

Running Head: EFFECTS OF PREKINDERGARTEN

State Efforts to Evaluate the Effects of Prekindergarten:

1977 to 2003 ^{1,2}

Walter S. Gilliam and Edward F. Zigler

Yale University Child Study Center

April 19, 2004

Address all correspondence to:

Walter S. Gilliam, Ph.D.

Associate Research Scientist

Yale University Child Study Center

230 South Frontage Road

New Haven, Connecticut 06520-7900

Phone: 203-785-3384; Fax: 203-785-7926

E-mail: walter.gilliam@yale.edu

State Efforts to Evaluate the Effects of Prekindergarten:**1977 to 2003**

In recent decades, state involvement in prekindergarten services has increased dramatically.³ In 1980, only 10 states provided publicly funded prekindergarten services, primarily targeting low-income children in the 3- through 4-year old range. Today, there are at least 55 statewide prekindergarten systems that are funded and administered by states, operating in 40 states (see Figure 1). Typically, there are two different ways that states invest in prekindergarten services: (a) developing, funding and administering their own unique prekindergarten systems and (b) providing state funds to increase the number of children served by Head Start in their state. Currently, 11 states both fund their own unique prekindergarten and supplement Head Start, 27 fund a state prekindergarten system, but do not supplement Head Start, 2 supplement Head Start, but do not fund a unique state prekindergarten system, and 10 states do neither (see Figure 2). Based on currently unpublished data from the ongoing Yale National Prekindergarten Study,⁴ state prekindergarten programs serve over 980,000 preschoolers in about 55,000 classrooms or groups of children across the nation. As of 2000, state-funded prekindergarten systems in aggregate surpassed Head Start as the leading provider of preschool services for children in the 3- to 4-year-old range. With an annual aggregate budget of \$2 to \$3 Billion, state-funded prekindergarten programs represent a massive public investment in early childhood education.

Figure 1.

Dramatic Increase in States Providing Prekindergarten Services

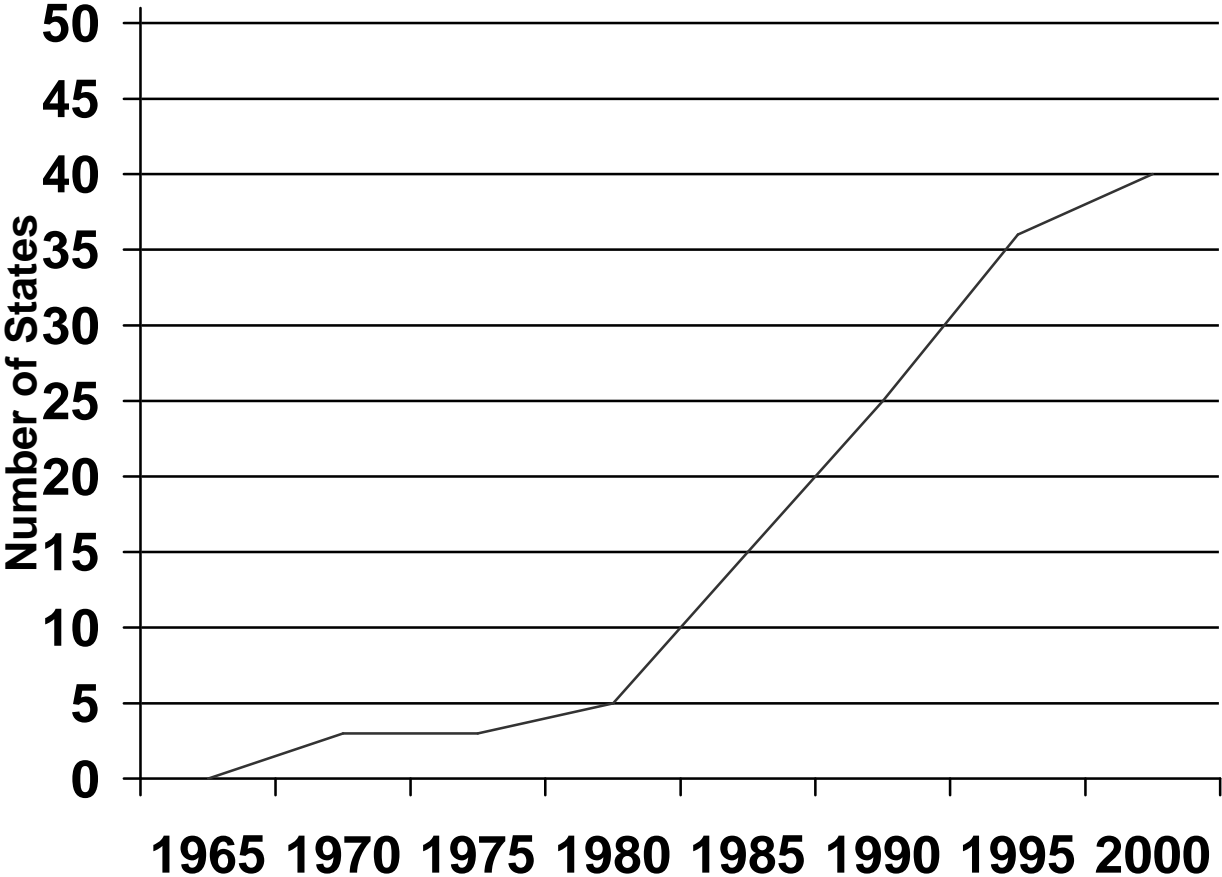
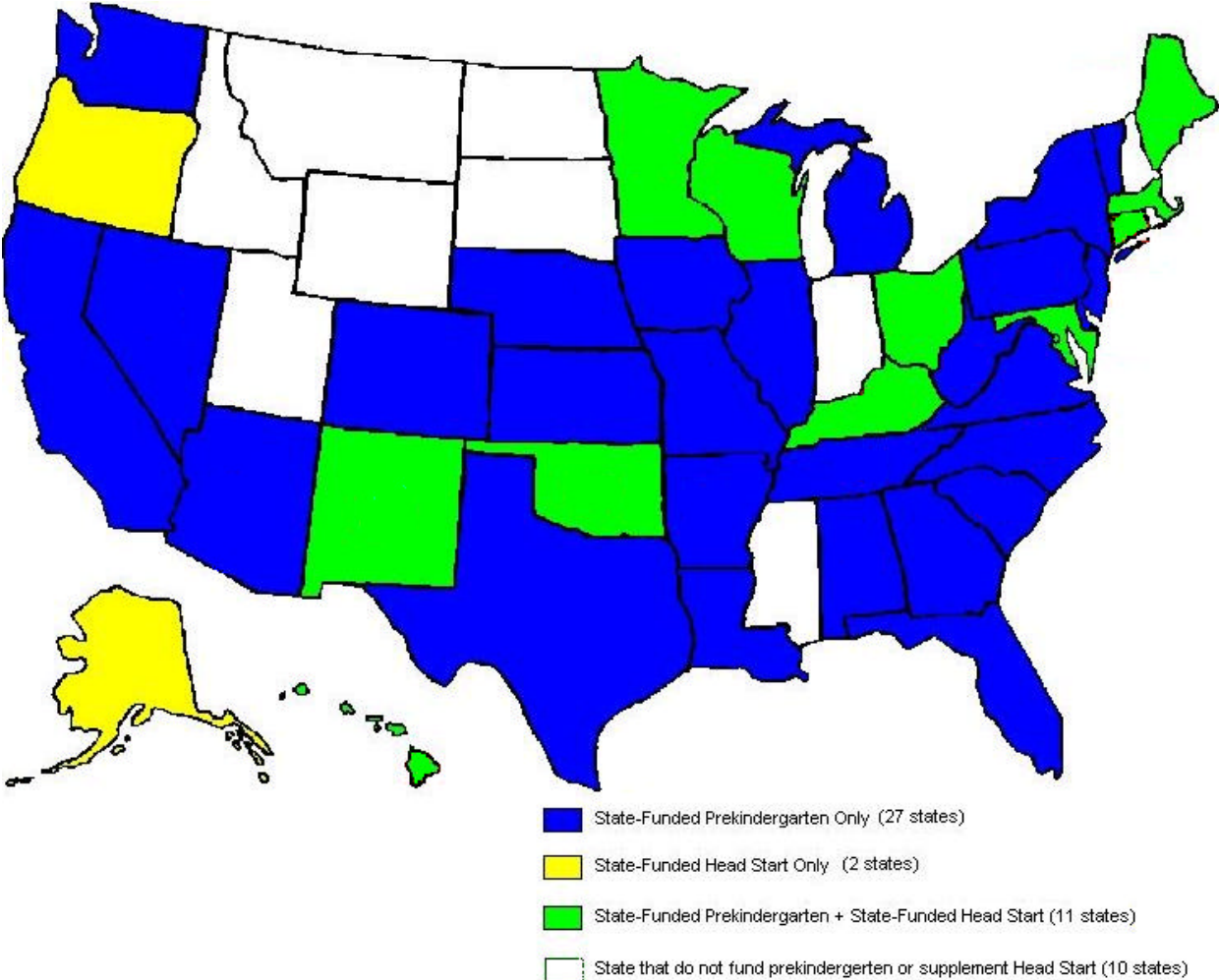


Figure 2.

States that Fund Prekindergarten (FY 2004)



These state-funded prekindergarten programs also are of immense early childhood policy importance for reasons beyond their current size and cost. As Congress debates granting as many as eight states control of Head Start programs operating in their state, these state-funded prekindergarten programs would be the likely recipients of any devolved Head Start funds, potentially expanding their scope dramatically. Also, as several states move closer to universal prekindergarten, it is state-funded prekindergarten systems that likely would be expanded to serve these children.

Programs that rely on public funds increasingly have been held accountable for demonstrating their effectiveness. State-funded prekindergarten initiatives are no exception. In many states, formal evaluation of program implementation and impact is mandated in the state legislation authorizing the program. However, most states have not formally evaluated the impacts of their state-funded prekindergarten programs, and most states that have conducted such evaluations have relied on less-than-rigorous methods of estimating program effects (Gilliam & Ripple, 2004; Gilliam & Zigler, 2001; Ripple, Gilliam, Chanana, & Zigler, 1999). An impact evaluation assesses the degree to which a program improves outcomes among its participants, in this case the children who attend the preschool program. In a review of the long-term cognitive and academic impacts of both model and large-scale public preschool programs (Barnett, 1995; 1998), it was found that in many cases public programs had weaker effects than the often higher quality and better implemented model programs, highlighting the need for scientifically valid evaluations of the effectiveness of these large-scale public programs.

Unfortunately, very little is known about these state-funded prekindergarten programs. What little published data exists focuses exclusively on cataloguing the number of children served, program costs, and a variety of service mandates (e.g., minimum teacher credentials,

maximum class size and child-staff ratios, etc.), with virtually no national data on classroom-level implementation or child-level impacts (Barnett, Robin, Hustedt, & Schulman, 2004; Gilliam & Ripple, 2004; Knitzer & Page, 1998; Mitchell, 2001; Mitchell, Ripple, & Chanana, 1998; Ripple et al., 1999; Schulman et al., 1999; Smith, Fairchild, & Groginsky, 1997). Of these 55 state-funded prekindergarten systems operating in 40 states,⁵ relatively few have been evaluated in terms of their impact on child outcomes, and only three of these evaluations have ever appeared in peer reviewed publications (Horan, Irvine, Flint, & Hick, 1980; Marcon, 1999; Xiang & Schweinhart, 2002).

By 1998, only 13 of the current 40 states that fund prekindergarten had completed and released an evaluation of the impact of their program on the participating children. The 60 reports that present the findings of these 13 prekindergarten evaluations were summarized and critiqued by Gilliam and Zigler (2001). Between 1998 and 2003, state-funded prekindergarten systems in an additional five states had been evaluated, plus two different prekindergarten systems in states that had already evaluated a different prekindergarten system within the same state. Two of these new reports focus on universal prekindergarten systems (Georgia and Oklahoma). This paper updates the original report by Gilliam and Zigler (2001) by adding these additional evaluations in order to provide a current discussion of the methods and findings of these evaluations. In order to be identified as an evaluation of a state-funded prekindergarten program, an evaluation must focus on a state-funded prekindergarten system (as defined earlier), provide data by which program impacts on child outcomes can be reasonably estimated (even if the methods are weak), and evaluate a statewide sample. The one exception to this rule is the Oklahoma Early Childhood Four-Year-Old Program evaluation (Gormley & Gayer, 2003), which evaluated the program only in the Tulsa Public Schools. Although results from a single

urban school system may not generalize to the entire state, this evaluation is included because it represents an evaluation of a universal prekindergarten program and because of the sophistication of its design and analyses, as described later.

Review Methods

The 60 evaluation reports summarized by Gilliam and Zigler (2001) were identified and obtained through several methods. First, state prekindergarten administrators were asked whether their programs had ever been evaluated formally during two surveys conducted in 1996 and 2000 (Gilliam & Ripple, 2004; Ripple et al., 1999). In total, 22 state administrators reported having completed a formal evaluation of the effectiveness of their program by 1998. Of these 22, only 13 reports were identified that provided data necessary to estimate the effects of the program on child outcomes -- 4 “evaluations” simply described activities such as site visits or provided statewide financial reports, 3 had not yet completed the evaluations and released a report, and 2 did not provide child-level outcome data that could be used to estimate the impact of the program. Other efforts (e.g., systematic searches of research databases, internet search engines, and consulting reference lists from other prekindergarten reports) to identify state programs with impact evaluations were not fruitful. Complete reports were obtained for all 13 evaluations, and program evaluators were contacted directly when additional information or clarification was needed. All reports were thoroughly reviewed, specifics about the study method and findings were recorded, and standardized effect sizes (though not presented in the present paper) were computed by the present authors from data in the evaluation reports and presented. (For more details, see Gilliam and Zigler, 2001).

Additional evaluation reports generated between 1998 and 2003 were obtained directly from state prekindergarten administrators by the current authors and staff at the National Institute

for Early Education Research (NIEER) and by internet searches. In all and from 1977 to 2003, evaluations were obtained for 20 different state-funded prekindergarten systems operating in 18 states. (Two different prekindergarten systems were administered by Louisiana. Also, evaluations for the Georgia Prekindergarten Program were obtained from when the program targeted low-income children and when the program was accessible to all 4-year-olds regardless of family income.) Additionally, follow-up data were obtained from one previously reported evaluation (Michigan). These state-funded prekindergarten programs are identified in Table 1, as well as the evaluation reports that are reviewed in this paper.

Table 1
Characteristics of State-Funded Prekindergarten Programs with Impact Evaluations

State	Program Name	Site ^a	Age Range	Length of Day ^b	Min. Teacher Education ^c	Required Training	Staff:Child Ratio ^d	Reports Reviewed (See Appendix)
AR	Arkansas Better Chance	V	3-4	S	BA	Yes	1:10	33
CT	Connecticut School Readiness Initiative	V	3-4	S	CDA/AA	Yes	1:10	1
DC	Public School Prekindergarten Program	PS	4	S	BA	Yes	1:10	24-30 ^f
DE	Early Childhood Assistance Program	V	4	H	CDA	Yes	1:10	3
FL	School Readiness Program	V	3-4	S	CDA	No	1:10	17-22
GA	Georgia Prekindergarten Program	V	4	S	AA	Yes	1:10	42-46
GA-UPK	Georgia Universal Prekindergarten Program	V	4	S	AA	Yes	1:10	9
KA	At-Risk Four-Year-Old Preschool Program	PS	4	H	BA	No	NA	31
KY	Kentucky Preschool Program	V	4	H	CDA	Yes	1:10	59-64
LA-8(g)	8(g) Student Enhancement Block Grant Program	PS	4	S	BA	No	1:15	23
LA-4	Louisiana Early Childhood Development and Enrichment Activity Classes (LA4)	PS	4	S	?	?	?	2
MD	Extended Elementary Education Program	PS	4	H	BA	No	1:10	32, 68
MI	Michigan School Readiness Program	V	4	H	CDA/AA/BA	No	1:8	7, 69
NJ-Abbott	New Jersey Abbott Preschool Program	V	3-4	S	BA	Yes	2:15	47
NY-EPK ^e	New York State Experimental Prekindergarten	PS	3-4	L	BA	Yes	1:9	4-6, 10-16, 65-67
OK	Early Childhood Four-Year-Old Program	V	4	L	BA	Yes	1:10	8
SC	Half-Day Child Development Program (4K)	V	4	H	BA	Yes	1:10	48-52, 54
TX	Public School Prekindergarten	V	3-4	H	BA	No	NA	55-58
VT	Early Education Initiative	V	3-4	L	BA	Yes	1:8	53
WA	Early Childhood Education and Assistance Program	V	4	L	AA/BA	No	1:9	34-41

Note. Data were obtained from evaluation reports reviewed in this paper or from a source that most closely matches the time at which the evaluations were conducted (Barnett et al., 2004; Gilliam & Ripple, 2004; Ripple et al., 1999; Schulman, 1999). (a) V = Various locations; PS = Public schools only. (b) Lengths of day reported are the states' minimum; S = School-day (about 6 or 7 hours; some offer extended-day hours up to 8-10 per day); H = Half-day; L = Locally determined. All state programs operated for at least a full public school year (9 months). (c) BA = Bachelor's Degree; AA = Associate's Degree; CDA = Child Development Associate Credential. (d) NA = State as no mandated minimum ratio. (e) Program data for the NY-EPK at the time of the impact evaluations were conducted are not available. Data are from Barnett et al., 2004. (f) Findings for the first grade and the second cohort of kindergartners in DC were obtained directly from the evaluator (R. A. Marcon, personal communication, March 7, 1999) and differ somewhat from those reported in Marcon, 1989. Data reported in the study used a much smaller sample of children than those reported in interview with the study author, due to casewise deletion of missing data when analyzing the data multivariately.

Characteristics of Evaluated Prekindergarten Programs

Most prekindergarten programs described in this paper were administered through a state department of education, with the stated goal of promoting school readiness. These programs, however, vary significantly in terms of their structure, accessibility, duration, classroom characteristics, comprehensiveness of services, and parent involvement efforts (Barnett et al., 2004; Gilliam & Ripple, 2004; Knitzer & Page, 1998; Mitchell, 2001; Mitchell et al., 1998; Ripple et al., 1999; Schulman et al., 1999). A few of the more easily quantified characteristics of these programs (the location of classrooms, duration in years and hours per day, required level of teacher education, in-service training requirements, and minimum staff-child ratio) are presented in Table 1.

An inspection of some of the characteristics of prekindergarten programs that have been evaluated (see Table 1) as compared to data across all current state-funded prekindergarten programs (Barnett et al., 2004) suggests that the prekindergarten programs that have been evaluated may differ significantly from prekindergarten programs that have not been evaluated. Relative to state-funded prekindergarten programs that have not been evaluated, those that have are about twice as likely to require the lead teacher to possess a bachelor's degree or higher (58% versus 30%) and to require full-day or school-day services of up to 6 hours per day (45% versus 20%). They were, however, far less likely to serve 3-year olds (35% versus 59%), focusing more exclusively on 4-year olds. There appeared to be no meaningful differences in their likelihood to have staff/child ratio mandates better than 1:10 (26% versus 27%). Since evaluated programs tend to have higher teacher education requirements and longer days, evaluation results from these programs may not generalize to other state-funded prekindergarten programs.

Methodologies Used to Evaluate State Preschool Programs

Of the 20 evaluations reviewed, 11 were conducted by third-party evaluators, sometimes in collaboration with the state department of education, and the rest were conducted solely by research staff at the state agency that administers the program. Third-party evaluators often were affiliated with one of the state's public universities. Private consultants and educational research foundations also were contracted by states to evaluate their programs. Several methodological characteristics of these studies are described below and presented in Tables 2 and 3.

Table 2
Methodological Characteristics of State-Funded Prekindergarten Evaluations

State	No. of Cohorts	Pretest	Grade Levels Assessed	Contrast Group	Type of Contrast Group ^a	Sample Size ^b
AR	1	Yes	PreK	No		P = 175 (108 Center-based; 67 Home-based)
CT	3	Yes ^l	PreK-3	Yes ^k	? ^k	P = 264
DC	4	No	PreK-5	Yes	M ^f	Cohort 1 = 22 pairs; 2 = 112 pairs; 3 = 234 pairs; 4 = 202 pairs
DE	1	Yes	PreK	No		P=490
FL	3	Yes	PreK-4	Yes	E _{no}	Cohorts 1-3: P ≈ 500-700, C ≈ 400-700
GA	2	No	K-1	Yes	M ^f	Cohort 1: P = 111, C = 111; Cohort 2: P = 267, C = 267
GA-UPK	1	Yes	PreK-K	Yes	NN ⁱ	P=353; Head Start=134; Child Care=143
KA	1	No	K-1	No		P=738
KY	6	Yes	PreK-4	Yes	Cohorts 1, 2, 6 = RC Cohorts 3-5 = E ^c	Cohorts 1-6: P ≈ 120-320, C ≈ 30-200
LA-8(g)	2	Yes	PreK	No		Nearly all enrolled children (~ 1500 each cohort)
LA-4	2	Yes	PreK	No		Cohort 1=1,358; Cohort 2=3,711
MD	1	No	K; 3; 5; 8; 9, 10	Yes	RC	P = 416, C = 476
MI	1	No	K-4	Yes	E _{no} ^d	P = 351, C = 279
NJ-Abbott	1	Yes	PreK	Yes	NN	P=1,488
NY-EPK	5	Yes	PreK-6	Yes	WL ^e	Cohorts 1-5: P ≈ 1000-1900, C ≈ 57-1700 ^h
OK	2	Yes	PreK-K	Yes	E ^j	Total N = 2,396
SC	3 ^f	No	1-3	Yes	Cohort 1 = E Cohort 3 = RC	Cohort 1: P ≈ 600-721, C ≈ 2500-4700; Cohort 3: P ≈ 3000, C ≈ 4000
TX	1	No	PreK-3	Yes	E	P ≈ 1499-46,379, C ≈ 396-43,589 ^g
VT	1	No	K-2	No		P = 280
WA	3	No	PreK-7	Yes	M	Cohort 1 = 250 pairs; 2 = 156 pairs; 3 = 946 pairs

Note. (a) M = Program eligible non-attendees matched to participants on some characteristics; E = Program eligible non-attendees; E_{no} = Program eligible non-attendees who have attended no other form of preschool; NN = Participants were compared to national norms from tests; WL = Program eligible children who were placed on a wait list; RC = Random classmates that may or may not be comparable to participants. (b) P = Prekindergarten Participants; C = Comparisons. (c) In cohorts 3 through 5, Kentucky recruited a comparison group of program eligible non-attendees. Due to difficulties with attrition, however, the eligible non-attendees were replaced in each cohort as needed with random classmates beginning in kindergarten, gradually transforming the comparison group from E to RC in follow-up years. (d) Participants must have attended the program 100 of 151 program days. (e) Two types of “wait listed” children were used depending on the variable and cohort (see Gilliam & Zigler, 2001). (f) Only results for the first and third cohorts of the South Carolina evaluation could be located. (g) Exact number of comparison children depended on the specific outcome being measured. (h) The Oklahoma study took advantage of an arbitrary age eligibility cutoff to use a quasi-experimental regression-discontinuity design, described further in the paper. (i) Comparisons were also made to gains by children in Head Start and in privately obtained child care. (j) The Oklahoma UPK evaluation in Tulsa used a sophisticated quasi-experimental regression-discontinuity design described in this report. (k) A kindergarten comparison group was added to the design for cohort 3 during the 2001-2002 school year. The obtained evaluation report only includes data for cohorts 1 and 2. (l) A pretest was only used in cohort 2 for an analysis using the Bracken Basic Concepts Scale.

Table 3
Outcomes Assessed and Instruments Used in Evaluations of State-Funded Prekindergarten Programs

State	Outcomes Assessed										
	Developmental Competence	Perceived Competence	Behavior Problems	Child Health	Attendance	Grades	Achievement Tests	Drop Out	Retention	Special Ed	Parent Involvement
AR	33										
CT	5		32, 52		32				32	32	
DC	58		58			48	8				
DE	14, 16, 23, 55, 61										
FL	6, 24, 32, 62		50		50	48	7, 8, 20, 39, 53		50	50	32
GA	25, 26				32		34		32	32	32
GA-UPK	19, 21, 27, 32, 41, 43, 54		32	32	32		60				32
KA	32										
KY	3, 5, 35, 51, 52	30	52		32					32	
LA-8(g)	23										
LA-4	23										
MD	37				50		7, 36	50	50	50	
MI	15, 31, 49				50		40		50	50	32
NJ-Abbott	10, 38, 43, 54, 60		4, 12, 31, 44, 52								
NY-EPK	17, 22, 43, 59				32		47		32	32	
OK	28										
SC	18, 24						2, 39, 42		50	50	
TX							57		32	32	32
VT	32					32			32	32	
WA	11, 24, 43		13	32		32				32	32

Continued.

Instruments Used (All current version)

- | | | | |
|----|--|----|---|
| 1 | <u>Basic Skills Assessment Program</u> | 42 | <u>Palmetto Achievement Challenge Test</u> |
| 2 | <u>Battelle Developmental Inventory</u> | 43 | <u>Peabody Picture Vocabulary Test</u> |
| 3 | <u>Behavior Problems Index</u> | 44 | <u>Personal Maturity Scale</u> |
| 4 | <u>Book Handling Test</u> | 45 | <u>Preschool Developmental Inventory</u> |
| 5 | <u>Bracken Basic Concept Scale</u> | 46 | <u>Preschool Language Scale</u> |
| 6 | <u>Brigance Diagnostic Inventory</u> | 47 | <u>Pupil Evaluation Program Test</u> |
| 7 | <u>California Achievement Test</u> | 48 | Report cards |
| 8 | <u>California Preschool Social Competency Scale</u> | 49 | <u>School Readiness Rating Scale</u> |
| 9 | <u>California Test of Basic Skills</u> | 50 | School records |
| 10 | <u>CAP Early Childhood Diagnostic Instrument</u> | 51 | <u>Sentence Repetition Test</u> |
| 11 | <u>Child Adaptive/Student Behavior Inventory</u> | 52 | <u>Social Skills Rating System</u> |
| 12 | <u>Child Behavior Checklist</u> | 53 | <u>Stanford Achievement Test</u> |
| 13 | <u>Child Behavior Inventory</u> | 54 | <u>Story and Print Concepts</u> |
| 14 | <u>Child Development and Learning Checklist</u> | 55 | <u>Telamon Outcomes Assessment Database-Early Childhood</u> |
| 15 | <u>Child Development Rating</u> | 56 | <u>Test of Early Language Development</u> |
| 16 | Class Progress Chart | 57 | <u>Texas Assessment of Academic Skills</u> |
| 17 | <u>Cognitive Abilities Test</u> | 58 | <u>Vineland Adaptive Behavior Scales [Classroom Version] (Age Normed)</u> |
| 18 | <u>Cognitive Skills Assessment Battery</u> | 59 | <u>Walker Readiness Test for Disadvantaged Children</u> |
| 19 | <u>Color Bears</u> | 60 | <u>Woodcock-Johnson Test of Achievement</u> |
| 20 | <u>Comprehensive Achievement Program</u> | 61 | <u>Work Sampling Developmental Checklist</u> |
| 21 | <u>Comprehensive Test of Phonological Processing</u> | 62 | <u>Yellow Brick Road</u> |
| 22 | <u>Cooperative Preschool Inventory</u> | | |
| 23 | <u>Creative Curriculum Checklist</u> | | |
| 24 | <u>Developmental Indicators for the Assessment of Learning-Revised</u> | | |
| 25 | <u>Developmental Profiles-II</u> | | |
| 26 | <u>Developmental Rating Scale</u> | | |
| 27 | <u>Draw a Person MSCD subtest</u> | | |
| 28 | <u>Early Childhood Skills Inventory</u> | | |
| 29 | <u>Early Screening Inventory</u> | | |
| 30 | <u>Harter Scale of Perceived Self-Competence</u> | | |
| 31 | <u>High/Scope Child Observation Record</u> | | |
| 32 | Homemade surveys, interviews, or tests | | |
| 33 | <u>Individual Developmental Early Assessment</u> | | |
| 34 | <u>Iowa Test of Basic Skills</u> | | |
| 35 | <u>Letter Recognition Test</u> | | |
| 36 | <u>Maryland Functional Test</u> | | |
| 37 | <u>Maryland Systematic Teacher Observation</u> | | |
| 38 | <u>McCarthy Scales of Children's Abilities</u> | | |
| 39 | <u>Metropolitan Achievement Test</u> | | |
| 40 | <u>Michigan Educational Assessment Program</u> | | |
| 41 | <u>OWLS</u> | | |

Number of Cohorts, Length of Follow-up and Number of Subjects

Most studies evaluated multiple cohorts of children. Kentucky had evaluated the most, with six different cohorts. Most states followed participants prospectively beginning in their prekindergarten year. Many, however, recruited children once they enrolled in kindergarten, when a comparison group could be located most easily. The median length of follow-up was second grade. Maryland followed children until their tenth grade, the farthest of the state evaluations. The number of subjects varied significantly by state, since some evaluations relied on individual assessment of representative samples, whereas others relied mostly on school-collected data that existed for all students. To recruit samples, evaluators typically selected school districts or buildings to represent various regions of the state and then randomly selected subjects at the classroom level. This created nested designs, but the analyses never accounted for this feature. Most study attrition rates ranged from less than 10% to about 25% per year, rather similar to the 20% rate typical for evaluations of programs serving at-risk families (Gomby, 1999).

Comparison Groups and Study Designs

No state evaluation randomly assigned children to program and control groups, and therefore all resorted to some less rigorous comparison group. All but 6 of the state evaluations used some form of comparison group, against which program impacts were estimated.

Evaluation Designs for Targeted Prekindergarten Programs. Of the evaluations of the 20 programs identified in Table 2, all but 2 were of programs that limit access to children who are from low-income families or have some other risk status for educational failure. These 18 evaluations of targeted prekindergarten programs used a variety of designs and comparison groups. All 6 of the evaluations without comparison groups (Arkansas, Delaware, Kansas,

Louisiana 8(g), Louisiana LA4, and Vermont), are severely limited methodologically. These designs used a single-group pretest-posttest (and in 2 cases no pretest was given) with non-nationally normed instruments in an attempt to demonstrate positive growth associated with participation in the program. In all 6 of these evaluations, children were determined to have improved in their skills, but, unfortunately, there is no way to attribute any of the growth to the effects of prekindergarten participation as opposed to simple maturation. In some cases, posttest results of prekindergarten participants were compared to some valued benchmarks for the program or the schools. This can be useful in order to determine whether children are leaving the program with the desired skill levels or to target services, but the methodology simply cannot support even the most tenuous of causal inferences. Several of these 6 evaluations were plagued by other methodological flaws. These flaws include the use of tests that have not been validated or normed, having the prekindergarten teacher administering the pretest and posttest protocol on their own children, and requesting kindergarten teachers who are rating participants' skills to choose a classmate by which comparisons will be made – the last two introducing the potential for significant bias in the ratings. Because of these extreme limitations, these 6 evaluations were excluded from further description and presentation of findings.

In the 12 remaining evaluations of targeted prekindergarten programs, at least one of five different types of comparison groups were used typically: waitlisted comparisons, matched program-eligible non-attendees, non-matched program-eligible non-attendees, program-eligible non-attendees who did not attend any other preschool program, and random classmates that may or may not have been eligible for the program. Some evaluations used multiple types of comparison groups. Of course the type of comparison group used has a great effect on the interpretation and validity of the findings.

Arguably, of the comparison groups used, the waitlist comparison provided the best test of the program, since comparison children and their families were both eligible for services and motivated to apply for the program. Unfortunately, the only program to employ this comparison was the evaluation of the New York State Experimental Prekindergarten Program, which was by far the most outdated state evaluation (University of the State of New York & New York State Education Department, 1977). Use of the three other comparison groups of program-eligible non-attendees may have introduced a self-selection bias to the results, since families of comparison children were not motivated to seek placement in the program being evaluated. The matched program-eligible non-attendee comparison attempts to methodologically control some of this bias by matching the two groups on related variables (e.g., gender; ethnicity; parent education and occupational level; and some type of proxy for family income, such as eligibility for free or reduced lunch at school). The program-eligible non-attendee comparison is less rigorous, since comparison children, though eligible for the program, may have significantly differed from participants in many ways. The comparison group that arguably provides the least stringent test of the program is the one that used program-eligible non-attendees who did not attend any other preschool program, since the families of comparison children were not motivated to seek preschool programming for their child. Furthermore, in Michigan the participants were required to have attended the program at least two thirds of the total number of program days (100 of 150 days). Stipulating that the participant children must have attended the program at least minimally and that the comparison children could not have participated in a similar program may bias studies toward finding positive results. However, one could also argue that it is unfair to test the effectiveness of a program by using participants who have not

participated in at least some minimal way or by comparing outcomes to children who may have attended similar programs (Gilliam, Ripple, Zigler, & Leiter, 2000).

Three evaluations used random elementary school classmates as comparisons. Despite this method of selecting comparisons, the two groups in Maryland were comparable in ethnicity, age, gender, family composition, and father's educational and occupational level, but differences in maternal educational and occupational level slightly favored preschool participants. In Kentucky and South Carolina, random classmates comparisons were used only for certain cohorts or at certain grade levels. Additionally, in 3 of Kentucky's 6 cohorts a program-eligible non-attendeer comparison group was initially recruited, but then changed to random classmates over the course of yearly follow-ups as a result of serious attrition rates in the comparison group. This practice, of course, makes it difficult to determine the nature of the comparison group in the follow-up years and virtually precludes a longitudinal treatment of the data.

Evaluation Designs for Universal Prekindergarten Programs. The two evaluations of universal prekindergarten programs – the Georgia and Oklahoma studies – deviated from the above types of contrast. Universal programs create a special challenge in terms of locating suitable comparison groups. Because the saturation rates of these universal programs are relatively high – of all state-funded prekindergarten programs, Oklahoma (55.6%) and Georgia (53.4%) serve the highest percent of their states' 4-year olds (Barnett et al., 2004) – the comparison pool is relatively small, and many of the children not served by these programs may be served by other preschool programs operating in the state.

The Oklahoma Early Childhood 4-Year-Old Program evaluation (Gormley & Gayer, 2003), conducted in the Tulsa Public Schools (TPS), used a very sophisticated design and analysis to deal with this problem. The TPS use an age cutoff to determine eligibility to enter the

program – only children with a birthday earlier than September 1 can enroll in the program, and children born after that date must wait one year. Taking advantage of this arbitrary entry criterion, the evaluators designed a quasi-experimental regression-discontinuity study that tested differences in predicted probabilities of test scores based on a bilateral quadratic parametric fit. This design allowed for a comparison of children who barely met the cutoff versus those that barely missed it, and preliminary analyses verified the demographic comparability of the groups. Given the difficulty of identifying and recruiting a suitable comparison group for an evaluation of a universally accessible program, this design provided an excellent means for estimating program impacts.

Rather than seek a method for providing a true counterfactual for participation in the prekindergarten program, the Georgia Universal Prekindergarten Program evaluation (Henry, Henderson, Ponder, Gordon, Mashburn, & Rickman, 2003) focused on contrasting the prekindergarten children with two samples of children served in Head Start and in privately obtained community-based preschool programs. Overall, two different types of analyses were employed to contrast these three groups: (a) comparing posttest scores between these three groups at the end of the preschool year and (b) comparing pretest to posttest gain scores during the preschool year and into the beginning of kindergarten. Also, by using several tests that have national norms the evaluators were able to benchmark children's performance in all three groups against national norms for several of the child outcomes being assessed.

Using the first analytic strategy described above, the Georgia evaluators attempt to document program effects by comparing at the end of the preschool year and the beginning of the kindergarten year child outcome scores for prekindergarten participants to those for children who attended Head Start and children who attended a private community-based child care

centers. At the end of the prekindergarten year and at the beginning of kindergarten, Georgia UPK children outscored Georgia Head Start students across most measures. However, children who attended private community-based child care centers typically outscored both groups. Unfortunately, simply comparing child scores at the end of a program says nothing about the degree to which these exit scores were the result of the intervention, and contrasting three service options (Georgia UPK, Head Start, and privately-obtained child care) results in a serious underestimation of the effects of the Georgia UPK because of the lack of a no-treatment comparison (Abbott-Shim & Lambert, 2003).

Since the posttest-only comparisons described above do not provide the data necessary to estimate changes due to the intervention, the evaluators used gain scores to attempt to document program impacts. Across the several measures used in the Georgia UPK evaluation, only four measures for which analyses were conducted have national normative data – the Peabody Picture Vocabulary Test – III (PPVT-III), Letter-Word Identification Subtest from the Woodcock-Johnson – III Tests of Achievement (WJ-III), Applied Problems Subtest of the WJ-III, and the Oral and Written Language Scales (OWLS). By using the national norms built into these tests, the evaluators were able to assess whether prekindergarten participants showed a gain in these areas during the course of the prekindergarten program that exceeded normative expectations. Statistically significant pretest to posttest differences were reported for the Georgia UPK participants across all four measures: pretest to the end of prekindergarten on the PPVT-III, Letter-Word Identification, and the OWLS, and from Pretest to kindergarten entry on the PPVT-III, Letter-Word Identification, and Applied Problems (Henry et al., 2003, Table 5.1, page 45). Although these differences were labeled “statistically significant,” the report does not indicate the type of analyses used, the test statistics, or the level of significance. Further, since standard

deviations were not reported anywhere in the report (Abbott-Shim & Lambert, 2003), it was not possible to determine accurately the standardized effect size. Assuming that the sample standard deviations are the same as the test standard deviations for the instruments used, standardized effect sizes⁶ ranged from 0.13 on Letter-Word Identification to 0.29 on the PPVT-III, both at the end of the prekindergarten year. Although apparently no statistical tests were conducted comparing the Georgia UPK gains to those in either Head Start or privately obtained child care, gain scores across these three groups appeared rather similar. One exception was a +7.04 point pretest to kindergarten entry gain on the PPVT-III for the Head Start sample, which results in an estimated effect size of 0.47 (assuming that the sample standard deviations approximate the test's normative standard deviation).

Significant pretest to posttest gains on non-normed measures were reported in the Georgia UPK evaluation, but it is impossible to know the degree to which these differences reflect simple maturation unrelated to prekindergarten participation. Since no significant differences on these non-normed tests were noted between Georgia UPK participants and participants in either Head Start or privately obtained child care, they are omitted from this review. This paper, therefore, only reports the analyses of gain scores from nationally normed measures used in the Georgia UPK evaluation.

Domains of Outcomes Assessed and Instruments Used

As in Gilliam & Zigler (2001), outcomes were categorized in one of 11 domains, and most states tracked outcomes in more than one domain (see Table 3). These outcome domains are indicated in Table 3, along with the tests or other measures used to assess these domains. A large number of tests and procedures were used in these 20 evaluations. Several of them are well-known, psychometrically valid instruments. In many cases, however, relatively unknown

tests were used, with little data provided in the evaluation reports regarding their reliability and validity.

Findings from State Prekindergarten Evaluations

Tables 4 and 5 indicate where statistically significant ($p < .05$)⁷ findings were found or not found, organized by domain of outcome, grade level and state. Null (denoted “0”), as well as negative (denoted “-”), findings are also noted. In many cases, findings from more than one cohort are represented in a single cell, starting with the earliest cohort. Only findings from the 14 state evaluations that used a comparison group of some kind and were able to conduct analyses that evaluate programmatic impact are presented. The six state evaluations that did not use a comparison group were described earlier in this paper. The findings presented in the tables and described below should be interpreted cautiously, bearing in mind the methodological limitations described throughout this paper.

Table 4
Statistically Significant Impacts of State-Funded Prekindergarten Programs Through Grade 4

	End PreK	K	1	2	3	4
<u>OVERALL DEVELOPMENT</u>						
CT	0(NN) ^k					
DC		+/0	+/0			
FL	+/+	+				
GA			0			
KY	+/+/+	+/0/0	0/0	0		
MD		+				
MI		+&+ ^h	+	+	+	+
NJ-Abbott	+(NN & FACES)					
NY-EPK	+&+/&+ ^h	+&0 ^h			0	
OK		+				
SC			+			
WA		NR	NR	NR		
<u>PERCEIVED COMPETENCE</u>						
KY	+	0	0			
<u>BEHAVIOR PROBLEMS</u>						
FL						+
KY	0/+0 ⁱ	0/0/0/0	0/0/0/-	-/0/0/0	-/0	0
NJ-Abbott	+(FACES)					
WA			0 ^a	0 ^a	0 ^a	
<u>CHILD HEALTH</u>						
WA			0 ^a	0 ^a	0 ^a	
<u>ATTENDANCE</u>						
FL		0/+	0/+	0/0	0/+	
GA		+	0/+			
KY		0 ^a /0 ^a	0 ^a /0 ^a	0 ^a /0 ^a	0 ^a /0 ^a	0 ^a /0 ^a
MI		+	+	+	+	
NY		+	+	+	+	+
<u>GRADES</u>						
DC (Reading)		+/0	0/0		0	0
DC (Math)		+/0	0/0		0	0
FL (Reading)		0	0	0	-	0
FL (Math)		0	0	0	-	0
WA (Reading)			0 ^a	0 ^a	0 ^a	
WA (Math)			+ ^a	0 ^a	0 ^a	

Continued.

	End PreK	K	1	2	3	4
<u>ACHIEVEMENT TESTS</u>						
DC (Reading)					0	
DC (Math)					0	
FL (Reading)		+	-/0	0/0	0/0	-
FL (Math)		+	0/0	0/0	0/0	-
GA (Reading)			+			
GA (Math)			+			
MD (Reading)					+	
MD (Math)					+	
MI (Reading)						+
MI (Math)						+
NY (Reading)					0	
NY (Math)					+	
SC (Reading)			+ ^b /0 ^c	0/0 ^c	+/0 ^c	
SC (Math)			+ ^b /0 ^c	0/0 ^c	+/0 ^c	
TX (Reading)					+	
TX (Math)					+	
<u>RETENTION</u>						
FL		+	0	-	0/.20 ^d	0
GA		+	0/0			
MD ^e					+	
MI ^e			+	+	+	+
NY		0	+	0	0	0
SC			+	0	- ^a	
TX				+		
<u>SPECIAL ED REFERRAL</u>						
GA		0	0/0			
MI		0	0	0	0	0
TX				+		
WA		0 ^a				
<u>SPECIAL ED PLACEMENT OR ACADEMIC ASSISTANCE</u>						
FL		0/0	0/0	0/0	0/0	0
MI		0	0	0	0	0
NY		0	0	0	0	0
SC			+	+	0	
TX				+		
WA		0 ^a				

Continued.

End PreK	K	1	2	3	4
PARENT INVOLVEMENT					
GA	0	0/0			
MI ^j	+	+	+		
TX		+			
WA ^g		0 ^a	0 ^a	0 ^a	

Note. A “+” indicates a statistically significant ($p < .05$) finding in favor of the prekindergarten participants, “0” indicates no significant difference between prekindergarten participants and comparison children; “-” indicates a statistically significant difference ($p < .05$) in favor of comparison children; “NR” indicates that not enough data could be obtained to determine statistical significance; FACES = Data from the Head Start Family and Child Evaluation Study (FACES); CPP = Community-based Preschool Program; P = Prekindergarten program being evaluated; HS = Head Start. An ampersand (&) is used to indicate that more than one measure of the same outcome was used in the same cohort, and a slash (/) is used to divide findings from different cohorts in the same evaluation.

- (a) Significance level was computed by this author using a Z -test of data presented in each state’s reports.
- (b) Comparison group consisted of program eligible non-attendees.
- (c) Comparison group consisted of random classmates who may or may not have been eligible for the program.
- (d) The positive finding was obtained in a second cohort of students who were analyzed for a grade retention that occurred anywhere from grade K to 3.
- (e) Maryland and Michigan data were analyzed cumulatively (e.g., at least one grade retention by third grade, by fifth grade, etc.).
- (f) Data were analyzed cumulatively (e.g., placement in special education by fifth grade).
- (g) Parent involvement was assessed by both parent and teacher report. In both cases results indicated no significant differences at all grade levels.
- (h) Multiple instruments were used to measure the same construct for a single cohort.
- (i) Effects in all 3 cohorts were based on parent reported behavior problems pretest to posttest.
- (j) Data were combined across grades kindergarten through two prior to analyses. Therefore, grade specific effects cannot be reported.
- (k) The evaluator reported this finding to have been significant at the $p < .10$ level (specifically, $p = .086$). Following the $p < .05$ convention for statistical significance used throughout this paper, we report this finding as being statistically non-significant.

Table 5
Statistically Significant Impacts in Various Subdomains of Developmental Competence Through Grade 5

	End PreK	K	1	2	3	4	5
<u>SOCIAL</u>							
DC		0/0	0/0				0
GA		+	0/0				
KY	+ / + / +	0/0/0/0/0	0/0/0/-	- / 0/0	- / 0		0
MI		+					
NJ-Abbott	+ (FACES)						
OK		0					
<u>SELF-HELP</u>							
DC		0/0	+ / 0				0
GA		+					
KY	+ / 0	+ / 0/0	0/0	0			
MI		+	+	+	+		
<u>MOTOR</u>							
DC (Overall)		0/0					
GA (Overall)		+					
KY (Gross)	0/0	0/0/0	0/0	0			
KY (Fine)	+ / +	0/0/0	0/0	0			
OK		0					
<u>LANGUAGE</u>							
DC (Overall)		+ / 0	+ / 0				0
GA (Overall)		+					
GA-UPK (Receptive)	+ (NN)	+ (NN)					
GA-UPK (Expressive)	+ (NN)						
KY (Receptive)	0/0	0/0/0	+ / 0	0			
KY (Expressive)	+ / 0	0/0/0	0/0	0			
MI (Overall)		0					
NY (Receptive)	0 / +	0			0		
OK		+					
WA (Receptive)	+ / + / +						

Continued.

	End PreK	K	1	2	3	4	5
<u>COGNITIVE</u>							
KY	+/0	0/0/0	0/0	0			
NJ-Abbott	+(FACES)						
NY					0		
OK ^a		+					
SC			+/-				
<u>LITERACY/NUMERARY</u>							
GA		+	+				
GA-UPK (Literacy)	+(NN)	+(NN)					
GA-UPK (Numerary)	0(NN)	+(NN)					
KY (Academic)		+/0/0/0	+/0/0	-/0	0		
KY (Literacy)	0/0	0/0	+/0				
MI		0	0	0	0	+	
NJ-Abbott	+(FACES) & -(NN)						

Note. A “+” indicates a statistically significant ($p < .05$) finding in favor of the prekindergarten participants, “0” indicates no significant difference between prekindergarten participants and comparison children; “-” indicates a statistically significant difference ($p < .05$) in favor of comparison children; “NR” indicates that not enough data could be obtained to determine statistical significance; FACES = Data from the Head Start Family and Child Evaluation Study (FACES); CPP = Community-based Preschool Program; P = Prekindergarten program being evaluated; HS = Head Start; NN = National norms from test. An ampersand (&) is used to indicate that more than one measure of the same outcome was used in the same cohort, and a slash (/) is used to divide findings from different cohorts in the same evaluation. (a) This test in Oklahoma measure cognitive and general knowledge skills combined.

Developmental Competence

Thirteen of the 14 state evaluations (all except Texas) presented in Tables 4 and 5 gathered at least some data on children's developmental competence at the end of the prekindergarten program or during early elementary school. Evaluation results are reported both for tests that address specific subdomains of developmental competence (social, self-help, motor, language, cognitive, and academic or literacy skills) and tests that provide an overall developmental score that combines two or more specific subdomains of development. Impacts on overall developmental scores are presented in Table 4, and impacts for specific subdomains are presented in Table 5.

In general, positive effects in overall developmental competence were commonly found. In all instances, except Connecticut, significant positive impacts were reported at the end of prekindergarten. Additionally, in all state evaluations where impacts were reported at the end of the prekindergarten year, positive effects were sustained to kindergarten for at least one cohort. Significant effects were inconsistent at first grade and significant effects beyond that point were only reported in the Michigan evaluation.

Data presented in Table 5 suggest that the positive impacts in overall developmental competence may be attributed to short-term impacts across a variety of developmental subdomains. At the end of prekindergarten and into kindergarten, at least one significant impact was found in almost every subdomain assessed. Significant effects were most commonly found in the domains of development in social, self-help, language, and literacy and numerary skills. Positive prekindergarten impacts on motor skills were rarely reported. Kentucky was the only evaluation that divided motor skills into gross motor (locomotion, balance, and large muscle skills) versus fine motor (eye-hand coordination and small muscle skills), and found impacts

only in the area of fine motor development. Several positive findings were reported in the area of language development, and no pattern of differential reported effects was found between receptive (ability to comprehend language) versus expressive language (ability to verbally express oneself). By first grade, significant effects were rare, but, when reported, most frequently occurred in the language and literacy/numerary domains.

Child Perceived Self-Competence

Kentucky was the only state to collect data on how preschool participants perceived themselves after participating in the program. At the end of the prekindergarten year, participants perceived themselves to be significantly more competent in the cognitive domain, relative to program eligible non-attendees. No significant differences, however, were observed in kindergarten or first grade, when participants were compared to random classmates.

Behavior Problems

Five states evaluated program impacts on children's behavior problems. Neither Kentucky, New Jersey-Abbott, nor Washington reported significant impacts beyond preschool using teacher rating scales, and DC (not indicated in the table) found no significant impacts at fifth grade. Florida, however, did report a significant impact as late as fourth grade. In contrast to other states, Florida relied on actual reported incidents of corporal punishment, in-school and out-of-school suspensions, and expulsions, as reported in school records. Combining data over four Florida counties, eligible non-attendees with no preschool experience were significantly more likely than participants to have been disciplined during the school year (32% versus 11%).

Child Health

Washington was the only state to collect data on this variable. (The Georgia-UPK evaluation also collected data on this variable, but the lack of a matched comparison group

precludes estimating program impacts.) In Washington, a 12-item questionnaire was used to ask parents about the health of their child. Z -tests computed by the present authors using reported data indicated non-significant differences between preschool participants and matched comparisons at first, second, and third grades ($p = .83$; $p = 1.00$; $p = .99$, respectively).

Attendance

With the exception of Kentucky, all states evaluating this outcome found significant impacts in one evaluation cohort. Furthermore, these effects persisted well beyond school entry. In addition to the effects shown in Table 5, New York found statistically significant impacts at fifth and sixth grades, and Maryland reported a sizable positive impact at tenth grade. The Kentucky analyses, in contrast, compared participants to random classmates who were known not to be comparable to the treatment group. The Kentucky evaluators interpreted this lack of significant difference as being indicative of a positive effect, since participants performed similarly to their less at-risk classmates. However, the Kentucky analyses only indicated that there was no significant difference between these two groups, not that the two groups were significantly equal (See Gilliam & Zigler, 2001 for more details on the analytic problems associated with attempts to equate groups on the basis of null significant differences.)

Grades

Although report card grades for a variety of subjects were reported by different evaluations, only grades in reading and math are reviewed here. At the younger grade levels, grades for subjects such as “verbal skills” were considered to be relatively synonymous to reading. Statistically significant impacts in DC were only found in kindergarten, and then only in the first of two cohorts. The second DC kindergarten cohort also found non-trivial standardized effects in both reading and math (Gilliam & Zigler, 2001), but sample sizes were too small ($n =$

47 pairs) to afford enough power to detect statistically significant difference. Washington found a significant impact only for math and only at the first grade. Statistically significant positive impacts were not reported for Florida from kindergarten through fourth grade. Interestingly, both DC and Washington found effects using a matched comparison group, whereas Florida, which did not find effects, relied on a comparison group that was not matched to participants.

School-Administered Academic Achievement Tests

Similar to report card grades, academic achievement test scores were available for a variety of domains and sub-domains. Again, for the sake of parsimonious presentation, only overall reading and overall mathematics test score findings are presented here. With the exception of DC's, all eight evaluations addressing this outcome reported statistically significant impacts on academic achievement tests occurring at one or more grade levels. Standardized effect sizes, however, in some cases were relatively low (Gilliam & Zigler, 2001). The relative consistency of statistically significant findings for this outcome may have been due, at least in part, to the large number of subjects involved in many of these analyses. For example, as previously reported in Gilliam & Zigler (2001) statistically significant findings were reported in both South Carolina's first grade and Texas' third grade, with standardized effect sizes ranging only from .07 to .09.² In both states, however, sample sizes were several thousand large. Conversely, DC reported no statistically significant impacts in third grade scores, despite standardized effect sizes about three times as large (.23). DC's sample size for this analyses, however, was only 29 matched pairs, resulting in weak statistical power.

Michigan reported a statistically significant effect in both literacy and math at fourth grade – 24% more prekindergarten participants passed the Michigan Educational Assessment Program literacy test and 16% more passed the mathematics test. Maryland and New York-EPK

evaluated impacts in this domain beyond fourth grade, finding statistically significant positive impacts at every grade level assessed by both states. Maryland found significant impacts in both reading and math in fifth, eighth, ninth, and tenth (for math only) grades. New York-EPK found significant impacts in both reading and math in sixth grade.

School Drop Out

The only state evaluation to collect data long enough after prekindergarten attendance to be able to address this important issue was Maryland. By tenth grade, 8.2% of the prekindergarten participants already had dropped out of school, compared to 11.3% of comparison children. Although the difference was not statistically significant, the standardized effect size approached a degree of meaningfulness (.18), as defined by Cohen.² Since drop out was not assessed beyond tenth grade, it is unknown whether the effect would have been significant by the end of what should have been each student's twelfth grade.

Retention Rates

In many ways positive impacts on this outcome may be one of the most robust findings for state programs, since all seven state prekindergarten programs for which this outcome was evaluated found a statistically significant impact at one or more grade levels, and effect sizes were almost always non-trivial by Cohen's conventions (as reported in Gilliam & Zigler, 2001).

All evaluations except Maryland, Michigan, and the second cohort of Florida were analyzed non-cumulatively.⁸ When data were analyzed non-cumulatively, positive impacts were always found in each state's evaluation at one grade level (either kindergarten, first grade, second grade, etc.). Additionally, when Maryland, Michigan, and Florida analyzed the data cumulatively significant positive impacts were always found. Therefore, the mixed findings at specific grade levels when analyzed non-cumulatively may say more about each state's practices

regarding when to retain children and less about the apparent robustness of the findings.

Maryland also evaluated cumulative retention rates beyond fourth grade, finding significant impacts at fifth, eighth, and tenth grades. Across these three states that reported cumulative retention rates, prekindergarten participants were found to be 31% (Maryland at grade 10) to 44% (Maryland at grade 5 and Michigan at grade 3) less likely to have been retained for at least one grade level.

Special Education Referral and Placement Rates

Overall, few significant differences were reported for this outcome – a stark contrast to evaluations of model preschool programs, which frequently find impacts in this area. Across 8 state prekindergarten evaluations, only 3 (Maryland, South Carolina, and Texas) reported any significant differences on either referral or placement rates. Maryland was the only state to examine this variable cumulatively. By fifth grade, only 13% of Maryland preschool participants reportedly had been placed in special education services at some point in their schooling, as compared to 24% of comparisons ($p < .01$). As was the case with grade retention data, cumulative analyses of this variable may be more likely to yield significant positive impacts than dividing the effects over successive grade levels in non-cumulative analyses.

Parent Involvement

Only four states collected data on parental involvement in their child's subsequent elementary education. The Michigan evaluation combined parent reports of involvement across kindergarten through grade two into one analysis, which showed that parents of prekindergarten participants reported greater involvement in child-related school activities ($p < .05$), adult-related school activities ($p < .01$), and teacher-parent communications ($p < .05$). No significant differences were found in parental expectations or home-based educational activities.

Additionally, Texas reported a statistically significant positive impact on parental involvement in school, and Georgia (with the same .15 standardized effect size as Texas on this variable, but a far smaller sample size) found no statistically significant effect.

Summary and Discussion of Prekindergarten Findings

In this section, findings from studies of the effects of state-funded prekindergarten are summarized, and these findings are placed within the context of other early childhood education data from model programs and Head Start.

Results of State Prekindergarten Evaluations Across 18 States

Several positive impacts were reported in a variety of outcome domains. Effects were found in several areas of developmental competence and achievement tests, extending into kindergarten and sometimes beyond. Both grade retention and special education are costly educational interventions for schools, and the potential for early education programs to reduce these expensive interventions has been documented in model early education interventions (Barnett, 1995; Karoly et al., 1998; Reynolds & Temple, in press). Reduced grade retention appeared to be a rather robust impact for state prekindergarten programs, with cumulative effects that may last well beyond elementary and middle school. Given that grade retention is often associated with a variety of poor academic outcomes for children (Heubert & Hauser, 1999) and is a highly expensive response to poor school achievement,⁹ this is a particularly noteworthy finding. Surprisingly, positive impacts were seldom observed in the domains of special education referral and placement, another expensive educational intervention.

The potential for preschool programs to serve as a preventive for later delinquency has attracted attention (Yoshikawa, 1995; Zigler, Taussig, & Black, 1992), and the general lack of positive impacts in this area for state prekindergarten programs is interesting. One possible

explanation is that behavior-rating scales may not be an adequate method for documenting this outcome. Indeed, as previously reported, two of the studies largely responsible for current interest in this area (Lally, Mangione, & Honig, 1988; Schweinhart, Barnes, & Weikart, 1993) relied on actual reports of subsequent delinquent behaviors and their disciplinary results, rather than more subjective teacher or caregiver ratings. Florida, the only state to find significant behavioral impacts beyond the program year, was the only state evaluation that used actual disciplinary records for behavioral problems, as opposed to teacher or caregiver ratings of child behavior.

Few positive effects were reported beyond first or second grade, and there were rarely statistically significant impacts *not* found at early grades, only to emerge later. When significant effects beyond second grade were reported, these were mostly limited to outcomes that related to the children's actual performance in school, rather than individually administered tests or rating scales. These impacts included reduced grade retention (observed in all eight evaluations that assessed this), school-administered educational achievement tests, school attendance rates, and behavioral problems that resulted in a school disciplinary action. It is possible that these prekindergarten programs, by helping children to enter kindergarten and first grade more ready for a classroom-based school experience, have set children on an educational pathway more conducive to improved classroom behavior, motivation, and academic performance – outcomes that might be measured better by these school-collected variables than individually administered test scores or ratings. Overall, the pattern of prekindergarten findings presented here more closely resembles the effects of other large-scale programs, like Head Start (McKey, Condelli, Ganson, Barrett, McConkey, & Plantz, 1985), rather than the relatively more impressive impacts of smaller-scale model programs.

Evaluations of Universal Prekindergarten Programs

Most of the evaluations reviewed in this paper were for state-funded prekindergarten programs that target low-income or otherwise at-risk preschoolers. The two notable exceptions are the evaluation of Georgia's UPK (Henry et al., 2003) and Oklahoma's UPK, as implemented by the Tulsa Public Schools (Gormley & Gayer, 2003). Due to the universal access of the Georgia and Oklahoma UPKs and the high proportion of 4-year olds served by these programs in these states (53% and 56%, respectively, Barnett et al., 2004), these evaluations are particularly noteworthy. Evaluating a universally accessible program presents major methodological challenges, since adequate comparison groups may be hard to locate.

The two UPK evaluations used very different approaches to dealing with the problem above. In Georgia, gain scores for UPK participants were compared to national norms on the tests used, an approach used in the Head Start FACES evaluation (Zill, Resnick, Kim, O'Donnell, Sorongon, McKey, et al., 2003) and to contrast groups of children who have attended either Head Start or privately obtained child care. The use of these national norms to serve as a sort of control group, though common in this type of research, has several limitations, with the most noteworthy being that there is little evidence to suggest that the national norms adequately reflect the learning trajectory of the particular students in the program's population pool that choose to participate in the program. This is a major problem for evaluations of programs that serve at-risk populations (e.g., Head Start and most of the state prekindergarten programs that target services to low-income children), because the at-risk populations served by these programs vary in significant ways from the national norms used in these tests – norms that were derived from a national sample of children across all socioeconomic levels. Though the use of these

national norms in evaluating universally accessible programs is not ideal, the population served by universally accessible prekindergarten programs may match these norms more closely.

The Oklahoma UPK evaluation, in contrast to the Georgia UPK evaluation, utilized the arbitrary birthday cutoff that the program uses (September 1) to analyze differences in children who were able to attend the program because their birthday fell before September 1 versus children who did not attend because their birthday fell after that date, and controlling for any age differences between the two groups. The Oklahoma design and analytic strategy allowed the evaluators to provide well-supported findings about the impact of the program, without either denying services to a control group or using national normative data that may not reflect adequately the sample being studied. This approach holds significant promise for evaluating UPK programs. Overall, both the Oklahoma and Georgia UPK evaluations provide findings that support the positive effects of universally accessible prekindergarten.

Limitations

As was previously found by Gilliam and Zigler (2001), these evaluation efforts ranged from simple to highly complicated. Overall, the 69 reports that document the 20 prekindergarten evaluations reviewed in this paper provide findings that may be used to build a growing body of evidence for prekindergarten effectiveness in certain outcome domains. However, as found in the original paper, there are many methodological weaknesses in these evaluations, and some are so flawed as to weaken significantly any ability to draw solid conclusions about program effects. These methodologically fatal flaws include practices that build in a strong evaluator bias (e.g., relying on prekindergarten teacher reports in a pretest-posttest design), as well as basing results solely on single-group pretest-posttest analyses without attempting to control for maturation and other factors. Gilliam and Zigler (2001) provide several recommendations for prekindergarten

program evaluation, including choosing appropriate and realistic outcomes and validated measures, obtaining the most comparable contrast group possible, using process evaluations to guide and contextualize outcome evaluations, and other suggestions that may help guide program evaluations in this area. These recommendations are again offered, and the interested reader is referred to the original paper for more details and suggestions.

Conclusions

The 20 evaluations of prekindergarten programs reviewed in this paper represent our best current knowledge of the child level impacts of these programs as they were implemented in 18 different states. In some states, the evaluation methods are too weak to estimate the effectiveness of these programs. In several other states, the studies offer some modest indication of the likely benefits of these programs, especially in terms of short-term developmental gains and longer-term indicators of educational progress (e.g., reduced grade retention, better school attendance, and improved school-administered test scores). In a few instances, the studies provide fairly solid data on the effects of these programs, using sound quasi-experimental evaluation methods to control for baseline differences between participants and contrasts. As the field of prekindergarten science grows, more and better quality research will be needed to assess the overall impacts of these programs.

Despite a significant degree of within-state variability in these state-funded prekindergarten programs (e.g., length of day, teacher education and credentials, classroom setting, types of children and families served, etc.), these evaluations almost never attempt to address fundamental questions regarding what types of prekindergarten services work best and under what conditions of implementation. Should prekindergarten programs serve both 3-year olds and 4-year olds, or just the latter? Is a half-day program enough? What level of teacher

education, training and credentials are necessary for beneficial impacts? What are the benefits of a prekindergarten system based only in the public schools, versus one that attempts to draw on the strengths of other systems, such as Head Start or community-based providers? What children benefit most from prekindergarten? It is the answers to these questions – rather than the simplistic “Does prekindergarten work?” – that will provide the early education field with the kind of data needed to inform sound policy formation. The extant literature addressing these important policy-relevant questions, comprised mostly of data from either high-quality model early childhood education programs or the broader field of child care, may not apply to prekindergarten programs implemented within the fiscal and monitoring constraints of state budgets. Thoughtful evaluations that address important policy-relevant issues about how prekindergarten is most effectively and efficiently implemented are needed to help guide prekindergarten policy development and to inform us of our successes – as well as our needs.

References

Abbott-Shim, M., & Lambert, R. (2003). Review of the “Report of the Findings from the Early Childhood Study.” Unpublished manuscript, Georgia State University and University of North Carolina at Charlotte.

Barnett, W. S. (1995). Long-term effects of early childhood programs on cognitive and school outcomes. The Future of Children, 5(3), 25-50.

Barnett, W. S. (1998). Long-term effects on cognitive development and school success. In W. S. Barnett & S. S. Boocock (Eds.), Early care and education for children in poverty (pp. 11-44). Albany, NY: State University of New York Press.

Barnett, W. S., Robin, K. B., Hustedt, J. T., & Schulman, K. L. (2004). The state of preschool: 2003 state preschool yearbook. New Brunswick, NJ: National Institute for Early Education Research.

Gilliam, W. S., & Ripple, C. H. (2004). What can be learned from state-funded preschool initiatives?: A data-based approach to the Head Start devolution debate. In E. Zigler & S. J. Styfco (Eds.), The Head Start debates (friendly and otherwise). Baltimore, MD: Paul H. Brookes Publishing.

Gilliam, W. S., Ripple, C. H., Zigler, E. F., & Leiter, V. (2000). Evaluating child and family demonstration initiatives: Lessons from the Comprehensive Child Development Program. Early Childhood Research Quarterly, 15, 41-59.

Gilliam, W. S., & Zigler, E. 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.

Gomby, D. S. (1999). Understanding evaluations of home visitation programs. The Future of Children, 9(1), 27-43.

Gormley, W. T., & Gayer, T. (2003). Promoting school readiness in Oklahoma: An evaluation of Tulsa's pre-k program (CROCUS Working Paper #1). Washington, DC: Georgetown University, Center for Research on Children in the United States. Available: www.crocus.georgetown.edu/working.paper.1.pdf

Henry, G. T., Henderson, L. W., Ponder, B. D., Gordon, C. S., Mashburn, A. J., & Rickman, D. K. (2003). Report of the findings from the Early Childhood Study: 2001-02. Atlanta, GA: Georgia State University, Andrew Young School of Policy Studies.

Heubert, J. P., & Hauser, R. M. (Eds.). (1999). High-stakes: Testing for tracking, promotion, and graduation. Washington, DC: National Academy Press.

Horan, M. D., Irvine, D. J., Flint, D. L., & Hick, T. L. (1980). A prekindergarten program: Policy implications of the research. Education and Urban Society, 12, 193-210.

Karoly, L. A., Greenwood, P.W., Everingham, S.S. et al. (1998). Investing in our children: What we know and don't know about the costs and benefits of early childhood interventions. Santa Monica, CA: RAND

Knitzer, J., & Page, S. (1998). Map and track: State initiatives for young children and families. New York: National Center for Children in Poverty.

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

Marcon, R. A. (1999). Differential impact of preschool models on development and early learning of inner-city children: A three-cohort study. Developmental Psychology, *35*, 358-375.

McKey, R., Condelli, L., Ganson, H., Barrett, B., McConkey, C., & Plantz, M. (1985). The impact of Head Start on children, families, and communities: Final report of the Head Start Evaluation, Synthesis, and Utilization Project. Washington, DC: U.S. Department of Health and Human Services.

Mitchell, A. (2001). Prekindergarten programs in the states: Trends and issues. Climax, NY: Early Childhood Policy Research. Available: www.nccic.org/pubs/prekinderprogtrends.pdf

Mitchell, A., Ripple, C., & Chanana, N. (1998). Prekindergarten programs funded by the states: Essential elements for policy makers. New York: Families and Work Institute.

Reynolds, A. J., & Temple, J. A. (in press). Economic returns of investments in preschool education. In E. F. Zigler, W. S. Gilliam, & S. M. Jones (Eds.), A Vision for Universal Preschool Education.

Ripple, C. H., Gilliam, W. S., Chanana, N., & Zigler, E. (1999). Will fifty cooks spoil the broth? The debate over entrusting Head Start to the states. American Psychologist, *54*, 327-343.

Schulman, K., Blank, H., & Ewen, D. (1999). Seeds of success: State prekindergarten initiatives 1998-1999. Washington, DC: Children's Defense Fund.

Schweinhart, L. J., Barnes, H. V., & Weikart, D. P. (1993). Significant benefits: The High/Scope Perry Preschool study through age 27. Monographs of the High/Scope Educational Research Foundation, *10*. Ypsilanti, MI: High/Scope Educational Research Foundation.

Smith, S. L., Fairchild, M., & Groginsky, S. (1997). Early childhood care and education: An investment that works (2nd ed.). Washington, DC: National Conference of State Legislatures.

University of the State of New York, & New York State Education Department. (1977). Preliminary report of findings: Evaluation of the New York State Experimental Prekindergarten Program. Albany: Authors.

Xiang, Z., & Schweinhart, L. J. (2002). The Michigan School Readiness Program evaluation through age 10. The Journal of At-Risk Issues, 8(1).

Yoshikawa, H. (1995). Long-term effects of early childhood programs on social outcomes and delinquency. The Future of Children, 5(3), 51-75.

Zigler, E., Taussig, C., & Black, K. (1992). Early childhood intervention: A promising preventative for juvenile delinquency. American Psychologist, 47, 997-1006.

Zill, N., Resnick, G., Kim, K., O'Donnell, K., Sorongon, A., McKey, R. H., et al. (2003, May). Head Start FACES 2000: A whole-child perspective on program performance – Fourth progress report. Washington, DC: U.S. Department of Health and Human Services, Administration for Children and Families. Available:
www.acf.hhs.gov/programs/core/ongoing_research/faces/faces_pubs_reports.html

Footnotes

¹ This paper is an updated revision of Gilliam & Zigler (2001). Therefore, much of the earlier work is contained within this report. This paper was supported by a grant to the lead author by the National Institute for Early Education Research (NIEER), which is funded by the Pew Charitable Trusts.

² Carol H. Ripple assisted with collecting many of the reports completed between 1977 and 1998, and NIEER staff assisted with collecting many completed between 1998 and 2003. Coordinating assistance from Crista M. Marchesseault is gratefully acknowledged, as well as assistance from the staff at the Yale National Prekindergarten Study.

³ For the purposes of this paper, we defined state-funded prekindergarten as those systems of early education that (a) serve children in the 3- through 4-year old age range, sometimes through targeting services to low-income or otherwise at-risk children but occasionally through universal access to all age-eligible children; (b) have a stated goal of promoting “school readiness” or some other positive outcome for the children being served; (c) provide at least some form of classroom-based, educational service directly to the children; (d) are administered at the state level, typically through a state department of education but occasionally through some other state department or the Governor’s Office; (e) receive at least some level of state funding; and (f) do not serve exclusively children with disabilities.

⁴ The National Prekindergarten Study (NPS) is an ongoing study of the implementation of state-funded prekindergarten programs in a random sample of about 3900 classrooms operating in 55 statewide prekindergarten programs across 40 states. The data being referenced here were obtained as part of the sampling process which required the investigators to document the location of over 55,000 prekindergarten classrooms across the nation. The NPS is funded by the

National Institute for Early Education Research (NIEER), which is supported by a grant from the Pew Charitable Trusts, and by the Foundation for Child Development.

⁵ Some states fund and administer more than one prekindergarten system.

⁶ A common convention for interpreting the magnitude of standardized effect sizes is to group them into one of four bands: trivial ($< .20$), small (.20 to .50), moderate (.50 to .80), and large ($> .80$; Cohen, 1962, 1988). However, even effect sizes that Cohen would categorize as “trivial” can be quite meaningful when the outcomes are highly valued (McCartney & Rosenthal, 2000; Rosenthal, 1993).

⁷ Although significance levels were not reported in the Georgia UPK evaluation (Henry et al, 2003), we assume that it follows the $p < .05$ convention as a minimum threshold for statistical significance.

⁸ Non-cumulative reporting of grade retention means that the data reflect the percentage of children retained at an actual grade level. Cumulative reporting, however, means that the data reflect the percentage of children that were retained by a given grade level.

⁹ Using a \$7,959 per child cost of retention, and the rates of associated school drop-out and expected lifetime earnings, Reynolds and Temple (in press) estimate the cost-benefit of grade retention at -\$3.13.

Appendix

State-Funded Prekindergarten Evaluation Reports Reviewed

1. Bond, J. T. (n.d.). Interim report: The evaluation of Connecticut's school-readiness program – Cohorts 1 & 2 through Spring 2000. Unpublished report, Connecticut Department of Social Services & Connecticut Department of Education, Hartford.
2. Brezausek, C. M., Greenleaf, W., Matthews, M. J., Ramey, C. T., & Stokes, B. R. (2003). LA 4 Prekindergarten Program evaluation 2002-2003. Baton Rouge, LA: Louisiana Department of Education.
3. Cornwell, J. R. & Kersteter, P. G. (2002). Head Start/ECAP outcomes report: State of Delaware. Dover, DE: Delaware Early Childhood Center.
4. Experimental Prekindergarten Program: Follow-up Study. (1985). Unpublished draft.
5. Flint, D. L., Hick, T. L., Irvine, D. J., Horan, M. D., & Kukuk, S. E. (1979). Effects of exposure to prekindergarten on five social competency factors (Technical paper #8). Albany: New York State Education Department, Prekindergarten Evaluation Unit.
6. Flint, D. L., Hick, T. L., Kukuk, S. E., Horan, M. D., & Irvine, D. J. (1978). Effects of prekindergarten on the cognitive performance of children at the end of prekindergarten: Wave II. In D. J. Irvine (Chair), Ongoing research on the New York State Experimental Prekindergarten Program. Symposium conducted at the annual meeting of the Northeastern Educational Research Association, Ellenville, NY.
7. Florian, J. E., Schweinhart, L. J., & Epstein, A. S. (1997). Early returns: First year report of the Michigan School-Readiness Program Evaluation. Ypsilanti, MI: High/Scope Press. [Available: www.highscope.org/MSRP/Support/Points%20of%20Light.PDF]
8. Gormley, W. T., & Gayer, T. (2003). Promoting school readiness in Oklahoma: An evaluation of Tulsa's pre-k program (CROKUS Working Paper #1). Washington, DC: Georgetown University, Center for Research on Children in the United States. Available: www.crocus.georgetown.edu/working.paper.1.pdf
9. Henry, G. T., Henderson, L. W., Ponder, B. D., Gordon, C. S., Mashburn, A. J., & Rickman, D. K. (2003). Report of the findings from the Early Childhood Study: 2001-02. Atlanta, GA: Georgia State University, Andrew Young School of Policy Studies.
10. Hick, T. L., Flint, D. L., Kukuk, S. E., Horan, M. D., & Irvine, D. J. (1978). Effects of prekindergarten on the cognitive performance of children in kindergarten. In D. J. Irvine (Chair), Ongoing research on the New York State Experimental Prekindergarten Program. Symposium conducted at the annual meeting of the Northeastern Educational Research Association, Ellenville, NY.
11. Hick, T. L., Irvine, D. J., Horan, M. D., Flint, D. L., Kukuk, S. E., & Fallon, E. (1979). Effects of parent involvement in a prekindergarten program on children's cognitive performance. Albany: New York State Education Department, Prekindergarten Evaluation Unit.
12. Hick, T. L., Kukuk, S. E., Horan, M. D., & Irvine, D. J. (1977). Effects of prekindergarten on three measures of cognitive development (Technical paper #3). Albany: New York State Education Department, Prekindergarten Evaluation Unit.
13. Hick, T. L., Kukuk, S. E., Horan, M. D., Irvine, D. J., & Flint, D. L. (1978). Effects of exposure to prekindergarten on the development of social competency in children (Technical paper #5). Albany: New York State Education Department, Prekindergarten Evaluation Unit.
14. Horan, M. D., Irvine, D. J., Flint, D. L., & Hick, T. L. (1980). A prekindergarten program: Policy implications of the research. Education and Urban Society, 12, 193-210.
15. Irvine, D. J., Flint, D. L., Hick, T. L., Horan, M. D., & Kukuk, S. E. (1981). Effects of an early childhood education program on children's verbal knowledge and general reasoning ability in third grade (Technical paper #10). Albany: New York State Education Department, Prekindergarten Evaluation Unit.
16. Irvine, D. J., Horan, M. D., Flint, D. L., Hick, T. L., & Kukuk, S. E. (1978). Obtaining control groups for evaluating the New York State Experimental Prekindergarten Program. Albany: New York State Education Department, Prekindergarten Evaluation Unit.
17. King, F. J., Cappellini, C. H., & Gravens, L. (1995). A longitudinal study of the Florida Prekindergarten Early Intervention Program: Part III. Tallahassee: Florida Department of Education.
18. King, F. J., Cappellini, C. H., & Rohani, F. (1995). A longitudinal study of the Florida Prekindergarten Early Intervention Program: Part IV. Tallahassee: Florida Department of Education.
19. King, F. J., & Gravens, L. (1994). A longitudinal study of the Florida Prekindergarten Early Intervention Program: Part II. Tallahassee: Florida Department of Education.
20. King, F. J., Rohani, F., & Cappellini, C. H. (1991). The third party evaluation report on the Prekindergarten Early Intervention Program, the Florida First Start Program, the Children's Early Investment

- Program, and the Community Resource Mother or Father Program. Tallahassee: Florida Department of Education.
21. King, F. J., Rohani, F., & Cappellini, C. H. (1993). A four-year longitudinal study of the Florida Prekindergarten Early Intervention Program. Tallahassee: Florida Department of Education.
 22. King, F. J., Rohani, F., & Morris, D. (1994). Parental involvement in the Prekindergarten Early Intervention Program. Tallahassee: Florida Department of Education.
 23. Louisiana Board of Elementary and Secondary Education. (n.d.). Unpublished raw data. Author: Baton Rouge, LA.
 24. Marcon, R. A. (1987). Evaluation of educational programs serving four-year-olds. Washington, DC: District of Columbia Public Schools.
 25. Marcon, R. A. (1987). Early learning and early identification study: 1986-1987. Washington, DC: District of Columbia Public Schools.
 26. Marcon, R. A. (1989). Early learning and early identification study: 1987-1988. Washington, DC: District of Columbia Public Schools.
 27. Marcon, R. A. (1990). Early learning and early identification: Final report of the three year longitudinal study. Washington, DC: District of Columbia Public Schools.
 28. Marcon, R. A. (1994). Early learning and early identification follow-up study: Transition from the early to the later childhood grades: 1990-93. Washington, DC: District of Columbia Public Schools.
 29. Marcon, R. A. (1994). Doing the right thing for children: Linking research and policy reform in the District of Columbia Public Schools. Young Children, 50(1), 8-20.
 30. Marcon, R. A. (1999). Differential impact of preschool models on development and early learning of inner-city children: A three cohort study. Developmental Psychology, 35, 358-375.
 31. Martinez, S., & Day, E. (2002). Four-Year-Old At-Risk Program final evaluation. Topeka, KA: Kansas State Department of Education, Planning and Research Section.
 32. Maryland State Department of Education. (1991). An analysis of the long-term effects of the Extended Elementary Education Prekindergarten Program. Baltimore, MD: Author.
 33. McKean, K. & Thistlethwaite, P. C. (1996). Arkansas governor's commission early childhood study: Year one 1995-1996.
 34. Northwest Regional Educational Laboratory. (1989). Tracking success for children and families: ECEAP longitudinal evaluation study year one report. Portland, OR: Author.
 35. Northwest Regional Educational Laboratory. (1991). Tracking success for children and families: ECEAP longitudinal evaluation study year 2 technical report. Portland, OR: Author.
 36. Northwest Regional Educational Laboratory. (1992). Tracking success for children and families: ECEAP longitudinal evaluation study year 3 technical report. Portland, OR: Author.
 37. Northwest Regional Educational Laboratory. (1993). 1992 ECEAP longitudinal study and annual report: An evaluation of child and family development through comprehensive preschool services. Portland, OR: Author. (ERIC Document Reproduction Service No. ED 360 422)
 38. Northwest Regional Educational Laboratory. (1994). 1993 ECEAP longitudinal study: Year 5 technical report: An evaluation of child and family development through comprehensive preschool services. Portland, OR: Author.
 39. Northwest Regional Educational Laboratory. (1995). Early childhood education and assistance program: An investment in children and families. 1994 longitudinal study: Year 6. Portland, OR: Author.
 40. Northwest Regional Educational Laboratory. (1998a). Early Childhood Education and Assistance Program: An investment in children and families: 1995 longitudinal study: Year 7. Summary. Portland, OR: Author.
 41. Northwest Regional Educational Laboratory. (1998b). Early Childhood Education and Assistance Program: An investment in children and families: 1995 longitudinal study: Year 7. Portland, OR: Author.
 42. Pilcher, L. C., & Kaufman-McMurrain, M. (1994). Georgia prekindergarten program evaluation (Georgia Department of Education, Contract No. 950969). Atlanta: Georgia State University, Department of Early Childhood Education.
 43. Pilcher, L. C., & Kaufman-McMurrain, M. (1995). The longitudinal study of Georgia's prekindergarten children and families: 1994-1995. Atlanta: Georgia State University, Department of Early Childhood Education.
 44. Pilcher, L. C., & Kaufman-McMurrain, M. (1996). The longitudinal study of Georgia's prekindergarten children and families: 1995-1996. Atlanta: Georgia State University, Department of Early Childhood Education.

45. Quay (Pilcher), L. C., & Kaufman-McMurrain, M. (1993). Georgia prekindergarten program evaluation (Georgia Department of Education, Contract No. 940996). Atlanta: Georgia State University, Department of Early Childhood Education.
46. Quay (Pilcher), L. C., Kaufman-McMurrain, M., Steele, D. C., & Minore, D. A. (1997). The longitudinal evaluation of Georgia's prekindergarten program. Paper presented at the biennial meeting of the Society for Research in Child Development.
47. Resnick, G. (2001). The evaluation of early childhood education programming in the 30 Abbott school districts: First-year report on program implementation and descriptions of children and families. Trenton, NJ: New Jersey Department of Human Services & New Jersey Department of Education.
48. South Carolina Department of Education. (1986). Early childhood development programs: Half-day programs for four-year olds. 1983-84 school year: Large sample evaluation report. Columbia: South Carolina Department of Education, Office of Research. (Available: South Carolina State Library, State Documents)
49. South Carolina Department of Education. (1987). Executive summary of the evaluation of South Carolina half-day programs for four-year-olds. Columbia: South Carolina Department of Education, Office of Research. (Available: South Carolina State Library, State Documents)
50. South Carolina Department of Education. (1988). The longitudinal evaluation of early childhood education: Interim report. The half-day child development program for four-year-olds (The Office of Research Report Series). Columbia: South Carolina Department of Education, Office of Research. (Available: South Carolina State Library, State Documents)
51. South Carolina Department of Education. (1990a). The longitudinal evaluation of early childhood education: Third interim report-revised. The half-day child development program for four-year-olds (The Office of Research Report Series). Columbia: South Carolina Department of Education, Office of Research. (Available: South Carolina State Library, State Documents)
52. South Carolina Department of Education. (1990b). The longitudinal evaluation of early childhood education: Fourth interim report. The half-day child development program for four-year-olds (The Office of Policy Research Report Series). Columbia: South Carolina Department of Education, Office of Policy Research. (Available: South Carolina State Library, State Documents)
53. Squires, J. H. (1995). Vermont early education initiative child follow-through survey: Preliminary results: Kindergarten, first and second grades. Montpelier: Vermont Department of Education.
54. Tenenbaum, I. M. (2002). What is the penny buying for South Carolina? Child development programs for four-year olds: Student and program characteristics, longitudinal study of academic achievement, and current parent perceptions. Columbia, SC: South Carolina State Department of Education, Office of Research.
55. Texas Education Agency. (1992). Texas evaluation study of prekindergarten programs: Preliminary findings (Publication No. GE2-091-08). Austin, TX: Author.
56. Texas Education Agency. (1993). Texas evaluation study of prekindergarten programs: Interim report (Publication No. GE3-410-09). Austin, TX: Author.
57. Texas Education Agency. (1995). Texas evaluation study of prekindergarten programs: Final report summary (Publication No. GE5-170-01). Austin, TX: Author.
58. Texas Education Agency. (1995). Texas evaluation study of prekindergarten programs: Final report (Publication No. GE5-170-02). Austin, TX: Author.
59. University of Kentucky. (1992). Third party evaluation: Kentucky Education Reform Act (KERA) Preschool Programs. Final report. Lexington: University of Kentucky, College of Human Environmental Sciences and College of Education.
60. University of Kentucky. (1993). Third party evaluation: Kentucky Education Reform Act (KERA) Preschool Programs. Final report. Lexington: University of Kentucky, College of Education and College of Human Environmental Sciences.
61. University of Kentucky. (1994). Third party evaluation of the Kentucky Education Reform Act Preschool Programs. Lexington: University of Kentucky, College of Education and College of Human Environmental Sciences.
62. University of Kentucky. (1995). Third party evaluation of the Kentucky Education Reform Act Preschool Programs. Lexington: University of Kentucky, College of Education and College of Human Environmental Sciences.
63. University of Kentucky. (1996). Third party evaluation of the Kentucky Education Reform Act Preschool Programs. Lexington: University of Kentucky, College of Education and College of Human Environmental Sciences.
64. University of Kentucky. (1997). Third party evaluation of the Kentucky Education Reform Act Preschool

Programs. Lexington: University of Kentucky, College of Education and College of Human Environmental Sciences.

65. University of the State of New York, & New York State Education Department. (1977). Preliminary report of findings: Evaluation of the New York State Experimental Prekindergarten Program. Albany: Authors.
66. University of the State of New York, & New York State Education Department. (1978). Second-year findings, with special emphasis on cognitive outcomes: Evaluation of the New York State Experimental Prekindergarten Program. Albany: Authors.
67. University of the State of New York, & New York State Education Department. (1982). Evaluation of the New York State Experimental Prekindergarten Program: Final report. Albany: Authors. (ERIC Document Reproduction Service No. ED 219 123)
68. Wode, J., Salehi, S., & Eckroade, G. (1992). An analysis of the long-term effects of the EEEP Program: The secondary years. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.
69. Xiang, Z., & Schweinhart, L. J. (2002). Effects five years later: The Michigan School Readiness Program Evaluation through age 10. Ypsilanti, MI: High/Scope Educational Research Foundation. Available: www.highscope.org/Research/MsrpEvaluation/msrp-Age10-2.pdf.