

Early Childhood Education: A Meta-Analytic Affirmation of the Short- and Long-Term Benefits of Educational Opportunity

Kevin M. Gorey
University of Windsor

Some scholars who emphasize the heritability of intelligence have suggested that compensatory preschool programs, designed to ameliorate the plight of socioeconomically or otherwise environmentally impoverished children, are wasteful. They have hypothesized that cognitive abilities result primarily from genetic causes and that such environmental manipulations are ineffective. Alternatively, based on the theory that intelligence and related complex human behaviors are probably always determined by myriad complex interactions of genes and environments, the present meta-analytic study is based on the assumption that such behaviors can be both highly heritable and highly malleable. Integrating results across 35 preschool experiments and quasi-experiments, the primary findings were: (a) preschool effects on standardized measures of intelligence and academic achievement were statistically significant, positive, and large; (b) cognitive effects of relatively intense educational interventions were significant and very large, even after 5 to 10 years, and 7 to 8 of every 10 preschool children did better than the average child in a control or comparison group; and (c) cumulative incidences of an array of personal and social problems were statistically significantly and substantially lower over a 10- to 25-year period for those who had attended preschool (e.g., school drop out, welfare dependence, unemployment, poverty, criminal behavior). The need for a very large, well-controlled, national experiment to either confirm or refute these provocative, review-generated findings is discussed.

Conventional wisdom certainly seems to support the notion that educational experiences early in childhood are beneficial. This is particularly true when considering compensatory preschool programs that are designed to serve children who, for any number of social or economic reasons, are at greater than average risk of

The author gratefully acknowledges the research assistance of Marcia J. Andkilde and Cynthia M. Meyer. Helpful critical commentary on manuscript drafts was provided by Nancy L. Richter, James Brophy, Margaret Keith, and Arthur G. Cryns.

Address correspondence to Kevin M. Gorey, School of Social Work, University of Windsor, 401 Sunset Ave., Windsor, Ontario, N9B 3P4, Canada. E-mail: gorey@uwindsor.ca.

experiencing learning difficulties. One envisions the educational intervention filling gaps, compensating for the relative lack of developmental opportunities experienced by children who, for example, live in extremely poor, segregated neighborhoods. Therefore, hypotheses typically advanced about expected intelligence and academic achievement gains and improved academic and other life successes because of various compensatory preschool interventions, hold a good deal of practical face validity. However, our understanding of the true effects of early childhood education arguably remains debatable and, unfortunately, the debate takes place more often in political rather than scientific forums. Given the lack of any recent integrative study of this topic, such a scientific investigation is needed to inform social scientists, educators, and policy makers. This meta-analytic study aims to fill this knowledge gap by reviewing the scientific evidence on the effectiveness of early childhood educational programs.

HERITABILITY AND MALLEABILITY OF INTELLIGENCE

The posing of research questions about the expected effects of early childhood education—large, moderate, small, or nonexistent—rests on the psychological and sociological evidence pertinent to the heritability and malleability of intelligence. As for heritability, some researchers have estimated that 50% or more of the variability in intelligence among populations of children can be because of genetic factors (Herrnstein & Murray, 1994; Jensen, 1985; Rushton, 1995; Rushton & Ankney, 1995). Others, adjusting for socioeconomic and other environmental factors, have concluded that the exclusive heritability or main effect of genes on intelligence is actually quite low, ranging from less than 1% to no more than 10% (Cryns & Gorey, 2000; Gorey & Cryns, 1995; 1999; Gould, 1995; Nisbett, 1998). This debate points toward the probable central importance of gene–environment interactions to ultimately determine such a complex behavior as intelligence. The abilities to think analytically, reason abstractly, and solve problems through the manipulation of words and numbers, tied as they are to the extraordinary complexities of the cerebral cortex and its associated central nervous system networks, obviously arise as an expression of genes. Indeed, these cognitive skills probably result from a constellation of many gene-by-gene interactions far more complex than can possibly be imagined at this time. The fact that analyses in some way accounted for the environment have produced very low estimates of the heritability of intelligence does not mean that genes do not matter; they most assuredly do. Rather, such analyses remind us that genes are expressed in environments, which addresses the issue of malleability.

Primary and secondary analysis of identical and fraternal twins and non-twin siblings, reared together or apart in the homes of either their biological or adoptive parents, have produced a wide range of results pertinent to intelligence, particularly its malleability (e.g., how much children's IQs can be influenced by the home and related environments). Malleability estimates have ranged from none to as many as 25 IQ points ($Mdn = 9$ IQ points; Bouchard,

Lykken, McGue, Segal, & Tellegen, 1990; Capron & Duyme, 1989; Cyphers, Fulker, Plomin, & DeFries, 1989; Devlin, Daniels, & Roeder, 1997; Locurto, 1990; Scarr & Weinberg, 1978; 1983; Tambs, Sundet, Magnus, & Berg, 1989; Turkheimer, 1991; Weinberg, Scarr, & Waldman, 1992). Their methodological and contextual variability notwithstanding, in aggregate these studies demonstrate that enriching manipulations of children's home environments can be associated with substantial score gains on standardized measures of intelligence. Because such manipulations of home environments are clearly analogous to planned manipulations of educational environments (e.g., preschool interventions), it seems reasonable to expect similar IQ gains with their attendance.

It should be noted that nearly all of the authors in this field have recognized at least implicitly the probable concomitant importance of genetic endowment and environments, but the possibly critical environmental elements have only been surmised to date. Other research has suggested that children's intelligence may be affected by a diverse array of environmental influences: the quality and amount of their own, their parents', and their grandparents' education; other socioeconomic characteristics of their parents and neighborhoods, including the prevalence of poverty; household size; academic and other stimulation received at home, school, and elsewhere; maternal health and related birthweights; prenatal alcohol exposure and exposure to other environmental hazards, such as lead or tobacco smoke (Bacharach & Baumeister, 1998; Brooks-Gunn, Klebanov, & Duncan, 1996; Christian, Morrison, & Bryant, 1998; Coles et al., 1991; Duncan & Brooks-Gunn, 1997; Jencks & Phillips, 1998; Johnson, Swank, Baldwin, & McCormick, 1999; Larroque & Kaminski, 1998; Schwartz, 1994). Similar to genetic effects, it seems probable that the environment also affects intelligence in complex ways. The environment—familial, educational, social, economic, political, and physical—is most assuredly strongly related to children's intellectual development. The environment matters, and the enriching opportunities provided by a high-quality preschool education probably matter very much, particularly among children who are socioeconomically or otherwise environmentally impoverished.

Currently there is a profound lack of coherence among the major theories of intellectual development and this lack of theoretical integration tends to reinforce an overly simplistic and erroneous world view. Some theorists tend to emphasize the importance of personal, genetic, or biological factors. The theories of others are dominated by social and environmental factors. Nature is continually pitted against nurture, as if interaction effects did not exist, and as if heritability and malleability were mutually exclusive phenomena. Perhaps a complex behavior such as intelligence can be understood to be highly heritable and highly malleable at the same time (i.e., perhaps it is the result of the interaction of myriad genetic and environmental characteristics). In fact, it is increasingly accepted among natural and social scientists that 100% of human disease (and probably most behavior) is genetically caused and 100% is caused by environmental factors (Rothman & Greenland, 1998; Wilson, 1998). This

theoretical acceptance of the centrality of gene–environment interactions can facilitate the unification of existing natural and social scientific knowledge and facilitate a more integrated understanding of human behavior. For example, implicit in many studies of the heritability of intelligence has been the notion that if intelligence is highly heritable it must not be very malleable: the greater the proportion of intellectual variance accounted for by genetic factors, the lesser the proportion accounted for by environmental factors. This mutually exclusive algorithm is based on the surely erroneous assumption of a main effects only causal model. These main effects, if perfectly measured, would provide a complete causal explanation of intellectual development. Interaction effects are typically not included in such models. However, if main effects are not mutually exclusive and if interaction effects are included in behavioral models, then the opportunities for practical interventional applications of scientific knowledge are potentiated. A heritable characteristic such as intelligence could indeed be amenable to change (Lehrman, 1970) through environmental manipulations. The institution of policies that provide for more equitable educational and other life opportunities for all children and the institution of programs that provide compensatory preschool opportunities for children of relatively deprived backgrounds are examples of such manifestations. The effects of American social policies have been empirically demonstrated by the incremental gains in IQ scores over the course of only a couple generations in addition to the diminishing African American–Caucasian IQ gap (Flynn, 1984; Vincent, 1991). However, such demonstrable verification of the effectiveness of preschool programs remains to be accomplished.

EFFECTIVENESS OF EARLY CHILDHOOD EDUCATIONAL INTERVENTIONS

Ongoing scientific inquiry on early childhood education is needed. Without it, the mere ebb and flow of political tides potentially have the power to overwhelm rational and empirical evidence that steer social and educational policies. For example, treatment of the topic of preschool interventions in *The Bell Curve* (Herrnstein & Murray, 1994) provocatively identified the highly heritable nature of intelligence and so assumed its relative unmalleability. Then, consistent with currently popular reform notions, the authors concluded that preschool interventions, like other purportedly wasteful social welfare policies, could be expected to have little impact on intelligence or other human behaviors. Herrnstein and Murray's treatise on compensatory preschool education cited only a handful of studies which, for the most part, secondarily analyzed data from the 1960s and 1970s (Adams, 1989; Brown & Campione, 1982; Ceci, 1991; Jensen, 1993; McLaughlin, 1977). Seminal work from the 1980s and 1990s was conspicuously absent. My review of recent reviews and editorials found as many encouraging as disparaging conclusions about the effectiveness of early childhood educational interventions (Barnett, 1998; Garber &

Hodge, 1991; Karoly et al., 1998; Locurto, 1991; Ramey, 1999; Raver & Zigler, 1997; Ripple, Gilliam, Chanana, & Zigler, 1999; Schweinhart & Weikart, 1991; Spitz, 1997). However, because they have typically only qualitatively summarized diverse outcomes observed by diverse studies of diverse preschool programs, it is exceedingly difficult to develop a clear picture of what is really happening. Preschool programs vary widely. For example, in terms of dollars invested per child served, various state Head Start programs vary by a factor of more than eight, ranging from annual funding levels of approximately \$1,000 to more than \$8,000 per child. Given that such investments are surely related to various potentially important program characteristics (e.g., teacher qualifications, teacher/student ratio, program intensity, and duration), our understanding of how they affect program outcomes would be served by a thorough synthesis of the relevant scientific research.

This study will translate scientific evidence about the effectiveness of early childhood education into practical evidence by the means of a meta-analysis. One can then judge for oneself: (a) are its compensating effects nonexistent, small, moderate, or large? and (b) which intervention characteristics are predictive of relatively larger effects?

METHODS

Selection of the Sample for Meta-Analysis

In January 2000 the following research literature data bases were searched (1990 to 2000): *ERIC*, *Education Index*, *PsycINFO*, *Social Work Abstracts*, *Sociological Abstracts*, *Social Science Index*, and *Index Medicus*. The key word search {preschool or compensatory education} and {intelligence, IQ, or cognition} and {effectiveness, effect, efficacy, evaluation, assessment, benefit, outcome, or follow-up} produced 49 conceptually relevant studies. Given that the focus was on rigorous evaluations of preschool interventions, studies had to meet the following additional methodological criteria to be included: (a) used a quasi-experimental comparison condition or randomized experimental control group(s); (b) assessed the pretest equivalence of their treatment and comparison or control groups and accounted for any observed nonequivalence (sample restriction, matching or mathematical models); and (c) presented their findings with sufficient detail so that corresponding effect sizes were calculable (minimally group means and standard deviations or statistical significance levels with group sample sizes). Because the focus was on early preventive cognitive interventions with socioeconomically but not otherwise disadvantaged children, studies were excluded if they sampled children older than 5 years of age with specific medical or organic problems (e.g., cognitive and other interventions with low birthweight infants or those with fetal alcohol syndrome), or if their primary focus was on intervention with already mentally retarded or otherwise developmentally or physically disabled children. A total of 21 of the originally selected studies met these

conceptual, methodological, and empirical criteria. Bibliographies of these 21 retrieved manuscripts were searched for earlier independent studies that met the same inclusion and exclusion criteria; 14 additional studies were selected. The 35 total studies selected for this analysis tested 80 hypotheses.¹ If a study used more than one measure of the same conceptual domain, the effects were combined (weighted average) into one independent hypothesis test. If a study used multiple conceptual measures (e.g., intelligence and academic achievement) they were treated as independent hypotheses for this meta-analysis.

Meta-Analysis and Effect Size Interpretation

Cohen's (1988) U_3 statistic was used as an index of effect size and practical significance. It is an intuitively appealing metric that compares all the scores of an intervention group's members on a dependent measure with a control group's average score at posttest. For example, a U_3 of 75% resulting from a 3-year, posttest comparison of a group of preschool participants with a nonparticipating control group on a standardized intelligence test would be interpretable as follows: 75% of the preschoolers scored higher than the average child in the control group. U_3 s were derived from d -indexes weighted by study sample sizes that are calculable from a variety of intervention-control outcomes statistics (group M s and SD s, t -test, F -ratio, χ^2 , and p level with group N s) and allow for ease of across-study comparison (Cooper, 1998; Cooper & Hedges, 1994). The statistical significance of aggregate cognitive and related effects of preschool intervention were tested with Rosenthal's (1978) weighted z s method of combining probabilities (weighted by study samples sizes). As a control for publication bias, fail-safe N s at $p < .05$ (Rosenthal, 1979) were computed for each major behavioral domain.

Although there are many extant effect size indexes, psychologists and other behavioral scientists will probably be most acquainted with Cohen's d -index. The d -index is interpretable as the number of outcome measure SD shifts experienced by an intervention group, compared with a comparison or control group. In fact, the d -index is often referred to erroneously as a synonym for study "effect size." I prefer another of Cohen's effect size metrics, the U_3 statistic, primarily because it tends to put the emphasis on people, rather than measures, and so in my view, lends itself well to the assessment of clinical and policy significance. Actually, being a direct function of d , U_3 basically provides the same information about the size of an effect. In the interest of clear communication, the following effect size landmark conversions may serve as an interpretive guide: (a) a d of 0.25 or a one-quarter SD shift is equivalent to a U_3 of 60% (e.g., six of every 10 participants in an intervention group score higher on the measure being studied than the average participant in a comparison group); (b) a d of 0.50 is equivalent to a U_3 of 70%; (c) a d of 1.00 equivalent to a U_3 of 85%; (d) a d of

1. References marked with an asterisk indicate studies included in the meta-analysis.

1.30 equivalent to a U_3 of 90%; and (e) a d of 1.65 is equivalent to a U_3 of 95%. For interpretive purposes, it will also be useful to note that a one SD change ($d = 1.00$, $U_3 = 85\%$) represents an actual score change of 15 IQ points on most standardized measures of intelligence.

Accounting for study validity. The central potential threat to the validity of this meta-analytic study (i.e., that any of the estimated effects of preschool intervention are really methodological artifacts rather than true treatment effects) proceeds directly from potential threats to the internal validity of its sample of studies. Because these 35 studies all attempted to observe cognitive or other behavioral change over time, in most cases years (even tens of years) after the intervention, and because their critical comparisons were between intervention and comparison or control groups at posttest, the most likely alternative explanations are that possibly the study groups were not equivalent at pretest or that some systematic loss to follow up occurred. For example, if significant numbers of initial study participants refuse or cannot be found for posttest measurement, and if these rates of attrition differ between the study groups and are also related to study outcomes (i.e., mortality is associated with the hypothesized independent {treatment} and dependent {outcome} variables), such differential mortality would probably fatally confound the study's findings.

Three methodologies are most critical in protecting against such threats to study internal validity: (a) the random assignment of participants to experimental groups (intervention and control) or some quasi-experimental approximation of this method (intervention and comparison groups); (b) the use of adequate numbers of participants to assure the pretest equivalence of study groups on the outcome variable of interest and other potential confounds; and (c) the use of various procedures to assure that adequate numbers of participants complete the critical posttest measurement. A summary index of internal validity, based on these three methodological factors, was computed for each of the 35 selected studies as follows: (a) pre-experiment, quasi-experiment, or true randomized experiment (scored respectively, 0, 1, or 2); (b) total initial study sample of 60 or more (sufficient statistical power to detect a d of 1.00; Fleiss, 1981), 200 or more (adequate to detect a d of 0.50), or 800 or more participants, which could detect a d of 0.25 (scored 0 to 2); and (c) attrition rate greater than 25%, 10% to 24%, or less than 10% (scored 0 to 2; missing data were scored as 0). This computed internal validity index has a theoretical score range of 0 to 6 and is keyed so that high scores are indicative of greater validity. The potential moderation of this meta-analytic review's main effects by the internal validity of its sample of studies will be empirically tested.

Hypotheses. Some social-environmental theories of intelligence acknowledge that, the strength of their heritabilities notwithstanding, cognitive abilities are expressed in diverse environments. Commensurate with these theories, the hypotheses of this meta-analytic study were that: (a) preschool program participants would enjoy significant cognitive and related benefits; and (b) better endowed programs would produce larger effects.

RESULTS

Sample Description

This review of the practical effects of early childhood education is based on 80 study outcomes of cognition, including intelligence and achievement test scores, and measures of school performance, in addition to other related measures of personal and social success. More than 18,000 children participated in these studies at more than 200 preschool sites. Approximately one-third of the research was accomplished by two groups; one in Ypsilanti, Michigan, and the other in Chapel Hill, North Carolina. Six national samples were also included. The remainder of the studies were diversely representative of predominantly inner-city areas. Fourteen states, one Canadian province, and Israel were also included in the aggregate sample. Furthermore, the studies were almost exclusively of children who could be deemed at-risk for school failure by any number of criteria: (a) they resided in single-parent, poor or very poor households; or (b) they scored well below the norm on standardized intelligence tests before their participation. More than three-quarters of the aggregate review sample was represented by racial or ethnic minorities, predominantly African American and Hispanic children. Among the 24 studies that reported sufficient descriptive detail, the aggregate intervention and comparison/control groups did not differ substantively on their racial composition; they included 73.2% and 71.8% African American children, respectively.

The research on compensatory preschool programs reflects their great diversity (left side of Table 1). Programs differed by major typology (Head Start, Perry Preschool, Abecedarian Project, etc.) and varied widely on such characteristics as the typical age participants enter them and the typical intensity and duration of their educational interventions. For example, Head Start programs may generally be placed at the low end of the continuum in terms of the amount of preschool intervention. Typically, children will participate for 1 to 2 school years between the ages of 3 and 5 in half-day classes with 10 to 20 or more children. Head Start teachers, who are required to have a 2-year community college or associate degree, are typically supported by volunteers. On the opposite end of the intervention continuum is the Abecedarian Project. Typical Abecedarian participants begin as infants in daycare-based, full-time programs (6 to 8 hours per day, 5 days per week for 2 to 5 years) with very low ratios of teachers with Masters degrees to children (ranging from one to three or six). The well-researched Perry Preschool program generally fits between these two extremes.

Major study design characteristics of the reviewed outcomes are also displayed in Table 1. They were predominantly (68%) randomized controlled experiments. The remaining research was accomplished with quasi-experiments that used nonrandomized comparison conditions, pre- and posttest measurements, and adjustment techniques such as sample restriction, matching, or mathematical modeling to control potentially potent threats to their internal validity.

TABLE 1. Descriptive Profile of the Preschool Programs and Study Designs

Program Characteristics	Outcomes	
	<i>N</i>	%
Type of Preschool Program		
Perry Preschool	21	26.3
Abecedarian Project	15	18.7
Head Start	11	13.7
Other Model Programs	13	16.3
Combined Models	8	10.0
With Follow Through	12	15.0
Age of Participants at Entry		
Infants	19	23.7
Three to four years of age	52	65.0
Five years of age or older	9	11.3
Intensity of Preschool Intervention ^a		
Low	5	6.3
Moderate	26	32.5
High	49	61.2
Duration of Preschool Intervention		
One year	12	15.0
Two years	34	42.5
Three or more years	34	42.5
Study Design Characteristics		
Type of Research Design ^a		
Experiment	54	67.5
Quasi-Experiment	26	32.5
Total Number of Study Participants		
29 to 100	25	31.2
100 to 199	33	41.3
200 to 4,787	22	27.5
<i>Mdn</i> = 155, <i>M</i> = 519, <i>SD</i> = 1,141		
Follow-Up Years		
One to four	25	31.2
Five to nine	22	27.5
10 to 25	33	41.3
<i>Mdn</i> = 7.0, <i>M</i> = 9.9, <i>SD</i> = 7.3		
Participants Lost to Follow-Up ^c		
2 to 9%	22	36.1
10 to 25%	33	54.1
26 to 55%	6	9.8
<i>Mdn</i> = 11.0, <i>M</i> = 14.7, <i>SD</i> = 8.8		

Note. Initial inter-rater agreement among three raters (two naive) who coded 25 variables from the study manuscripts was 93.5%. Disagreements were discussed so that the ratings were ultimately unanimous (100% agreement).

^aExamples: Low = half day classes of 10 to 20 or more children with paraprofessional teachers holding two-year community college or associate degrees; High = six to eight hours per day, five days per week with very low teacher (Masters degree) to child ratios (1:3 to 1:6); and Moderate = intermediate programs.

^bApproximately half (53%) of the researchers were not institutionally affiliated with the preschool programs being evaluated; 10% were evaluating their own programs, and about a third (37%) of the collaborative teams were comprised of both internal and external evaluators (Gorey, 1996).

^cNot codable from the original study manuscripts for 19 outcomes.

Two-thirds of the preschool evaluations were based on initial samples of more than 100 children (69%), and one of every four of them (27%) used much larger samples ranging in size from 200 to nearly 5,000. Moreover, their investigators seem to have done a remarkable job of successfully following most of the participants. One-third of the evaluation studies (36%) located more than 90% of their original participants at follow up and almost all of them (91%) located 75% or more of their original participants at follow up. Given the very long time that had transpired between project initiation and follow-up assessment for many of these studies—10 to 25 years among nearly half of them—it is clear that these are exceptionally high completion rates.

Cognitive Effects of Preschool Programs

With only one noted exception, all of the sample size-weighted combined probabilities of this study's meta-effects were minimally significant at $p < .05$. This is not surprising because of the 80 original study outcomes, 73 were minimally statistically significant at the .05 α -level criterion, three could be categorically described as having approached significance ($p < .10$), and only four were not significant in a statistical sense. Relatedly, fail-safe N s at $p < .05$ for intelligence, academic achievement, and other behavioral effects of preschool intervention were 109, 148, and 231, respectively.

Aggregated effects of early childhood educational interventions on predominantly standardized measures of intelligence and academic achievement are displayed in the top entry of Table 2. The average intervention effects of preschool on IQ and academic achievement can best be characterized as quite large. At follow up, three-quarters of the children who experienced some preschool educational intervention scored higher on IQ and achievement tests than the average child in a comparison/control group (76% and 78%, respectively; both $p < .05$). Reading down Table 2, the effect of preschool intensity and duration on the magnitude of IQ and achievement outcomes shows a clear relationship between the investment in early childhood education programs and the strength of their compensatory effect. Statistically significant and clinically meaningful effects, all of them quite large by any standard, were observed among the more intensive and extended programs, where, for example, approximately 80% or more of the program participants scored better than the average child in a control group. As a further practical interpretive aid, it should be noted that effects of this size are equivalent to an average IQ difference of 13 or 14 points.

Next, the length of time that had passed since their preschool experience was found to be significantly associated with youngsters' scores on standardized tests of cognition. They tended to perform more poorly after the passage of five years than they did during the preceding post-preschool years. However, on both IQ and achievement tests, many preschool participants (69% and 74%, respectively), even after five or more years, scored better than the average child who did not experience an enriching preschool education. There clearly is some evi-

TABLE 2. Cognitive Effects of Preschool Programs: Weighted Average Effect Sizes

Outcome Measures	<i>N</i>	<i>U</i> ₃	<i>SD</i>		<i>N</i>	<i>U</i> ₃	<i>SD</i>
Intelligence ^{a,c}	23	76.5	8.9	Academic Achievement ^{b,c}	17	78.2	11.0
By Intensity of Preschool Intervention							
Low to Moderate	8	65.3	5.3	Low to Moderate	8	74.0	6.9
High*	15	82.0	7.3	High*	9	82.1	8.7
By Duration of Preschool Intervention							
One or two years	12	69.9	6.7	One or two years	7	74.0	9.1
Three or more*	11	79.0	9.0	Three or more	10	80.9	9.5
By Years of Follow Up							
Five or less	14	80.1	6.1	Five or less	9	88.8	9.2
Greater than five*	9	69.4	7.8	Greater than five*	8	73.9	8.3
By Intensity of Preschool Intervention (Greater Than Five Years of Follow Up)							
Low to Moderate	4	60.7 ^d	5.9	Low to moderate	3	61.9	10.8
High*	5	73.9	9.5	High*	5	79.5	8.7

Note. *N* = number of study outcomes. With three exceptions, each study provided only one intelligence and/or academic achievement-related outcome; one study provided three outcomes by assessing intelligence at three different follow-up points (Martin, Ramey, & Ramey, 1990), and two studies provided two outcomes each by assessing reading and mathematics achievement separately (Reynolds, 1994; Sprigle & Schaefer, 1985).

^aStanford Binet Intelligence Scale (38%), Wechsler Scales (26%, Wechsler Intelligence Scale for Children (WISC), WISC-R {revised}, Wechsler Preschool and Primary Scale of Intelligence, Wechsler Adult Intelligence Scale), McCarthy Scales of Children's Abilities (19%), Bagley Tests of Infant Development (7%), and others (10%, none accounting for more than 2%) (Canivez & Watkin, 1998; Dacey, Nelson, & Stoeckel, 1999; Kaplan, 1996; Karr et al., 1993; Krohn & Lamp, 1999; Slate, 1995; Slate & Jones, 1995).

^bWoodcock-Johnson Psycho-Educational Battery, Part 2, Tests of Achievement (23%), Iowa or Comprehensive Test of Basic Skills (23%), California Achievement Test (14%), grades (9%), and others (31%, none accounting for more than 4%) (Daly, Wright, Kelly, & Martens, 1997; DiPerna & Elliott, 1999; Duran & Powers, 1993; Powers, Escamilla, & Haussler, 1986; Shull, Weatherly, Morgan, & Bradley, 1995).

^cBoth distributions of intelligence and academic achievement effects (*U*₃s) were more heterogeneous than expected due to random sampling variability alone (Hedges & Olkin's (1985) *Q* statistics calculated from corresponding *d*-indexes, *p* < .05).

^dCombined probability not significant. All of the other effects are statistically significant (*p* < .05).

**p* < .05, one-tailed independent samples *t*-test.

dence of the fade-out effect. However, it seems that although, on average, the differences between preschool and comparison/control groups do diminish over time, even years later appreciable numbers of preschool youngsters still demonstrate practically significant IQ and achievement test performance benefits.

Adding a level of complexity to this analysis, with longer term studies the amount of preschool investment was significantly related to outcome (bottom of Table 2). The cognitive effects of intensive preschool programs tend not to fade-out very much, whereas, less intensive programs do show a marked fade-out effect. The percentage of children in intensive preschools scoring better than their average counterpart in comparison/control groups five or more years later was 74% and 80% on IQ and achievement tests, respectively. This represents an av-

erage long-term IQ gain of approximately nine points, for example. Although this meta-analysis lacks sufficient statistical power to test even longer term cognitive effects, they were essentially undiminished among the two evaluations of highly intensive educational interventions that assessed intelligence and academic achievement more than 10 years after preschool participation (U_3 s of 71% and 75%, respectively). After accounting for program intensity, years of follow up, and their interaction (low to moderate or high intensity by ≤ 5 years or > 5 years), none of the other programmatic or study design characteristics displayed in Table 1 could further account for the variability in cognitive effects: type of program, type of research design, internal or external evaluation (dummy variables), typical participant age at entry, study sample size, attrition rate, and program duration (three-level ordinal variables). After program intensity, years of follow up, and their interaction term were force-entered into separate linear regression models with intelligence or academic achievement effects (d -indices) as the dependent variable, multiple-partial F -values (1, 18) for each of the seven predictors of intelligence ranged from 0.13 to 1.54 and similar F -values (1, 12) for academic achievement predictors ranged from 0.51 to 1.72. None of these stepwise regression F statistics even approached statistical significance at the .10 α -level criterion.

Other Personal and Social Effects of Preschool Programs

Many of the studies in this field used categorical measures of various personal and social problems hypothesized to be associated with low intelligence rather than using standardized cognitive measures. Their cognition-related and generally long-term outcomes are displayed in Table 3. A few studies assessed differences between preschool participants and nonparticipants on their assignment to "special" learning or educational statuses and included participants identified as borderline mentally retarded or placed in other special education classes (top of Table 3). For the label of borderline mentally retarded, at an average of eight years after preschool, only one of 10 children (11%) had been so labeled, whereas four of every 10 children (40%) who did not go to preschool had been so labeled, a more than three-fold differential. Preschool intervention offered a nearly three-quarters protective effect (rate ratio = $11.3/39.7 = 0.29$). It prevented three of every four participants from being assigned the label of borderline mentally retarded, a label they would probably otherwise have received at some point during their elementary school years had they not attended preschool.

Standardized cognitive assessment is only one limited way of evaluating early childhood education programs. From the lifespace perspective of children and their families, successful school performance, because it is linked to so many other avenues of success, is another, probably even more meaningful outcome measure (School Performance, Table 3). Relatively long-term follow up of grade failure (22% versus 43%) and high school drop out rates (26% versus 48%)

TABLE 3. Other Personal and Social Effects of Preschool Programs: Aggregate Rates and Rate Ratios

Type of Outcome Measures	Average Years of Follow Up	N	Rate (%) at Follow Up		
			Preschool	Control or Comparison	Rate Ratio
Categorical Cognitive-Related					
Identified borderline mentally retarded	8	3	11.3	39.7	0.29
Ever assigned to special education	11	5	17.4	36.4	0.48
School Performance					
Ever held back a grade	9	8	22.3	43.0	0.52
Not a high school graduate	16	7	26.0	48.4	0.54
Welfare Dependence					
Currently receives welfare assistance	15	2	12.6	32.3	0.39
Ever received welfare as an adult	25	1	59.5	80.5	0.74
Economic Well-Being					
Currently unemployed	17	2	50.2	68.8	0.73
Earnings below poverty criterion	20	2	56.6	84.7	0.67
Not a home owner	25	1	63.1	86.5	0.73
Delinquent and Criminal Behavior					
Ever engaged in delinquent behavior	11	3	30.2	74.0	0.41
Ever arrested	18	4	23.9	51.1	0.47
Arrested five or more times	25	1	6.6	35.8	0.18
Teenage Pregnancy	15	1	43.4	57.8	0.75*

Note. N = number of study outcomes. Rate ratio = preschool rate/control-comparison rate.

* $p < .10$. All of the other critical comparisons (aggregate preschool versus control and comparison samples) were significant at $p < .05$ (χ^2 test statistic (degrees of freedom = 1)).

among preschool and control group youngsters were demonstrative of statistically significant preschool advantages. An enriching preschool education decreased the chances that a child would fail a grade (rate ratio = 0.52) and decreased the chances that a child would drop out of high school before graduation (rate ratio = 0.54), both by nearly 50%. The remainder of the findings displayed in Table 3 are consistent with what is well known about the associations of school failure with various other life difficulties. All of the preschool control group rate ratios on welfare dependence, low socioeconomic status, and criminal behavior were both statistically and practically significant. Over the typical course of near generational follow ups, the cumulative lifetime rates of experiencing such problems were much lower among adults who had attended preschool. For example, compared with their nonpreschooled counterparts, participants were 26% less likely to have ever received welfare, 33% less likely to be poor, less than half as likely to have ever engaged in delinquent or criminal behavior, and strikingly, far less likely (82%) to have developed a criminal lifestyle. Although some of these aggregate findings are based on only a few study outcomes, together the consistency observed across their 40 comparisons certainly seems to indicate that preschool education has a very positive influence

on the lives of its participants transcendent of their scores on tests of intelligence or academic achievement.

Lack of Moderation of Effects by Study Characteristics

The computed index of study internal validity (actual score range of two to five), a function of major research design type, sample size, and attrition, was not associated with effect size. Within intelligence, academic achievement, and other behavioral domains, respective internal validity *d*-index *r*s (degrees of freedom) were $-.19$ (21), $-.31$ (15), and $-.13$ (38). It should be noted that all of the possible associations—treating the index as a categorical, ordinal, or continuous variable, and testing its association with all 80 study outcomes and with the subsamples of intelligence, academic achievement, and other behavioral outcomes—were explored completely with standard correlational and analysis of variance statistics (Pearson's *r*, *t*-test, and *F*-ratio). None of these test statistics even approached statistical significance at the .10 α -level criterion. It seems that the original study sampling scheme was successful in selecting a relatively homogeneous group of fairly rigorous studies: 27 of 35 (77.1%) scored 4 or 5 on the index ($M = 4.16$, $SD = 0.90$). In other words, there is not a great deal of internal validity variance among this meta-analytic sample, so it is not very surprising that the internal validity index itself it is not predictive of effects. Also, representing the sampling scheme's success, all 35 of the studies assessed race, socioeconomic status, or other family background characteristics (e.g., maternal education) at pretest and accounted for them in some way in their analyses. All but three of the studies did the same for cognitive characteristics of children or mothers. Methodological problems, particularly those related to study group nonequivalence at pretest (possible selective group assignment) or posttest (possible selective attrition or follow up), do not seem to be potent alternative explanations for the observed intervention effects of this meta-analysis.

DISCUSSION

This meta-analysis of experimental and quasi-experimental research showed strong support for the idea that early childhood education is a highly effective preventive intervention. Its observed cognitive and behavioral meta-effects were synthesized from a sample of studies that were largely of high internal validity. They offered a good deal of methodological assurance that their observed treatment effects were indeed the effects of planned preschool educational interventions, rather than the mere spurious results of methodological artifacts such as, for example, uncontrolled selection bias. In light of its methodological strengths, perhaps this review's most telling finding is that as preschool intervention intensity increases, so does the magnitude of its positive effects. In fact, this relatively strong relationship between program endowment and outcomes observed among generally rigorous studies seems indicative of the causal nature of the relation-

ship between preschool programs and their outcomes. Previous analysts (Cicarelli, Evans, & Schiller, 1969; Datta, 1976; Haskins, 1989; Herrnstein & Murray, 1994) who did not account for this key characteristic of program endowment erroneously emphasized the fading-out of compensatory program effects after a few years. In sharp contrast, the latest long-term evaluations demonstrate in fairly unequivocal fashion that the positive cognitive and related lifespace effects of more intensive early childhood educational programs are maintained well into adulthood.

This review's null findings probably contribute as much interpretive power to the integrative knowledge about the cognitive and related effects of early childhood education as do the significant ones. Perhaps most critically, major research design type—experiment or quasi-experiment—did not significantly confound or moderate the central hypothesized effects of this meta-analytic study. Because of the possibility that meta-analysis may merely compound erroneous inferences drawn from research that is not well controlled and because some analysts have observed the effects of nonrandomized trials to be systematically larger than those of otherwise similar randomized trials, numerous caveats have been offered regarding the use of meta-analysis with nonrandomized trials (Pettiti, 2000; Stroup et al., 2000). For these reasons, one could make a rational argument for the exclusion of quasi-experimental primary research from this meta-analytic study's sample. In this instance, however, such an exclusion criterion was not applied because empirical arguments were consistently discordant with the rational one. Not only did the design distinction of quasi-experiment versus randomized experiment not confound or moderate the observed aggregate effects of early childhood education, neither did other methodological characteristics such as study attrition rates or even a computed index of study internal validity. This study's exercise of due empirical caution provided ample assurance that its inclusion of quasi-experiments did not produce conclusions substantively different than those expected from the synthesis of randomized experiments only.

Possible Meta-Analytic Review Limitations

The decision to make published research this review's focus was based on the notion that professional peer and editorial review would bolster confidence in the validity of its aggregate findings. However, because it was based on published journal articles, books, and reports, the findings of this meta-analysis may be confounded by publication bias, although it seems that such intrusion is highly unlikely. As was noted, fail-safe N s at $p < .05$ for the intelligence, academic achievement, and other behavioral effects of preschool intervention were found to be 109, 148, and 231, respectively. These are the estimated number of studies with null findings that would have to exist in researchers' or practitioners' file drawers to change this review's central conclusions. The computed fail-safe N s are nearly five, nine, and six times the number of respective study outcomes included in this review ($N = 23, 17, \text{ and } 40$). This review seems highly

resistant to the potential impact of unretrieved null results, thus, publication bias is probably not a potent alternative explanation for its findings.

A systematic research review such as this one necessarily produces review-generated findings. Although sample of studies reviewed were rigorous experiments or quasi-experiments, at the level of meta-analysis the research design is essentially cross-sectional. Studies were sampled and analyzed at one point in time during January 2000. All of the review-generated findings of this meta-analysis are most appropriately thought of as screened hypotheses awaiting the confirmation or refutation of future primary research. However, given the homogeneously high internal validity of the studies analyzed, it also seems appropriate to label this review's conclusions as strong hypotheses that one can be quite confident will be affirmed with superior primary research methodologies.

Possible Policy and Research Implications

Some, such as *The Bell Curve's* authors (Herrnstein & Murray, 1994) and allied scholars, have disparaged "the spending of billions on run-of-the-mill" (p. 403) preschool programs, yet they claim the impossibility of implementing intensive, Abecedarian-like programs for all of the nation's disadvantaged children. Such a scientific and policy catch-22 is not helpful; as with many extreme notions, it is not consistent with the true complexity of the human condition. Such notions are based on a rather antiquated and overly simplistic main effects only model of cognitive development. Conversely, this review's interactional perspective, emphasizing the importance of both nature and nurture, rather than the relative importance of either nature or nurture, strongly suggests that the probable high heritability of complex behaviors such as intelligence notwithstanding, they can be highly malleable at the same time. It seems then that a balanced reading of its findings along with all of the extant scientific evidence would lead to the implementation policies that support, if not the very best imaginable, then at least better preschool programs for more, if not all, of the children who live in our most impoverished environments.

Such policy decisions are necessarily not only based on beneficiary considerations, but also costs. As taxpayers and their policy-making representatives will ultimately pay for all social programs, they want and should have information about their value. As for the relative benefits and costs of early childhood educational programs, the scientific evidence gathered so far strongly suggests that, in addition to their noted benefits for children through young adults, taxpayers definitely seem to be getting their money's worth. Consistent with the findings of this meta-analysis, for example, benefit-cost analyses of the Perry Preschool Program over 15 to 25 years have estimated that it returns between two and seven dollars to society for each dollar initially invested (Barnett, 1985; 1993). These are returns in the form of additional tax revenues from the lifetime incomes of preschool program participants, in addition to decreased social welfare, criminal justice, and related costs. Even if it is assumed that the more liberal of these estimates are doubt-

ful, a rational benefit-cost argument could certainly be made to double or even quadruple funding for compensatory education. Why not offer the very best preschool programs possible, up to the societal break-even point?

It seems quite clear what type of future research is needed in this field. A study is needed of diverse preschool interventions that includes the following characteristics: (a) experimental manipulation of the preschool control group variable; (b) a very large sample of participants; and (c) procedures that ensure near complete follow up of the original intervention and control samples. The present meta-analytic study showed that although most of this field's extant primary studies are categorically high in internal validity, including two of three of these research design characteristics, none of them has yet to incorporate all three rigorous procedures. Recall that most studies scored four or five of six points on a computed internal validity index (none achieved the maximum possible score of six). The power provided by very large, perhaps national samples, would allow not only for confident assurance of intervention control pretest equivalence, but also for the accomplishment of direct primary analysis of the potential moderating effect of various child and program characteristics. This would simultaneously allow for the potentially confounding or interacting effects of other key variables, such as various characteristics of the home environment. Moreover, although the research accomplished to date has demonstrated a clear positive relationship between program endowment and its effects (i.e., generally, the more money invested in a given preschool educational intervention, the larger will be its cognitive and other behavioral effects), we still know next to nothing about threshold effects or the relative importance of particular program investments. For example, what is the overall point of diminishing returns on investment in preschool education, above which further cognitive or other behavioral gains will not be realized? Is it \$3,000 per child-year, \$5,000, or \$10,000 or more? What are the threshold effects of specific program characteristics: teacher credentials, teacher/preschooler ratios, program intensity, and duration? These questions have great scientific and policy merit. Their answers, achievable with the best possible research, will be well worth society's investment.

REFERENCES¹

- Adams, M. J. (1989). Thinking skills curricula: Their promise and progress. *Educational Psychologist*, 24, 25–77.
- Bacharach, V. R., & Baumeister, A. A. (1998). Effect of maternal intelligence, marital status, income, and home environment on cognitive development of low birthweight infants. *Journal of Pediatric Psychology*, 23, 197–205.
- Barnett, W. S. (1985). Benefit-cost analysis of the Perry Preschool Program and its policy implications. *Educational Evaluation and Policy Analysis*, 7, 333–342.
- *Barnett, W. S. (1991). Benefits of compensatory preschool education. *Journal of Human Resources*, 27, 279–312.

¹References marked with an asterisk indicate studies included in the meta-analysis.

- Barnett, W. S. (1993). Benefit-cost analysis of preschool education: Findings from a 25-year follow-up. *American Journal of Orthopsychiatry*, *63*, 500–508.
- Barnett, W. S. (1998). Long-term cognitive and academic effects of early childhood education on children in poverty. *Preventive Medicine*, *27*, 204–207.
- Bouchard, T. J., Jr., Lykken, D. T., McGue, M., Segal, N. L., & Tellegen, A. (1990). Sources of human psychological differences: The Minnesota study of twins reared apart. *Science*, *250*, 223–228.
- Brooks-Gunn, J., Klebanov, P. K., & Duncan, G. J. (1996). Ethnic differences in children's intelligence test scores: Role of economic deprivation, home environment, and maternal characteristics. *Child Development*, *67*, 396–408.
- Brown, A. L., & Campione, J. C. (1982). Modifying intelligence or modifying cognitive skills: More than a semantic quibble? In D. K. Detterman & R. J. Sternberg (Eds.), *How much can intelligence be increased?* (pp. 215–230). Norwood, NJ: Ablex.
- *Burchinal, M. R., Campbell, F. A., Bryant, D. M., Wasik, B. H., & Ramey, C. T. (1997). Early intervention and mediating processes in cognitive performance of children of low-income African American families. *Child Development*, *68*, 935–954.
- *Campbell, F. A., Breitmayer, B., & Ramey, C. T. (1986). Disadvantaged single teenage mothers and their children: Consequences of free educational day care. *Family Relations*, *35*, 63–68.
- *Campbell, F. A., & Ramey, C. T. (1990). The relationship between Piagetian cognitive development, mental test performance, and academic achievement in high-risk students with and without early educational experience. *Intelligence*, *14*, 293–308.
- *Campbell, F. A., & Ramey, C. T. (1994). Effects of early intervention on intellectual and academic achievement: A follow-up study of children from low-income families. *Child Development*, *65*, 684–698.
- *Campbell, F. A., & Ramey, C. T. (1995). Cognitive and school outcomes for high-risk African-American students at middle adolescence: Positive effects of early intervention. *American Educational Research Journal*, *32*, 743–772.
- Canieuz, G. L., & Watkin, M. W. (1998). Long-term stability of the Wechsler Intelligence Scale for Children-third edition. *Psychological Assessment*, *10*, 285–291.
- Capron, C., & Duyme, M. (1989). Assessment of effects of socio-economic status on IQ in a full cross-fostering study. *Nature*, *340*, 552–554.
- Ceci, S. J. (1991). How much does schooling influence general intelligence and its cognitive components? A reassessment of the evidence. *Developmental Psychology*, *27*, 703–722.
- Christian, K., Morrison, F. J., & Bryant, F. B. (1998). Predicting kindergarten academic skills: Interactions among child care, maternal education, and family literacy environments. *Early Childhood Research Quarterly*, *13*, 501–521.
- *Cicarella, V. G., Evans, J. W., & Schiller, J. S. (1969). *The impact of Head Start: An evaluation of the effects of Head Start on children's cognitive and affective development*. Athens, OH: Westinghouse Learning Corporation and Ohio University.
- *Clark, S. H., & Campbell, F. A. (1998). Can intervention early prevent crime later? The Abecedarian project compared with other programs. *Early Childhood Research Quarterly*, *13*, 319–343.
- Coles, C. D., Brown, R. T., Smith, I. E., Platzman, K. A., Erickson, S., & Falek, A. (1991). Effects of prenatal alcohol exposure at school age. I. Physical and cognitive development. *Neurotoxicology and Teratology*, *13*, 357–367.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Mahwah, NJ: Erlbaum.
- Cooper, H. (1998). *Synthesizing research: A guide for literature reviews* (3rd ed.). Thousand Oaks, CA: Sage.
- Cooper, H., & Hedges, L. V. (Eds.) (1994). *The handbook of research synthesis*. New York: Russell Sage Foundation.

- *Copple, C. E., Cline, M. G., & Smith, A. N. (1987). *Path to the future: Long-term effects of Head Start in the Philadelphia school district*. Washington, DC: US Department of Health and Human Services.
- Cryns, A. G., & Gorey, K. M. (2000). *Rushton's impermeable belief in the validity of his own evolutionary paradigm may prevent his consideration of alternatives: An update meta-analysis and rejoinder to Rushton and Ankney*. Manuscript submitted for publication.
- *Currie, J., & Thomas, D. (1995). Does Head Start make a difference? *American Economic Review*, *85*, 341–364.
- Cyphers, L. H., Fulker, D. W., Plomin, R., & DeFries, J. C. (1989). Cognitive abilities in the early school years: No effect of shared environment between parents and offspring. *Intelligence*, *13*, 369–386.
- Dacey, C. M., Nelson, W. M., III, & Stoeckel, J. (1999). Reliability, criterion-related validity and qualitative comments of the fourth edition of the Stanford-Binet Intelligence Scale with a young adult population with intellectual disability. *Journal of Intellectual Disability Research*, *43*, 179–184.
- Daly, E. J., III, Wright, J. A., Kelly, S. Q., & Martens, B. K. (1997). Measures of early academic skills: Reliability and validity with a first grade sample. *School Psychology Quarterly*, *12*, 268–280.
- Datta, L. (1976). The impact of the Westinghouse/Ohio evaluation on the development of project Head Start: An examination of the immediate and longer-term effects and how they came about. In C. C. Abt (Ed.), *The evaluation of social programs* (pp. 129–181). Beverly Hills, CA: Sage.
- Devlin, B., Daniels, M., & Roeder, K. (1997). The heritability of IQ. *Nature*, *388*, 468–471.
- DiPerna, J. C., & Elliott, S. N. (1999). Development and validation of the Academic Competence Evaluation Scales. *Journal of Psychoeducational Assessment*, *17*, 207–225.
- Duncan, G. J., & Brooks-Gunn, J. (Eds.) (1997). *Consequences of growing up poor*. New York: Russell Sage Foundation.
- Duran, R., & Powers, S. (1993). Reliabilities of the Iowa Tests of Basic Skills for Hispanic- and Anglo-American students. *Psychological Reports*, *73*, 64–66.
- *Farnworth, M., Schweinhart, L. J., & Berrueta-Clement, J. R. (1985). Preschool intervention, school success and delinquency, in a high risk sample of youth. *American Educational Research Journal*, *22*, 445–464.
- Fleiss, J. L. (1981). *Statistical methods for rates and proportions* (2nd ed.). New York: Wiley.
- Flynn, J. R. (1984). The mean IQ of Americans: Massive gains 1932 to 1978. *Psychological Bulletin*, *95*, 29–51.
- Garber, H. L., & Hodge, J. D. (1991). Understanding intervention, inoculation, and risk for intellectual deceleration: A reply to Locurto. *Intelligence*, *15*, 317–323.
- Gorey, K. M. (1996). Social work intervention efficacy research: Comparison of the findings from internal versus external evaluations. *Social Work Research*, *20*, 119–128.
- Gorey, K. M., & Cryns, A. G. (1995). Lack of racial differences in behavior: A quantitative replication of Rushton's (1988) review and an independent meta-analysis. *Personality and Individual Differences*, *19*, 345–353.
- Gorey, K. M., & Cryns, A. G. (1999). The bell curve: Race, social class, and social work. *Social Work*, *44*, 586–589.
- Gould, S. J. (1995). Mismeasure by any measure. In R. Jacoby & N. Glaubergerman (Eds.), *The bell curve debate: History, documents, opinions* (pp. 3–13). New York: Times Books.
- *Gray, S. W., Ramsey, B. K., & Klaus, R. A. (1982). *From 3 to 20: The Early Training Project*. Baltimore: University Park Press.
- Haskins, R. (1989). Beyond metaphor: The efficacy of early childhood education. *American Psychologist*, *44*, 274–282.
- Hedges, L., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando, FL: Academic Press.

- Herrnstein, R. J., & Murray, C. A. (1994). *The bell curve: Intelligence and class structure in American life*. New York: Free Press.
- *Hubbell, V. R. (1978). The developmental continuity consortium study: Secondary analysis of early intervention data. In B. Brown (Ed.), *Found: Long-term gains from early intervention* (pp. 111–128). Boulder, CO: Westview Press.
- Jencks, C., & Phillips, M. (Eds.) (1998). *The black-white test score gap*. Washington, DC: Brookings Institution Press.
- Jensen, A. R. (1985). The nature of the black-white difference on various psychometric tests: Spearman's hypothesis. *Behavioral and Brain Sciences*, *8*, 193–263.
- Jensen, A. R. (1993). Psychometric g and achievement. In B. R. Gifford (Ed.), *Policy perspectives on educational testing* (pp. 117–127). Boston: Kluwer Academic Publishers.
- Johnson, D. L., Swank, P. R., Baldwin, C. D., & McCormick, D. (1999). Adult smoking in the home environment and children's IQ. *Psychological Reports*, *84*, 149–154.
- *Jordan, T. J., Gallo, R., Deutsch, M., & Deutsch, C. P. (1985). Long-term effects of early enrichment: A 20-year perspective on persistence and change. *American Journal of Community Psychology*, *13*, 393–415.
- Kaplan, C. (1996). Predictive validity of the WPPSI-R: A four year follow-up study. *Psychology in the Schools*, *33*, 211–220.
- Karoly, L. A., Greenwood, P. W., Everingham, S. S., Houbé, J., Kilburn, M. R., Rydell, C. P., Sanders, M., & Chiesa, J. (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.
- Karr, S. K., Carvajal, H. H., Elser, D., Bays, K., Logan, R. A., & Page, G. L. (1993). Concurrent validity of the WPPSI-R and the McCarthy Scales of Children's Abilities. *Psychological Reports*, *72*, 940–942.
- Krohn, E. J., & Lamp, R. E. (1999). Stability of the SB:FE and K-ABC for a group of children from low-income families: A 5-year longitudinal study. *Journal of School Psychology*, *37*, 315–332.
- Larroque, B., & Kaminski, M. (1998). Prenatal alcohol exposure and development at preschool age: Main results of a French study. *Alcoholism: Clinical and Experimental Research*, *22*, 295–303.
- *Lee, V. E., Brooks-Gunn, J., Schnur, E., & Liaw, F. (1991). Are Head Start effects sustained? A longitudinal follow-up comparison of disadvantaged children attending Head Start, no preschool, and other preschool programs. *Annual Progress in Child Psychiatry and Child Development*, 600–618.
- Lehrman, D. S. (1970). Semantic and conceptual issues in the nature-nurture problem. In L. R. Aronson, E. Tobach, D. S. Lehrman, & J. S. Rosenblatt (Eds.), *Development and evolution of behavior: Essays in memory of T. C. Schneirla* (pp. 17–52). San Francisco: Freeman.
- Locurto, C. (1990). The malleability of IQ as judged from adoption studies. *Intelligence*, *14*, 275–292.
- Locurto, C. (1991). Beyond IQ in preschool programs? *Intelligence*, *15*, 295–312.
- *Luster, T., & McAdoo, H. (1996). Family and child influence on educational attainment: A secondary analysis of the High/Scope Perry Preschool data. *Developmental Psychology*, *32*, 26–39.
- *Madden, J., O'Hara, J., & Levenstein, P. (1984). Home again: Effects of the Mother-Child Home Program on mother and child. *Child Development*, *55*, 636–647.
- *Martin, S. L., Ramey, C. T., & Ramey, S. (1990). The prevention of intellectual impairment in children of impoverished families: Findings of a randomized trial of educational day care. *American Journal of Public Health*, *80*, 844–847.
- McLaughlin, D. H. (1977). *Title I, 1965–1975: A synthesis of the findings of federal studies*. Palo Alto, CA: American Institutes for Research.
- *Miller, L. B., & Bizzell, R. P. (1984). Long-term effects of four preschool programs: Ninth- and tenth-grade results. *Child Development*, *55*, 1570–1587.
- Nisbett, R. E. (1998). Race, genetics, and IQ. In C. Jencks & M. Phillips (Eds.), *The black-white test score gap* (pp. 86–102). Washington, DC: Brookings Institution Press.

- Petitti, D. B. (2000). *Meta-analysis, decision analysis, and cost-effectiveness analysis: Methods for quantitative synthesis in medicine* (2nd ed.). New York: Oxford University Press.
- Powers, S., Escamilla, K., & Haussler, M. M. (1986). The California Achievement Test as a predictor of reading ability across race and sex. *Educational and Psychological Assessment, 46*, 1067–1070.
- Ramey, S. L. (1999). Head Start and preschool education: Toward continued improvement. *American Psychologist, 54*, 344–346.
- *Ramey, C. T., & Campbell, F. A. (1984). Preventive education for high-risk children: Cognitive consequences of the Carolina Abecedarian Project. *American Journal of Mental Deficiency, 88*, 515–523.
- *Ramey, C. T., & Ramey, S. L. (1998). Prevention of intellectual disabilities: Early intervention to improve cognitive development. *Preventive Medicine, 27*, 224–232.
- Raver, C. C., & Zigler, E. F. (1997). Social competence: An untapped dimension in evaluating Head Start's success. *Early Childhood Research Quarterly, 12*, 363–385.
- *Reynolds, A. J. (1994). Effects of a preschool plus follow-on intervention for children at risk. *Developmental Psychology, 30*, 787–804.
- *Reynolds, A. J., & Temple, J. A. (1995). Quasi-experimental estimate of the effect of a preschool intervention: Psychometric and econometric comparisons. *Education Review, 19*, 347–373.
- *Reynolds, A. J., & Temple, J. A. (1998). Extended early childhood intervention and school achievement: Age thirteen findings from the Chicago Longitudinal Study. *Child Development, 69*, 231–246.
- 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.
- Rosenthal, R. (1978). Combining results of independent studies. *Psychological Bulletin, 85*, 185–193.
- Rosenthal, R. (1979). The “file drawer problem” and tolerance for null results. *Psychological Bulletin, 86*, 638–641.
- Rothman, K. J., & Greenland, S. (1998). *Modern epidemiology* (2nd ed.). Philadelphia: Lippincott Raven.
- *Royce, J. M., Lazar, I., & Darlington, R. B. (1983). Minority families, early education, and later life chances. *American Journal of Orthopsychiatry, 53*, 706–719.
- Rushton, J. P. (1995). *Race, evolution, and behavior: A life history perspective*. New Brunswick, NJ: Transaction Publishers.
- Rushton, J. P., & Ankney, C. D. (1995). Fallacious analyses cannot test for racial differences: A reply to Gorey and Cryns. *Personality and Individual Differences, 19*, 355–357.
- Scarr, S., & Weinberg, R. A. (1978). The influence of “family background” on intellectual attainment. *American Sociological Review, 43*, 674–692.
- Scarr, S., & Weinberg, R. A. (1983). The Minnesota adoption studies: Genetic differences and malleability. *Child Development, 54*, 260–267.
- Schwartz, J. (1994). Low-level lead exposure and children's IQ: A meta-analysis and search for a threshold. *Environmental Research, 65*, 42–55.
- *Schweinhart, L. J., & Wallgren, C. R. (1993). Effects of a Follow Through Program on school achievement. *Journal of Research in Childhood Education, 8*, 43–56.
- Schweinhart, L. J., & Weikart, D. P. (1991). Response to “Beyond IQ in preschool programs?” *Intelligence, 15*, 313–315.
- *Schweinhart, L. J., & Weikart, D. P. (1993). Success by empowerment: The High/Scope Perry Preschool study through age 27. *Young Children, 49*, 54–58.
- *Schweinhart, L. J., Weikart, D. P., & Larner, M. B. (1986). Consequences of three preschool curriculum models through age 15. *Early Childhood Research Quarterly, 1*, 15–45.
- Shull, S. S., Weatherly, M., Morgan, S. K., & Bradley, J. S. (1995). Stability reliability for elementary-age students on the Woodcock-Johnson Psychoeducational Battery-Revised (Achieve-

- ment section) and the Kaufman Test of Educational Achievement. *Psychology in the Schools*, *32*, 89–92.
- Slate, J. R. (1995). Two investigations of the validity of WISCII. *Psychological Reports*, *76*, 299–306.
- Slate, J. R., & Jones C. H. (1995). Preliminary evidence of the WISC-III for African American students undergoing special education evaluation. *Educational and Psychological Measurement*, *55*, 1039–1046.
- Spitz, H. H. (1997). Some questions about the results of the Abecedarian early intervention project cited by the APA task force on intelligence. *American Psychologist*, *52*, 72.
- *Sprigle, J. E., & Schaefer, L. (1985). Longitudinal evaluation of the effects of two compensatory preschool programs on fourth- through sixth-grade students. *Developmental Psychology*, *21*, 702–708.
- Stroup, D. F., Berlin, J. A., Morton, S. C., Olkin, I., Williamson, G. D., Rennie, D., Moher, D., Becker, B. J., Sipe, T., & Thacker, S. B. (2000). Meta-analysis of observational studies in epidemiology: A proposal for reporting. *Journal of the American Medical Association*, *283*, 2008–2012.
- Tambs, K., Sundet, J. M., Magnus, P., & Berg, K. (1989). Genetic and environmental contributions to the covariance between occupational status, educational attainment, and IQ: A study of twins. *Behavior Genetics*, *19*, 209–222.
- Turkheimer, E. (1991). Individual and group differences in adoption studies of IQ. *Psychological Bulletin*, *110*, 392–405.
- *Tzuriel, D., Kaniel, S., Kanner, E., & Haywood, H. C. (1999). Effect of the “Bright Start” program in kindergarten on transfer and academic achievement. *Early Childhood Research Quarterly*, *14*, 111–141.
- Vincent, K. R. (1991). Black/white IQ differences: Does age make the difference? *Journal of Clinical Psychology*, *47*, 266–270.
- *Wasik, B. H., Ramey, C. T., Bryant, D. M., & Sparling, J. J. (1990). A longitudinal study of two early intervention strategies: Project CARE. *Child Development*, *61*, 1682–1696.
- *Weikart, D. P. (1998). Changing early childhood development through educational intervention. *Preventive Medicine*, *27*, 233–237.
- Weinberg, R. A., Scarr S., & Waldman, I. D. (1992). The Minnesota transracial adoption study: A follow-up of IQ test performance at adolescence. *Intelligence*, *16*, 117–135.
- *Whitehurst, G. J., Zevenbergen, A. A., Crone, D. A., Shultz, M. D., Velting, O. N., & Fischel, J. E. (1999). Outcomes of an emergent literacy intervention from Head Start through second grade. *Journal of Educational Psychology*, *91*, 261–272.
- Wilson, E. O. (1998). *Consilience: The unity of knowledge*. New York: Vintage.
- *Wright, M. J. (1983). *A Canadian approach: Compensatory education in the preschool*. Ypsilanti, MI: The High/Scope Press.
- *Wu, P., & Campbell, D. T. (1996). Extending latent variable LISREL analyses to the 1969 Westinghouse Head Start evaluation to blacks and full year whites. *Evaluation and Program Planning*, *19*, 183–191.

Action Editor: Terry B. Gutkin

Acceptance Date: August 4, 2000