of the language skill assessments.)<sup>136</sup>

Although pre-k participation was associated with more positive outcomes than other preschool experiences on 11 of 16 measures used, these differences were not statistically significant by first grade. Further, when statistical controls were employed to account for relevant differences in the populations of children attending the different programs, then the growth in children's skill levels largely paralleled one another. One important exception to this pattern, however, was that children from working poor or very low-income families posted greater gains if they attended state pre-k, as opposed to Head Start or private preschool.<sup>137</sup>

Variables with the greatest effect on child outcomes included level of maternal education, family structure, race, and income. For example, nearly one-third of children whose mothers did not complete high school repeated either kindergarten or first grade. Children whose mothers had completed higher levels of education had better language, communication, applied problem solving, and math skills than children whose mothers had completed lower levels of schooling. As noted above, however, significant differences existed among low-income children in the study sample, as measured by both child outcomes and interview data concerning parental beliefs and attitudes deemed critical to children's academic success.

---

#### NCEDL MULTI-STATE STUDY OF PRE-KINDERGARTEN & STUDY OF STATE-WIDE EARLY EDUCATION PROGRAMS (SWEEP)

Summary: The Multi-State Study of Pre-Kindergarten and the Study of State-Wide Early Education Programs (SWEEP) represent by far the most comprehensive examinations of state-funded pre-kindergarten in the U.S. to date. The primary goal of both studies is to

---

<sup>135</sup> One conclusion stated in the study's Executive Summary is that "traditional environmental measures of child care quality are not properly aligned with current Pre-K program goals. We need measures that capture aspects of quality that are related to children's skill development" (Ibid., ix). However, none of the study findings elaborate this point. Further, the report does not discuss how the classroom study sample scored on the three classroom quality assessments employed.

<sup>136</sup> Ibid., iv – v.

<sup>137</sup> Ibid., vii – viii, 80.

understand how important variations in pre-k programs relate to child outcomes at the end of pre-k and in kindergarten. Together, these studies include a sample of over 2,900 pre-kindergarten children enrolled in 705 classrooms in 11 states across the country. The majority of these children are low-income and non-white. Data are collected using child assessments (direct and teacher-administered); parent, teacher, administrator, and principal questionnaires; parent interviews; and videotaped parent-child interactions. Although data analysis is ongoing, the most striking finding to date is that the quality of most pre-k classrooms is significantly lower than expected, particularly with regard to focused instruction designed to engage children in learning.

Duration: 2000 – 2005

Principal investigators: *Multi-State Study:* Richard Clifford and Donna Bryant (Co-Directors); Lynette Aytch and Diane Early (Assistant Directors); Oscar Barbarin, Margaret Burchinal, Carollee Howes, Robert Pianta, and Pam Winton (Principal Investigators).<sup>138</sup>

Primary research questions: Examples of key questions to be addressed include:

- (1) Are a minimum number of hours needed for children to achieve learning goals?  
Are better child outcomes associated with school- or community-based pre-k?
- (2) Do low- or middle-income children gain more from attending pre-k?
- (3) Is higher classroom quality positively associated with better child outcomes?
- (4) What are the kindergarten outcomes of children who attended pre-k?<sup>139</sup>

Notable characteristics: Methodologically, the NCEDL Multi-State Pre-K and SWEEP studies are particularly notable for the careful attention paid to the measurement of dynamic classroom quality. Both studies employ a newly developed instrument, the Classroom Assessment Scoring System (CLASS), to measure the social/emotional climate of the classroom, the nature and quality of teacher-child interactions, and the quality of instruction in participating classrooms. (The studies are additionally using the ECERS-R and another newly developed classroom quality observation instrument, the Emerging Academic SNAPSHOT.)<sup>140</sup> The CLASS is the only standardized instrument available to assess dynamic classroom quality from the pre-kindergarten level through third grade in a way that can be directly related to child outcomes.<sup>141</sup>

Research on the CLASS demonstrates that the quality of the “actual interactions between children and teachers in the classrooms” represents the most important component of the ECE program experience for children. Structural indicators of classroom quality, such as teacher education requirements or class size, cannot be used as

<sup>138</sup> [http://www.fpg.unc.edu/~ncedl/pages/pre-k\\_study.cfm](http://www.fpg.unc.edu/~ncedl/pages/pre-k_study.cfm). Personnel information for SWEEP is not listed on the project website.

<sup>139</sup> Pre-k education in the states. *Early developments* (Spring 2005) 9, no. 1, p. 9.

<sup>140</sup> The Emerging Academic SNAPSHOT uses trained observers to record children’s activities at timed intervals. Nothing has been published on it yet. *Early developments* (Spring 2005) 9, no. 1, pp. 23, 26.

<sup>141</sup> Pianta, Robert C. (Oct. 2003). Standardized classroom observations from pre-k to third grade: A mechanism for improving quality classrooms experiences during the P-3 year. Available: <http://www.fcd-us.org/uploaddocs/standardizedclrmobsfrompre-kto3rdfinal.doc> (July 2005).

proxies for dynamic quality.<sup>142</sup> While such structural measures may create the conditions under which dynamic quality may be strengthened, they do not guarantee high-quality interactions among teachers, children, and staff. This is particularly true in such settings as state-funded pre-kindergarten programs, which, in contrast to many child care centers, typically exhibit a fairly well-organized and adequately resourced physical environment.<sup>143</sup>

Study sample: The Multi-State Study of Pre-Kindergarten includes a carefully selected sample of pre-k classrooms in six states: California, Georgia, Illinois, Kentucky, New York, and Ohio. SWEEP includes an additional five states: Massachusetts, New Jersey, Texas, Washington, and Wisconsin.<sup>144</sup> In combination, these studies include 705 classrooms in 240 sites and over 2,900 pre-kindergarten children. In 2001-02 (when the Multi-State Study began), the 11 states included in the two studies accounted for 79% of all children enrolled in state-funded pre-k programs nationwide, as well as 83% of the state dollars spent on these programs nationally.<sup>145</sup>

To participate in these studies, children could not have an Individualized Education Plan (IEP) and had to speak English or Spanish well enough to understand simple instructions. Of the selected sample, 57% of children lived in families with annual incomes of \$30,000 or less. Fifty-five percent were in families whose annual incomes were less than or equal to 150% of the federal poverty income guidelines. Eighty-six percent spoke English, and 26% frequently spoke Spanish at home. Forty-one percent of the children's mothers completed high school; 17% did not complete it. Thirty-five percent of the participating children were white, 28% Hispanic, and 22% African-American.<sup>146</sup>

Data collection: The primary categories of data being collected for these studies include:

- (1) individual assessments of the literacy, language, and mathematics skills (for four children randomly selected from each participating classroom),
- (2) multiple days of classroom observation by trained evaluators,
- (3) questionnaires of parents, administrators/principals, pre-k and kindergarten teachers, and

---

<sup>142</sup> Ibid., 4, 9.

<sup>143</sup> LaParo, Karen M., Pianta, Robert C., Stuhlman, Megan. (2004). The classroom assessment scoring system: Findings from the prekindergarten year. *Elementary School Journal* 104, no. 5, 412. See also Love, Are they in any real danger?, 3-4, 27-30.

<sup>144</sup> States were selected from among those that had committed significant resources to a pre-k initiative to maximize diversity with regard to geography, program settings (public school or community setting), program intensity (full-day vs. part-day) and educational requirements for teachers. In most states, a stratified random sample of 40 centers/schools was selected from the complete list provided by the state board of education. In four states, less populous regions were omitted due to time and budgetary constraints. Early, Diane et al. (May 2005). Pre-kindergarten in eleven states: NCEDL's Multi-State Study of Pre-Kindergarten and Study of State-Wide Early Education Programs (SWEEP). Preliminary descriptive report. NCEDL working paper downloaded July 7, 2005 from [http://www.fpg.unc.edu/~ncedl/pages/products.cfm#sweep\\_ms](http://www.fpg.unc.edu/~ncedl/pages/products.cfm#sweep_ms), 4-6.

<sup>145</sup> Early et al, Pre-kindergarten in eleven states, 3.

<sup>146</sup> Ibid., 7-9. In addition, 10% of children identified as multi-racial, 3% as Asian/Pacific Islander, and 1% as Native American.

- (4) for about half of the participating families, individual home-based interviews and videotaped parent-child interactions.<sup>147</sup>

Data collection schedule: For the Multi-State Study, individual child assessments were conducted in the fall and spring of the pre-k and kindergarten years. Multiple days of classroom observation were also conducted in the fall and spring of the preschool and kindergarten years. SWEEP followed the same data collection schedule in the pre-k year but relied on teacher questionnaires to assess children's academic and social progress in kindergarten.<sup>148</sup>

Reports available: Two articles based on data from the Multi-State Pre-K Study were published in the July 2005 issue of *Applied Developmental Science*.<sup>149</sup> An initial descriptive report of the Multi-State Pre-K and SWEEP studies is also available at [http://www.fpg.unc.edu/~ncedl/pages/products.cfm#sweep\\_ms](http://www.fpg.unc.edu/~ncedl/pages/products.cfm#sweep_ms).<sup>150</sup> The Spring 2005 issue of *Early developments* also provides summary information on these studies (available at <http://www.fpg.unc.edu/ncedl/pages/ED9.cfm>).<sup>151</sup>

Primary research findings: From the fall to the spring of the pre-k year, children's receptive vocabulary score as measured by the PPVT-III rose from 93.2 to 95.4. Expressive language, as measured by the OWLS rose from 91.2 to 93.0. Early math skills, as measured by the W-J III Applied Problems subtest, rose from 98.1 to 98.7. Given that the majority of children in the study sample were low-income and/or otherwise at-risk, it is not surprising that they began pre-k with scores below the national average.<sup>152</sup> Although children did not collectively achieve the national norm by the end of the program year, this modest increase in test scores demonstrates that more learning occurred during this period than would be expected from maturation alone.

On average, teachers rated children's social skills and behavior as good in the fall of the pre-k year and reported that they had remained the same or improved slightly by the spring. Key measures of social skills began and remained in the 3 – 4 point range on a five point scale, in which "1" was the least and "5" the most desirable score. Similarly,

<sup>147</sup> *Early developments* (Spring 2005) 9, no. 1, pp. 13, 26.

<sup>148</sup> Ibid.

<sup>149</sup> Clifford, Richard M. et al. (2005). What is pre-kindergarten? Characteristics of public pre-kindergarten programs. *Applied developmental science* 9, no. 3, 126-143; and Pianta, Robert et al. (2005). Features of pre-kindergarten programs, classrooms, and teachers: Do they predict observed classroom quality and child-teacher interactions? *Applied developmental science* 9, no. 3, 144-159.

<sup>150</sup> Early, Diane et al. (2005). Pre-kindergarten in eleven states: NCEDL's multi-state study of pre-kindergarten & study of state-wide early education programs (SWEEP): Preliminary descriptive report. Chapel Hill, NC: NDEDL working paper, FPG Child Development Institute, University of North Carolina, Chapel Hill.

<sup>151</sup> *Early developments* (Spring 2005) 9, no. 1.

<sup>152</sup> Key demographic findings include (1) 55% of children in the study sample were from families with annual incomes less than or equal to 150% of poverty line, (2) 45% had mothers whose highest level of education was a high school diploma, (3) 17% had mothers who did not achieve high school graduation, (4) 35% were white, 28% Hispanic, and 22% African American, (5) 86% spoke English at home, while 26% frequently spoke Spanish. Early et al, Pre-kindergarten in eleven states, 7-8.

key measures of behavior problems began and remained very close to 1.5 on a scale in which “1” was the most and “5” the least desirable score.<sup>153</sup>

The average level of classroom quality as scored on the 7-point ECERS-R was 3.86. Eighty-one percent of classrooms were in the minimal quality range (3 – 5 points); 11% were inadequate (0 – 3 points); and 8% were good-to-excellent (5 – 7 points). This average level and range of quality is lower than found by other large-scale studies of ECE programs, including FACES and the Cost, Quality, and Outcomes (CQO) project. Researchers hypothesize that this may be due to the fact that (1) the study used the revised ECERS, which was not used by the 1997 cohort of FACES or the CQO study and includes items on diversity, math, and science that may lower overall scores, and (2) the classroom sample included a large number of part-day programs, which spend a substantial proportion of time on routine activities such as providing snacks.<sup>154</sup>

Consistent with the latter theory, researchers found that meals and routine activities typically accounted for 36% of preschool day. During this time, children were generally not engaged with adults, having learning-oriented conversations, singing, playing number games, or engaged in other learning-directed activities. The average quality of the classroom instructional climate was typically low, rated at 2.47 on the 7-point CLASS scale. In contrast, the average emotional climate scored a substantially higher 5.22 on this instrument. This disparity indicates that while the average pre-k classroom provides a relatively warm and nurturing environment, relatively little time is spent in meaningful instructional activities designed to engage children in learning.<sup>155</sup>

It is important to note that this relatively low rating of global quality and even lower assessment of instructional quality was attached to a study sample composed of programs characterized by high levels of structural quality. For example, class size and teacher – child ratios typically met or exceeded recommended standards, and teachers were on average better educated and more highly paid than is common among early childhood educators. This finding, in conjunction with analyses of which variables best predict classroom quality, suggest that quality in the pre-k context is most closely related to proximal teacher and child characteristics, rather than distal features of programs and teachers.<sup>156</sup>

#### **(4) Overview of research questions and methods**

As demonstrated by the initiatives discussed in the preceding section, the most important research on ECE programs during the past decade focused on such large-scale, commonly utilized programs as Head Start, child care, and state pre-k, as contrasted to the small model programs that were the primary subject of more rigorous studies several decades ago. Moreover, the basic structure of research on these various types of large-scale programs has become more similar. Differences among formerly distinct areas of

---

<sup>153</sup> Ibid., 29-30.

<sup>154</sup> *Early developments*, 18; Clifford et al, What is pre-kindergarten?, 136-137, 141.

<sup>155</sup> *Early developments*, 18.

<sup>156</sup> Pianta et al, Features of pre-kindergarten programs, classrooms, and teachers, 144, 155-157.

inquiry (e.g., cognitively-oriented studies of preschool as opposed to attachment-centered examinations of child care) have been replaced by a common emphasis on examining the relationship among program structures, classroom quality, teacher characteristics, child and family characteristics, and child outcomes.

All of the studies examined in this report share this basic agenda, despite significant differences in method, scope, and types of programs studied. Consequently, while the specific research questions addressed vary, most include some variation of the following:

- (1) What is the relationship between program and teacher characteristics, classroom quality, and child outcomes?
- (2) What is the relationship between ECE program experience, child and family characteristics, and child outcomes?
- (3) Is preschool participation associated with improved outcomes on measures of children's cognitive and social-emotional development?
- (4) What is the long-term relationship between preschool participation and children's later school achievement?

The research projects discussed in this report employ a wide variety of methods to answer such core questions. Most fundamentally, these may be divided into the three primary categories of experimental, quasi-experimental, and non-experimental designs.<sup>157</sup> Across these different categories, researchers employed a wide range of data collection methods. The most important and frequently used include the following:

- (1) *Preschool-year child assessments.* These assessments are most commonly administered during the fall and spring of the preschool year. All of the studies surveyed employed standardized, normed instruments, most commonly to assess cognitive development. Most also included teacher-administered instruments, particularly to assess children's social-emotional development and classroom behavior.
- (2) *Post-preschool child assessments.* The more rigorous longitudinal studies continue to employ direct and teacher-administered assessments, typically in the spring of the academic year under consideration.
- (3) *Other child-level information.* Demographic information (race, gender, home language, special needs, etc.) is commonly collected on study children. Less commonly, risk indices are constructed that summarize data on multiple child and family variables, a practice that is strongly recommended by some researchers. In longitudinal studies, data on whether study children are retained in grade or placed in special education have been found to be particularly useful.

---

<sup>157</sup> To review, experimental and quasi-experimental designs are distinguished by their use of treatment and control or comparison groups. Only the former, however, randomly assigns study subjects to either participate or not participate in a given intervention. Non-experimental designs, in contrast, do not employ a control or comparison group. As such, they cannot assess causal relationships in the same manner as an experimental or quasi-experimental study. However, if non-experimental studies collect sufficient data to control for factors expected to significantly affect outcomes, they may assess the statistical significance of critically important relationships, such as the association among ECE program quality, child and family characteristics, and child outcomes.

- (4) *Family characteristics.* All but one of the studies surveyed collected demographic information on study children's parents or guardians (e.g., education, employment, marital status, family income). Some additionally surveyed parents on their attitudes, beliefs, and practices with regard to child rearing. Some questioned parents regarding their involvement in their child's classroom and/or satisfaction with her ECE program. Some explored issues of maternal depression, which has been found to have a particularly important impact on child development. A few included observational and/or videotaped data on the quality of the home environment and parent/child interactions.
- (5) *Teacher characteristics.* Most of studies surveyed collected data on teacher characteristics (e.g., demographics, education and training, teaching experience). Some also surveyed teachers to ascertain their attitudes, beliefs, and/or practices with regard to teaching young children.
- (6) *Classroom quality.* The majority of the studies surveyed included structural and dynamic measures of program quality, collected directly by trained observers at least one time per academic year.
- (7) *Preschool program characteristics.* All of the studies surveyed collected data on such program characteristics as duration, auspice, staffing, and curriculum. Some additionally surveyed and/or interviewed non-teaching program staff.
- (8) *School and classroom characteristics (post-preschool).* Longitudinal studies that tracked study children beyond the preschool year directly assessed classroom quality and/or collected data on school characteristics (e.g., percentage of students receiving free or reduced-price lunches, average levels of student achievement in key subject areas, attendance and dropout statistics). Some additionally surveyed and/or interviewed teachers, principals, and staff.

In all cases, sampling techniques were employed to reduce a larger universe of children and classrooms to smaller number of representative cases. Sampling protects individual children and teachers involved in a study by ensuring that child outcome data in particular cannot be inappropriately used to evaluate them on an individual basis. Sampling also reduces the cost of a given evaluation by minimizing the number of children, teachers, and classrooms that must be examined in order to evaluate the larger program. It should be noted, however, that it is generally difficult for a single project to have a sample that is large and diverse enough to allow its results to be generalized both to the general U.S. population and particular demographic groups such as linguistic and ethnic minorities. Typically, study samples are dominated by children with a particular demographic profile (e.g., low-income for both Head Start studies, middle-income for the NICHD). Consequently, when interpreting research findings, it is critical to consider the nature of the study sample in conjunction with the larger structure of the research design.

### **(5) Synthesis of research findings**

Considered as a whole, the primary research findings of the ten studies detailed in Section 3 paint a complex and at times inconsistent picture. Again, it is important to emphasize that this report does not attempt to contextualize these findings within the much larger body of literature that examines the relationship between ECE program

participation and children's development. Instead, it presents a more focused examination of both research methods and findings, emphasizing the importance of each. Although, as discussed above, it is true that the ten studies presented share certain common methodological characteristics, it is also true that there are substantial differences among them. Any consideration of their respective findings should be informed by the particularities of the method employed; all, to greater and lesser degrees, have their own strengths and limitations.

This synthesis of research findings focuses on children's short- and long-term cognitive and social-emotional outcomes.<sup>158</sup> This emphasis reflects both the general focus on child outcomes that dominates discussions of findings of ECE program research in the public policy and early childhood fields, as well as their centrality to issues of ECE program accountability in particular. Although issues of program quality are equally important, in this report they are discussed primarily in terms of their relationship to child outcomes. This focus parallels the project descriptions presented above, as the primary research findings of nine of the ten projects emphasize child outcomes and/or the relationship between quality and outcomes.

*Short-term cognitive outcomes.* All of the studies included in this report found preschool participation to be associated with improvements in children's levels of cognitive skill. Standardized measures showed progress relative to national norms and/or statistically significant gains for the typical child in each study. Several studies that compared the relative gains of children from more or less economically advantaged families also found that low-income children reaped comparatively larger cognitive gains. However, the estimated size of these associations or effects varied significantly among different projects, both for the typical child and for specific subgroups. Further, the particular domains in which gains occurred varied among the different studies. Finally, the different methods employed make this research difficult to compare in some cases.

For example, the Head Start Impact Study found small positive effects (ranging in size from 0.10 – 0.24) on letter-word identification and letter naming (both 3- and 4-year-olds), spelling (4s only), and vocabulary, color naming, and perceptual motor skills (3s only). Analyses of ECLS-K data similarly found small positive effects (ranging in size from 0.10 – 0.15) on children's pre-reading and math skills associated with participation in a non-Head Start center-based program. However, while the former employed a rigorous experimental design that compared the impact of a Head Start-eligible child being admitted or denied entry upon application to the program, the latter used a non-experimental design that retrospectively analyzed ECE program effects based on kindergarten test score data (and did not collect entry-level data at preschool). Given these very different methodologies, the reported effect sizes – while both small – cannot be considered equivalent.

---

<sup>158</sup> Short-term effects describe those occurring during the preschool year(s), including as measured at the beginning of kindergarten. Long-term effects, which may include other educationally relevant variables measured, include effects through the spring of kindergarten or beyond.

Alternately, both the Oklahoma Universal Pre-K and the NIEER Five-State Pre-K studies found moderate-to-large positive effects on the cognitive measures employed. Specifically, the former found effect sizes of 0.79 for letter-word identification, 0.64 for spelling, and 0.38 for math; while the latter found effect sizes of 0.64 for print concepts, 0.26 for vocabulary, and 0.28 for math. Since both studies employed a regression-discontinuity design, these findings are relatively comparable (although the Oklahoma study sample drew only from Tulsa while the NIEER study sampled children across five states). The question of how these findings should be considered vis-à-vis those emanating from research that did not employ this method, however, is debatable. Particularly given that both studies constructed comparison groups using children whose birthdays could be up to a full year short of the pre-k entry cut-off date, it is not unreasonable to question whether these unusually large effect sizes might be in part attributable to the method employed.<sup>159</sup>

Several studies found that children from economically disadvantaged families reaped comparatively larger short-term cognitive gains from ECE program participation than their more affluent peers.<sup>160</sup> In the Oklahoma study, children who qualified for free lunch posted larger effects from pre-k participation on each of the three cognitive assessment used. In the Cost, Quality, and Outcomes Study, there were generally larger positive associations between program participation and cognitive outcomes for children whose mothers had low levels of education. Several analyses of ECLS-K data found that cognitive and particularly pre-reading gains were larger and longer-lasting for low-income children. Magnuson et al., for example, found an effect size of approximately 0.30 for pre-reading skills associated with preschool or pre-k attendance for this population.<sup>161</sup>

The FACES study found significant short-term cognitive outcome differences between children whose early language and number skills were in the highest- or lowest-scoring quartile of their cohort at program entry. Those entering with lower levels of knowledge and skill showed the largest gains. While those entering with higher levels of skills posted higher scores in the spring than the fall, their standing in terms of national norms declined slightly for 7 of the 9 measures used. For example, while the vocabulary scores of the overall study sample experienced a mean gain of 3.8 points on the PPVT-III, scores of the bottom quartile rose 8.4 points (more than half a standard deviation), while those of the top quartile declined .5 points relative to national norms.

---

<sup>159</sup> An alternative hypothesis is that the association between participation in state pre-k programs and relatively large cognitive gains is due to the fact that these programs are of comparatively higher quality. However, NCEDL findings regarding the lower than expected quality of state pre-k classrooms advise against this explanation.

<sup>160</sup> The NIEER Five-State Pre-K Study found comparatively larger, statistically significant positive effects for children with free- or reduced-price lunch status in Oklahoma and South Carolina only. Barnett et al., *The effects of state pre-kindergarten programs*, 13. The NICHD study found only “limited evidence to suggest child care experiences moderate the negative associations between family risk and child outcomes.” However, the study sample was constructed in a way that limited the representation of low-income, high-risk families. NICHD Early Child Care Research Network, *Child care and child development*, 171. Alternately, both the Head Start Impact Study, FACES, and the Chicago Child-Parent Centers project had samples drawn almost exclusively from low-income and/or high-risk families.

<sup>161</sup> Magnuson et al., *Inequality in preschool education and school readiness*, 142.

*Long-term cognitive and educational outcomes.* Several studies document the continuation of small, statistically significant positive gains on cognitive measures attributable to ECE program participation through the early elementary school years. Analysis of ECLS-K data found that while 60% of the cognitive gains associated with participation in non-Head Start center-based programs had faded out for the typical child by the spring of first grade, this left a small effect equivalent to almost one additional question answered correctly on relevant assessments.<sup>162</sup> Such effects were larger and longer-lasting, however, for children living in low-income households or with less highly educated parents. For the average child living in poverty or with a parent with relatively low levels of formal education, attending pre-k raised the level of kindergarten reading skills from the 33<sup>rd</sup> to the 44<sup>th</sup> percentile. Further, reading effects remained large and math effects remained significant for children in families receiving TANF support through the spring of first grade.<sup>163</sup>

The Georgia Early Childhood Study found that children who participated in the state pre-k program gained relative to national norms from beginning of preschool to end of first grade on measures of receptive language (mean scores increased from 92.9 to 98.0), letter-word recognition (102.7 vs. 111.1), expressive language (90.7 vs. 98.8), and problem-solving (96.9 vs. 109.3). However, the same general pattern of gains was found for the entire study sample, including children who did not attend preschool at all.

Other educationally-relevant long-term outcomes were measured by indicators other than cognitive assessments. Analysis of ECLS-K data, for example, found a 2% reduction in the rate of kindergarten retention (from 7.5% to 5.5%, which represents a 27% reduction overall).<sup>164</sup> Most notably, the Chicago Child-Parent program study found that by age 15, former program participants had a five-month gain in both reading and math achievement over the comparison group. By age 18, 14% had received special education services, as contrasted to 25% of the comparison group; 23% had repeated a grade, compared to 38% of the comparison group. By age 22, 65% of program participants had completed high school, as compared to 54% of the comparison group.<sup>165</sup>

These findings, however, represent the average effects of 1 – 7 years of program participation and exhibit a strong dosage effect. Consequently, they cannot be considered directly analogous to participation in a traditional one- or two-year preschool program. It is also important to note that analyses of why CPC participation was associated with these outcomes identified early cognitive development as only one of several major factors, including parent involvement in school, quality of the post-preschool learning environment, and avoidance of school mobility. These findings suggest that cognitive gains should be viewed neither as the most important indicators of long-term ECE program effects, nor as a direct or exclusive means of leveraging such positive effects.

---

<sup>162</sup> Ibid., 135.

<sup>163</sup> Magnuson et al, *Inequality in children's school readiness and public funding*, 17.

<sup>164</sup> Magnuson et al, *Inequality in preschool education and school readiness*, 135-136.

<sup>165</sup> *Newsletter of the Chicago Longitudinal Study* 1 (Aug. 2000) & 2 (June 2002);

<http://www.waisman.wisc.edu/cls/index.html>.

*Classroom quality and cognitive outcomes.* Both the NICHD and the Cost, Quality, and Outcomes study found significant positive relationships between classroom quality and cognitive outcomes.<sup>166</sup> Controlling for the study's full set of child and family covariates, NICHD researchers found that children in classrooms rated in the highest tercile of quality obtained higher scores on assessments of pre-academic and language skills than those in the bottom tercile (effect sizes of .24 and .15, respectively). Use of a more limited set of covariates increased effect sizes to .39 and .29). By the end of third grade, having participated in higher quality child care continued to have a small positive effect on cognitive outcomes as measured by standardized tests of math, reading achievement, and memory, with effect sizes ranging from 0.07 – 0.09.<sup>167</sup>

The Cost, Quality, and Outcomes Study found that preschool quality, considered independently of K – 2 classroom quality, had a modest positive association with children's receptive language, early math skills, cognitive/attention skills through kindergarten; and on math and cognitive attention skills through second grade. In general, these relationships were stronger for children whose mothers had low levels of education. After controlling for the quality of children's K – 2 classroom experiences, the relationship between preschool quality and math skills remained significant through second grade.<sup>168</sup> Like the CPC study, these findings suggest that while preschool quality may have a significant effect on children's cognitive outcomes through the early elementary years, the quality of their later early education experiences is also important, both independently and with regard to building on initial ECE program gains.

*Social-emotional outcomes.* Findings on the short- and longer-term relationship between ECE program participation and children's social-emotional outcomes varied significantly among the eleven studies featured in this report. Some found small-to-moderate positive associations or effects; other found negative relationships of a similar magnitude. Outcomes tended to vary significantly among subgroups and by number of hours spent in center-based care. Although some consistent findings were reported across studies, others seem to contradict one another. Overall, it appears that the relationship between ECE program participation and children's social-emotional development is significantly more complex, varied, and potentially troubling than is true with regard to their cognitive development.

---

<sup>166</sup> The FACES study found use of the High/Scope curriculum, higher teacher salaries, teachers' educational credentials (B.A., A.A., or higher), and provision of preschool services for a longer period each day associated with small increases in entry-graduation achievement gains in some cognitive outcomes. However, it also found higher scores on the ECERS-R Language subscale associated with lower scores on children's dictation measures. In addition, children who had teachers with higher scores on the ECE practices measure had significantly smaller gain scores in book knowledge ( $b = -.08, p < .05$ ) and design copying ( $b = -.09, p < .01$ ) than those taught by teachers with lower scores. (No association was found between cognitive outcomes and class-level scores on the Assessment Profile Individualizing Scale or the Caregiver Interaction Scale (CIS).) Zill et al, *FACES 2000 Technical Report*, Chap. 6.

<sup>167</sup> NICHD Early Child Care Research Network, *Child care and child development*, 389.

<sup>168</sup> Peisner-Feinberg et al, *The children of the cost, quality, and outcomes study go to school: Technical report*, 34.

The NICHD study presented the most troubling short- and long-term findings. This study emphasized the negative effects of cumulative time spent in non-maternal care during the first four years of life: the more this number increased, the more that mothers, caregivers, and/or teachers reported that children exhibited externalizing problems and conflict with adults at 4½ years and in kindergarten. These effects remained significant after controlling for the quality, type, and instability of child care, as well as for maternal sensitivity and other family background factors.<sup>169</sup> By the end of third grade, spending over 30 hours per week in child care had a small negative association with work habits and social skills (effect sizes ranging from 0.01 – 0.12). However, the relationships between amount of time in child care, and externalizing behaviors and teacher-child conflict decreased during the primary grades and were insignificant by Grade 3.<sup>170</sup>

Findings from the ECLS-K largely paralleled those of the NICHD. Loeb et al., for example, found that while 15 – 30 hours per week in center-based care had a negative association with behavioral outcomes of -0.10 SD, 30 or more hours per week increased this effect to -0.25 SD. These effects varied by subgroup: Middle-class and affluent children in care for 30 or more hours per week exhibited the largest negative outcomes (-0.28 and -0.29 SD respectively). Low-income children showed less negative results (-0.12 SD), which were not statistically significant when broken down into dosage components. Hispanic children, in contrast, demonstrated no statistically significant effects at all.<sup>171</sup>

Magnuson et al.'s analysis of ECLS-K data found pre-kindergarten attendance associated with small levels of increased externalizing behavior (effect size of approximately 0.11) and decreased self-control (effect size of -0.07). Alternatively stated, pre-k participation is predicted to raise children from the median to the 54<sup>th</sup> percentile of externalizing behavior, and lower them to the 47<sup>th</sup> percentile of self-control.<sup>172</sup> Although derived using a completely different set of research and analytic methods, this finding is inconsistent with those of the NCEdL, which found teacher-reported social skills and behavior problems to have remained the same or improved slightly between the fall and spring of the pre-k year. Given that pre-k children are typically in half-day programs (although how the rest of their weekdays are spent is not reported), the NCEdL's finding of neutral-to-small positive changes could be hypothesized to be in keeping with the NICHD and ECLS-K findings that negative social-emotional effects are associated with higher cumulative and/or absolute hours in center-based programs. However, it is in conflict with Magnuson et al.'s findings regarding the small negative effects of pre-k participation on children's externalizing behavior and self-control.

The FACES study presents an interesting comparison case, as it found both positive and negative social-emotional change for Head Start children, depending on where they stood on relevant assessments at the time of program entry. Children who demonstrated

---

<sup>169</sup> NICHD Early Child Care Research Network, *Child care and child development*, 315.

<sup>170</sup> NICHD Early Child Care Research Network, *Early child care and children's development in the primary grades*, 537-570.

<sup>171</sup> Loeb et al, *How much is too much?*, 13, 23, 26.

<sup>172</sup> Magnuson et al, *Does pre-kindergarten improve school preparation and performance?*, 17.

the lowest levels of social skills and the highest levels of problem behaviors at this time showed the greatest gains. For example, the mean “cooperative behavior” score for those in the bottom quartile rose from 8.72 at program entry to 15.40 at graduation. However, the corresponding scores of children who scored in the top quartile of social skills and minimal problem behaviors at program entry declined from 20.47 to 20.00.<sup>173</sup> This pattern parallels the study’s findings regarding cognitive outcomes, in which children who scored in the lowest quartile at program entry benefited significantly, while those who entered in the highest quartile did not.<sup>174</sup>

The Cost, Quality, and Outcomes Study found that preschool quality had a modest positive association with children’s sociability through kindergarten and on problem behaviors through second grade. After controlling for K – 2 classroom quality, children of mothers with fewer years of education continued to show a significant reduction in problem behaviors. In addition, children in preschool classrooms with higher levels of peer interaction during play had significantly better relationships with their peers in second grade. In contrast, children in preschool classrooms with high incidence of problem behaviors and low levels of teacher-child closeness demonstrated higher levels of aggression and disruptive behavior at this time.<sup>175</sup> These findings present an interesting contrast to those of the NICHD, which found that classroom quality did not mediate the relationship between cumulative hours in child care and negative social-emotional outcomes. Given that the Cost, Quality, and Outcomes Study did not analyze cumulative time in care, one possible hypothesis is that while quality may have a positive effect up to a certain threshold, it may be insufficient to compensate for high cumulative hours in child care, at least for the typical child.

## **(6) Conclusion: Implications for ECE program accountability issues**

The ten studies highlighted in this report offer important lessons to early childhood professionals and policy makers concerned with ECE program accountability issues. In terms of both methods and findings, they collectively illustrate the complexities of child development and the challenges involved in attempting to analyze its relationship to ECE program participation. In particular, findings regarding the effects of participation on children’s cognitive and social-emotional development demonstrate the importance of examining the relationships among program structure and quality, child and family characteristics, and multiple and diverse measures of child outcomes before rendering judgment about the strengths and limitations of any given ECE program.

The field needs to grapple more seriously with research findings that suggest that in at least some very important cases, children’s social-emotional development may be negatively affected by ECE program participation, even while their cognitive skills are improving. Clearly, any program evaluation or accountability strategy that exclusively

---

<sup>173</sup> Zill et al, *FACES 2000 Technical Report*, 3-6.

<sup>174</sup> This finding could be interpreted as evidence of regression to the mean, a common phenomenon in intervention studies.

<sup>175</sup> Peisner-Feinberg et al, *The children of the cost, quality, and outcomes study go to school: Technical report*, 32-34.

examines cognitive outcomes will not allow the possibility of examining social-emotional development at all. Given the intensive and singular focus on cognitive outcomes in the upper elementary and secondary grades driven by the *No Child Left Behind Act*, there is ample reason to be concerned that such methods may be pushed down to the early childhood level. In addition to violating well-established understandings of the nature of child development and the most scientifically credible means for its assessment, such an approach would ignore the cautionary lessons represented by the NICHD and ECLS-K findings in particular.

At a minimum, any program evaluation or accountability method should include appropriate measures of both cognitive development, including vocabulary, pre-reading, and math skills, and social-emotional development, including externalizing and internalizing behaviors and classroom adjustment. Although multiple measures make the assessment process more time-consuming and expensive, it is necessary to gain reasonably valid and comprehensive information. The fact that only a sample of children should be assessed in this way alleviates such challenges to some extent. However, it remains an inescapable fact that the complexities of child development cannot be meaningfully reduced to a single set standardized test scores.<sup>176</sup>

The significant differences in cognitive and social-emotional outcomes among the various subgroups of children discussed in this report underscore the importance of collecting data on relevant child and family characteristics in the program assessment process. Indicators of educational disadvantage, such as low family income and/or low levels of maternal education, are particularly important to examine, given findings that children living in such households typically reap comparatively greater gains from ECE program participation, as well as the well-deserved emphasis in policy circles on closing the early education gap. Similarly, findings regarding significant racial/ethnic subgroup differences, as well as the historic legacy of racial discrimination and recent growth in immigrant and ESL populations, underscore the importance of structuring assessments to account for such critical markers of diversity. It is also true, however, that the expansion of universal pre-k programs raises questions regarding how ECE program participation affects the typical child to a new level of policy importance.

Findings from the NCEDL, NICHD, and Cost, Quality, and Outcome studies underscore the importance of including assessments of classroom quality in ECE program evaluation and accountability procedures. The NCEDL finding that 81% of the classrooms in the study sample (which represents almost 80% of state pre-k programs nationwide) received only a minimal quality rating on the ECERS-R and a low instructional climate rating on the CLASS indicates that the quality of many publicly-funded ECE programs needs to be improved. This is particularly true given the NICHD's finding that higher classroom quality has a significant positive effect on children's pre-academic and language skills, as well as the Cost, Quality, and Outcomes Study's findings regarding the longer-term positive associations between classroom quality and child outcomes.

---

<sup>176</sup> Meisels, *Accountability in early childhood*.

It is also important to keep in mind, however, that the one existing study of a large-scale, publicly funded ECE program that tracked children from preschool to adulthood found that the cognitive boost provided by program participation was neither a direct nor exclusive predictor of long-term positive outcomes. Instead, the CPC study demonstrated that what Reynolds (2000) termed “cognitive advantage” interacted with the effects of parent involvement, the quality of the post-preschool learning environment, and avoidance of school mobility to produce important long-term effects including increased rates of high school graduation and decreased rates of grade retention, special educational placement, and juvenile arrests. This finding once again emphasizes the fact that ECE programs should not be assessed on the basis of children’s cognitive score gains alone. Indeed, factors such as a program’s ability to increase parental engagement with their children’s out-of-home education experiences may be equally important predictors of such long-term positive outcomes.

The findings of the CPC study regarding the key mechanisms that best explain the long-term benefits of ECE program participation are reinforced by the NICHD’s emphasis on the importance of family experiences in determining children’s cognitive and social-emotional outcomes, as well as their experience of out-of-home care. They are also in line with the findings of the Cost, Quality, and Outcomes Study, which showed that while preschool quality can affect key child outcomes through at least second grade, the quality of their K – 2 classroom experiences also plays an important role. These findings, as well as our more general understanding of child development, suggest that ECE programs will be most effective at producing positive cognitive and social-emotional outcomes for participating children when they function in harmony with their families, communities, schools, and other educationally-relevant institutions. Properly conceived, ECE program evaluations and accountability procedures should help programs to achieve the challenging but fruitful task of playing a maximally supportive role within the larger fabric of children’s lives.