Contextual Effects of Prekindergarten Classrooms for Disadvantaged Children on Cognitive Development: The Case of Chapter 1

Valerie E. Lee; Susanna Loeb; Sally Lubeck

Child Development, Vol. 69, No. 2 (Apr., 1998), 479-494.

Stable URL:
http://links.jstor.org/sici?sici=0009-3920%28199804%2969%3A2%3C479%3ACEOPCF%3E2.0.CO%3B2-2

Child Development is currently published by Society for Research in Child Development.

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at http://www.jstor.org/about/terms.html. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at http://www.jstor.org/journals/srcd.html.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is an independent not-for-profit organization dedicated to creating and preserving a digital archive of scholarly journals. For more information regarding JSTOR, please contact support@jstor.org.

http://www.jstor.org/
Wed Jun 8 11:00:20 2005

Child Development, April 1998, Volume 69, Number 2, Pages 479–494

Contextual Effects of Prekindergarten Classrooms for Disadvantaged Children on Cognitive Development: The Case of Chapter 1

Valerie E. Lee, Susanna Loeb, and Sally Lubeck

This study explores the effects of the social context of Chapter 1 prekindergarten classrooms on children's learning. Chapter 1 (also called Title I) is a federal government preschool program directed at children in low-income schools who are at risk of later school failure. Using hierarchical linear modeling (HLM) and a sample of 677 4-year-olds in 55 1990–1991 Chapter 1 prekindergarten classes in 5 states, the study explores factors that influence gains on the Preschool Inventory (PSI) over the preschool year. Social context is defined here mainly in terms of the cognitive and social composition of the classroom. Contextual factors defined in terms of demographics are shown to be related to learning, but the average cognitive level of the class is not. On average, children learn less in classrooms with high concentrations of minorities, children with special needs, recent immigrants, and children whose mothers have little education. The study explores differential effects of racial concentration on race differences in learning. Policy implications are discussed.

INTRODUCTION

This article describes a study of the differential effects of Chapter 1 prekindergartens on the cognitive development of the children who attend them. Studies of preschool programs for children from economically disadvantaged families have focused almost entirely on Head Start. Evaluations typically have compared the cognitive and social development of program participants to that of demographically equivalent children without preschool experience. These studies have ignored other publicly supported preschool programs for disadvantaged children. In addition, many of the studies have used small and local samples, seldom considering the possibility of differential effects in a wider sample of children and settings. Moreover, and perhaps more importantly, the evaluations have viewed the preschool program as a "black box," with children either experiencing it or not, but with "it" remaining unexplored.

Very little research has considered the context or environment of preschools, an oversight with both methodological and conceptual origins. Researchers typically have viewed preschool participation as a "treatment" for individuals, not recognizing that children experience such programs in groups. Failure to consider contexts has substantially limited our understanding of the effects of these interventions, as well as excluding from study the possibility that the same classroom has different effects, depending on children's relative social and intellectual status. Rather than the "efficacy" approach to studying disadvantaged children's early schooling, we approach the subject by focusing on the social context of preschools. We examine how contextual differences between classrooms influence cognitive development over the course of the preschool year.

Using data from a large study sponsored by the U.S. Department of Education, we address several of the shortcomings described above. Rather than Head Start, we focus on the prekindergarten (pre-K) component of the major federally funded compensatory education program, Chapter 1—in essence, the federal government's *other* preschool intervention.[1] We look within Chapter 1 pre-K classrooms to capture several elements of their context, using longitudinal data on a large and geographically diverse sample (677 children in 55 preschool classrooms in five states). We employ a research design that captures the grouped nature of the preschool experience and statistical methods that are designed to evaluate group effects on individuals. Our analysis of classroom context focuses on demographic, academic, and social composition, although we also explore such el-

1. Chapter 1 represents the federal government's major program for compensatory education. The large majority of Chapter 1 funds, which are allocated first to schools with substantial proportions of poor children and then targeted at children with low measured performance, flow to elementary schools. The very large majority of U.S. school districts (over 90%) have at least one of these schools and thus receive Chapter 1 funds. Much smaller amounts accrue to high school Chapter 1 programs. Even smaller amounts go to preschool programs, and individual school districts may decide whether or not to provide Chapter 1 preschools.

© 1998 by the Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/98/6902-0009$01.00

ements as the economic conditions of neighborhoods, class size, leadership characteristics of program directors, and location.

BACKGROUND

The Context of Preschool Classrooms

What is context? We define context broadly as the elements of a preschool program that make a difference in children's cognitive or social development. Context, as such, can take many forms, only a small subset of which we explore. In particular, we focus on classroom composition (based on the social and cognitive characteristics of children) and neighborhood. Our study also examines such contextual elements as curriculum (e.g., what is taught and by whom), structure (e.g., full-day versus half-day, full-year versus 9 month year, or mixed age grouping), and the resource base (e.g., teacher qualifications, staff-child ratio, annual expenditures per child), although these elements are not our major focus.

The research base. Although a few researchers have examined the preschool classroom context by assessing curricular, interactive, or instructional dynamics (Bronson, Tivnan, & Seppanen, 1995; Frede & Barnett, 1992; Stipek, Daniels, Galluzzo, & Milburn, 1992), studies with a primary focus on the context of preschool programs for economically disadvantaged children are rare.

An example of research that considers (but does not focus on) context is a longitudinal study of 1,539 African American and Hispanic Chicago children, most of whom had experience in federally funded preschool programs (Reynolds, 1991). With four waves of data, from pre-K, kindergarten, and grades 1 and 2, Reynolds used a complex path analytic design (but not multilevel methods). The study took account of two contextual factors: the socioeconomic status (SES) of the school (not the classroom) and the achievement level of the child's kindergarten class. An earlier study by Reynolds (1989) investigated the effects of the same contextual/compositional measures on reading, mathematics, and socioemotional maturity outcomes in grade 1. The two Reynolds studies, which included context measures for statistical control purposes, found indirect (but not direct) effects of these measures on student learning.

Another study, in one school district (Pittsburgh), considered neighborhood SES as a covariate in studying children's academic success in elementary school (Bickel, Zigmond, & Strayhorn, 1991). Although neighborhood SES was statistically related to first-grade achievement, it seems that this contextual element was used mainly as a proxy for more precise knowledge of family economic condition, which was not included in the model. Because the research base on preschool social context is thin, we expand our discussion of existing research on educational contexts.

School Effects Research

Conceptualization. Beginning in the late 1970s, sociologists renewed their interest in how schools influence their students. A conceptual and empirical link was forged between classroom processes and school organization. Schools began to be seen as internally differentiating institutions. This work views social context through two lenses: (1) externalities that influence schools (e.g., community or neighborhood conditions, nature of parent/community involvement, location) and (2) the internal organization of schools (e.g., student social and academic composition, size, curriculum, facilities, classroom structure, and climate). Both types of context influence both students and teachers.

Although some important school effects studies have been conducted in elementary schools (e.g., Barr & Dreeben, 1983), most of this research focuses on high schools. Lee, Bryk, and Smith (1993) review this work. Most early empirical research focused on linking external factors and student outcomes (especially achievement) directly. These factors also affect students indirectly, through their impact on schools. That is, contextual elements influence the internal organization of schools, which, in turn, affects teacher and student outcomes. As new statistical methods to conceptualize these models have become available, the number and complexity of empirical studies in this genre have grown recently. However, the theoretical literature on context effects is not new (e.g., Bidwell, 1972; Bidwell & Kasarda, 1980; Brown & Saks, 1985).

School effects research investigates how school organization and context can influence both the operation of schools and, more importantly, how students develop. Much of this research focuses on students' social or academic disadvantage and how structural characteristics of the school can influence social differentiation of educational outcomes within schools. We focus further discussion on studies of school composition, particularly ability and racial composition.

Ability composition. In a well-known study of working-class secondary schools in London, Rutter, Maughan, Mortimore, Outson, and Smith (1979) concluded that schools need a core of motivated and academically able students to provide a stable base for instruction. A nucleus of children of at least average ability characterized schools showing better examination scores and lower delinquency rates. Other work has focused on how compositional factors influence how teachers organize instruction. In their study of reading in first-grade classrooms, Barr and Dreeben (1983) illustrated how the characteristics of students influence teachers' work. High proportions of low-ability students were shown to make it more difficult for teachers to organize effective reading instruction. Barr and Dreeben suggest that classroom composition (students' background, ability, and interests) represents a direct resource for student learning.

Vacancy theory provides one explanation for the relation between a school's ability composition and a student's opportunity to learn. The relative competence of the classmates with whom a student must compete determines, to some extent, his or her access to a desirable educational environment (Sorensen, 1987). However, several studies have shown that the structure of the school, as well as the composition of the student body, is important. Several studies of high schools have demonstrated that varying structures respond differently to students with similar entry characteristics (e.g., Garet & Delaney, 1988; Hallinan & Sorenson, 1983; Lee & Bryk, 1988, 1989). That is, some high schools are considerably more effective in providing an equitable learning environment to students with academic and social disadvantage than others.

Racial composition. The findings of Coleman and his colleagues (1966) that the achievement of minority students was higher in racially integrated schools, combined with a national commitment to the racial integration of schools 2 or 3 decades ago, stimulated research on the relation between school racial composition and student achievement. A review of the desegregation and achievement literature by Mahard and Crain (1983) echoed Coleman et al.'s (1966) findings. However, another important review (Schofield & Sagar, 1983) reported that intergroup relations were better in schools with higher proportions of minorities. Case studies of newly desegregated schools provide evidence of significant changes in the organization of schools. One response to integration was to add nondemanding courses to the curriculum, thereby encouraging the resegregation of students within schools (Cusick, 1983; Eyler, Cook, & Ward, 1983; Grant, 1988).[2]

The Chapter 1 Education Program

An alternative preschool program for disadvantaged children. Chapter 1, formerly called Title I (and now again using this label), represents the major federal investment in compensatory education. Through Title I of the 1965 Elementary and Secondary Education Act (ESEA), Congress sought to narrow the achievement gap between economically advantaged and disadvantaged students by targeting funds to low achieving students in schools with high proportions of low-income students. The program served five million children at a cost of $6.1 billion in 1992 (Madaus, 1993). Sinclair and Gutmann (1990) estimated that 42% of children receiving Chapter 1 services in 1987–1988 came from low-income families and that 58% were considered "educationally deprived" relative to other children in the same districts.

Aimed at "prevention" rather than "remediation," pre-K was included in the Title I/Chapter 1 legislation from its inception, although preschool programs represent only a small part of the federal Chapter 1 initiative (Sinclair & Guttmann, 1990).[3] In 1988–1989, only 1.5% of the children receiving Chapter 1 services were preschoolers (LeTendre, 1991). Only about 10% of school districts receiving Chapter 1 funds support pre-K or full-day kindergarten programs (Seppanen, Godin, Metzger, Bronson, & Cichon, 1993). Although the numbers served are relatively small compared to Head Start, Chapter 1 pre-Ks constitute an important alternative form of publicly supported preschool education because, unlike Head Start, they are staffed by certified teachers and are typically housed in public schools.

Despite such obvious and important policy variations, there is almost no research either evaluating Chapter 1 preschools or comparing these two federally funded programs for preschool-aged children. Moreover, the small volume of research on Chapter 1 pre-K programs is of weak design. Most is in the form of program descriptions (e.g., Shiminski, 1992; Trangmoe, 1988), accounts from the National Diffusion Network, or local process/product evaluations

2. By contrast, Bryk, Lee, and Holland (1993), investigating the role of the curriculum in providing opportunities to students of varying abilities, showed that Catholic high school students evidence greater academic preparation, in large part because of the limited curricular offerings in these schools, coupled with adult encouragement of academic pursuits by all students.

3. In 1986–87, 57,200 were enrolled; in 1987–88, 68,200.

(Claus & Quimper, 1991, 1992; New York City Board of Education, 1989). These studies do not appear in peer-reviewed journals, and results are frequently inconclusive (May & Farha, 1989; Seawell & Ross, 1992; Sevigny, 1987).

A federally funded study of Chapter 1 preschools. To compensate for this lack and to assess Chapter 1 program effects more broadly, the U.S. Department of Education funded a large cross-site study of Chapter 1 early childhood programs in 1990–1991. Seppanen and her colleagues (1993) studied 55 pre-K classes funded solely or in part by Chapter 1 funds. Two purposes guided the study: (1) to evaluate effects of program characteristics on child outcomes and (2) to assess how children's experiences in Chapter 1 pre-K programs changed between prekindergarten and kindergarten. Data collection included observations and interviews as well as pre- and posttests of children's cognitive and behavioral development over the course of the preschool year. There was no control group; all study participants were in Chapter 1 preschool classrooms. The study described here is a secondary analysis of data from the first year of the Seppanen et al. (1993) study.

In general, the Chapter 1 preschool study did not realize its major purpose. The Seppanen et al. report (1993) found almost no effects of classroom characteristics on children's learning during the preschool year. However, another study using the same data but only bivariate correlations (rather than multivariate analyses) reported associations with change in task behavior of two classroom conditions related to observations of time spent on (1) cognitive concepts and (2) activities devoted to mathematics and language (Bronson et al., 1995). The majority of observed classroom activities considered in that study were uncorrelated with learning. Neither study used methods that captured the nested nature of the data and the questions (children in classrooms).

Research Questions

Our aim in this study is to improve on the Seppanen et al. (1993) study in several respects using the same data. We do not assess the efficacy of Chapter 1 preschools on program participants, even though we examine children's cognitive development over the preschool year. Rather, we investigate how particular elements of the classroom context influence participant children's learning. Thus, the study design is not framed within the traditional comparative framework, where participants are compared to nonparticipants. Instead, we make comparisons among pre-K classrooms. Within this framework, we pose three research questions, with the answer to each depending on a positive answer to the previous one.

Question 1. Does learning vary systematically between classrooms? We hypothesize that even when the cognitive and social characteristics of participants are taken into account, children's learning varies between preschool classrooms.

Question 2. Do contextual elements of the preschool classrooms affect student learning? We hypothesize that classroom context, defined in both cognitive and social terms, influences systematic differences in learning between preschool classrooms.

Question 3. Are there differential contextual effects in the same classroom? We hypothesize that classroom racial composition will differentially influence the learning of children of different race/ethnicity.

METHOD

Sample and Data

Data. Data for the study were drawn from the *Observational Study of Early Childhood Programs,* a research effort begun in 1989 by the U.S. Department of Education (Seppanen et al., 1993, p. xi). We use one of the two related research efforts that constituted the observational study: a study of 55 pre-K classrooms whose financial support came totally or partially from Chapter 1 funds. A larger descriptive study conducted in more than a hundred preschool classrooms (of which "our" 55 are a substudy) was concerned chiefly with describing interactions in classrooms and did not include the collection of outcome data. A major purpose of the substudy, on the other hand, was to investigate the association between class characteristics and child outcomes. Thus, the researchers collected information on children's cognitive and behavioral development in those pre-K classes.

Sample. The sampling strategy, intended to maximize variability between classes, included stratification on several levels: state, school district, school, and classroom. The first level included five states (California, Florida, Michigan, New Jersey, and Texas) that provided geographic diversity. In each state, from two (California and Florida) to five (Michigan) school districts were selected from among those offering the pre-K option within Chapter 1 programs. This yielded a total of 16 school districts. Between one and six schools were selected in each district, based on available Chapter 1 pre-K, for a total of 49 schools. Most sampled schools offered a single Chapter 1 pre-K class. This resulted in 55 classrooms, each with a different teacher. Because our study focused

on children's cognitive development, our sample included only children in sampled classes with scores on both the fall 1990 and spring 1991 administration of the Preschool Inventory (PSI). This selection resulted in a sample of 677 children of the 783 originally drawn (86.5%).[4] Within-class samples averaged 12.3 children. A large majority came from low-income homes.

The classes had a strong human resource base. Teachers were well educated (all had bachelor's degrees and 47% had master's degrees). All reported having been trained for teaching in early childhood classrooms, 71% had state licensure in early childhood, and 87% were certified in elementary education. Most classrooms had an assistant or aide (93%). The average teacher had almost a decade of experience in early childhood teaching and had taught for over 6 years in the current program. Except for a modest bias against rural location, the sample is reasonably representative of children and classrooms in all Chapter 1 pre-K programs in the 1990–1991 school year. Seppanen et al. (1993) provide full information about the sampling rationale, strategy, and limitations.

Measures

Variable sources. We selected variables from an array of measures in the original study. Information about children came from two sources: (1) three assessment scores (a test of cognitive achievement measured at two timepoints and a behavioral assessment measured at the beginning of the preschool year) and (2) descriptions of children and families provided by teachers. Data on classes came from three sources: (1) descriptions of classrooms, drawn from interviews with teachers and program directors, (2) classroom aggregates of data about children, and (3) data from the 1990 Census about school districts, drawn from the U.S. Department of Education's Common Core of Data.

The dependent variable. As our outcome, we used the spring 1991 administration of the Preschool Inventory (PSI), a 32-item assessment designed specifically for measuring the intellectual competency of disadvantaged children from 3 to 6 years. Meisels (1987) describes tests like the PSI as readiness tests, intended to assess child characteristics as they enter school rather than to predict child outcomes later in school. The PSI was developed to evaluate Head Start. Its psychometric properties are respectable. Reliability is high (.86 to .92; Caldwell & Freund, 1980). Consistency across time, represented by the correlation between pre- and posttest scores in this sample, is also strong, $r(676) = .73$. Although middle-class children might evidence ceiling effects on the PSI, target children (similar to our sample) seldom do. Another study that examined 1-year cognitive gain for disadvantaged preschoolers showed larger effects on the PSI than more common cognitive assessments such as the Peabody Picture Vocabulary Test (Lee, Brooks-Gunn, & Schnur, 1988). More detail about the PSI, as well as the other variables used in this study, is provided in the Appendix.

Independent variables describing children. Our measures describe children along four dimensions: cognitive status, behavior, demographic and family characteristics, and special needs. We include the fall 1990 PSI as a measure of preprogram cognitive ability. We also utilized children's scores on the Child Behavior Rating Scale (CBRS), a composite of 32 items assessing the child's social and task behavior scored by the teacher in fall 1990. Demographics include age (in months), gender, race/ethnicity, and mother's education. We divided children into three categories: African American, Hispanic, and European American. With two dummy-coded variables, we compare African Americans and Hispanics separately to European Americans.[5] Because the distribution of mother's education was positively skewed for this sample, we created three categories based on the highest level of education obtained: less than a high school education, a high school diploma, and education beyond high school. We created two dummy variables, separately comparing the first two to the third. Unfortunately, there was no information on family income. Teachers indicated whether each child or his or her family had special needs beyond the academic criteria that would qualify him or her for Chapter 1 (e.g., whether they evidenced behavioral or emotional problems, learning or psychological problems, or substance abuse). We created separate dummy variables for children's and families' special needs.

Variables describing classrooms. In selecting classroom measures, we dropped variables that we found had little between-class variability or virtually no effect in the Seppanen et al. (1993) study. Thus, we considered, and subsequently eliminated, such

4. The attrition rate (13.5%), a bit higher than the researchers had expected, prompted them to examine whether those children not in the classrooms at the time of posttesting were different from the initial sample in terms of cognitive achievement or child/family background characteristics. They were not (Seppanen et al., 1993, p. 74).

5. As only a small proportion of the sample (7%) fell into "other" ethnic categories (i.e., Asians, Native Americans), these children were grouped with European Americans (the comparison group).

potentially interesting context factors as parent involvement, length of the class day (most were half-day), the frequency and schedule of classroom activities, teachers' self-reported management behaviors, physical resources, and adult/child ratio. Most of our final context measures were aggregates of characteristics of children in each class. For example, an aggregate of the fall 1990 PSI score measured the ability context of the class. The construct of classroom social composition was tapped with aggregates that measured the proportions of children in each class (1) whose mothers had less than a high school education; (2) who were recent immigrants; (3) who needed extra help for behavioral, emotional, or mental problems; and who were (4) African American and (5) Hispanic.

We linked 1990 census data about each district to our data file of classrooms to consider other aspects of the social context: proportions of residents living in poverty and on public assistance (55 classrooms in 16 districts). Other measures tapped other classroom characteristics: for example, teachers' assessment of the effectiveness of leadership of the school's principal and class size. We examined systematic variability by state with four dummy-coded indicators to tap the five states where the sampled schools and classrooms were located. Classroom variables are described more fully in the Appendix.

Analytic Approach

Measuring change over time. There are two common alternatives to exploring cognitive change over time or learning. Some researchers (e.g., Bereiter, 1963; Linn & Slinde, 1977) favor a covariance approach, where analytic models use the posttest as an outcome and the pretest as a predictor. Other experts on measuring change over time (e.g., Rogosa & Willett, 1985; Willett, 1988, 1994) favor simple gain scores (posttest minus pretest). We decided on the covariance approach. Unless the scores are subjected to some sort of scaling to put them on a true interval scale, gain scores are typically negatively correlated with initial status (so the children who started the lowest artifactually "gain" the most). We were unable to rescale the PSI scores ourselves, as item-level data were unavailable.[6]

Descriptive analysis. We describe the sample to familiarize readers with characteristics of children and classrooms in the Chapter 1 pre-K program. Means and standard deviations of continuous variables and proportions for each category for categorical variables are presented. For variables describing children, means were computed for 677 children. Classroom means were computed for the 55 sampled classrooms. The variables are presented in their original metric in descriptive analyses, but in multivariate analyses continuous variables were rescaled to z scores ($M = 0$, $SD = 1$), and all categorical variables were dummy coded (0, 1).

Multilevel multivariate analyses. Because our data are nested and our research questions are multilevel, we employ Hierarchical Linear Models (HLM; Bryk & Raudenbush, 1992). The first step in any HLM is to partition the variance in the dependent variable into its within- and between-group components and to estimate the proportion of its variance that lies systematically between groups (here, classrooms). In our application of HLM, the next step is to estimate a set of regression coefficients in each classroom (level 1). These coefficients become outcomes to be simultaneously explained as functions of measured context differences between classrooms (level 2). Readers unfamiliar with (or interested in) more detail about a substantive application of HLM to an educational problem are referred to Bryk and Raudenbush (1992) or Lee and Bryk (1989).

RESULTS

Description of the Sample

The children. Descriptive information on children and classrooms is presented in Table 1. There appear to be neither ceiling nor floor effects on either the pretest or posttest PSI. On average, children score 20.3 on the 32-item test at the end of the year, and 15.5 at the beginning of the year, averaging about 5 points of gain (a 31% improvement). Standard deviations for the PSI at both timepoints are close to 5 points. On the CBRS, children average about 4 points, with a small *SD*. Equally represented by gender, sample children were on average close to 4 years old at the

6. We were predisposed to analyze these data with gain scores as outcomes. Unfortunately, we could not obtain (from the RMC Corporation) item-level data for the PSI scores. We were, thus, unable to subject the PSI scores to a scaling procedure (such as Item Response Theory or Rasch scaling) that would correct for the difficulties associated with simple gain-score analysis (particularly regression to the mean). When we explored a simple gain score analysis with a similar HLM analytic model, nevertheless, we found that the PSI gain was rather strongly correlated (negatively) with the pretest. Although a gain score outcome has considerable construct validity (i.e., it measures exactly what we want to measure: learning), no other measure of cognitive ability was available to control for the bias toward larger gains for initially lower-scoring children. Reluctantly, we abandoned this strategy.

Table 1 Descriptive Statistics for Variables Considered in Preschool Context Study

Variable Name	*M*	*SD*
Dependent variable (*N* = 677 children):		
Posttest (spring PSI)	20.34	4.95
Independent variables measured on children (*N* = 677 children):		
Pretest score (fall PSI)	15.53	5.16
Behavior score (fall CBRS)	3.90	.54
Age (in months)	46.71	5.54
Race (%):		
African American	47.56	...
Hispanic	30.58	...
European American	21.86	...
Gender (%):		
Female	51.56	...
Male	48.44	...
Mother's education (%):		
Less than high school	25.26	...
High school graduate	40.51	...
Beyond high classroom	34.23	...
Free or reduced lunch (%):		
Eligible	90.75	...
Not eligible	9.25	...
Child needed help (%):		
Needs help	12.20	...
Does not need help	81.80	...
Family needed help (%):		
Needs help	14.16	...
Does not need help	85.84	...
Independent variables measured on classrooms (*N* = 55 classes):		
Average pretest	15.34	2.15
Percent low mother's education	26.77	19.87
Percent children needing help	12.24	10.61
Percent recent immigrants	5.77	13.15
Percent poverty	18.63	12.64
Percent welfare	12.30	9.87
Percent African American	43.31	35.26
Percent Hispanic	27.23	34.15
Principal leadership	−.06	1.03
Class size	19.13	4.35
State location (%):		
New Jersey	18.18	...
California	18.18	...
Florida	16.36	...
Texas	23.64	...
Michigan	23.64	...

start of the preschool year. Most were minority: 48% African American and 31% Hispanic.

Reflecting the target group for Chapter 1, sample children were socially disadvantaged in several respects. Almost all qualified, by virtue of low income, for free or reduced lunch. Many mothers were not highly educated. Teachers rated some children as needing some outside help for noncognitive behaviors (beyond what could be provided in the classroom). They also classified some families as needing outside help for some type of family problem (such as substance abuse or substantial economic need).

The classrooms. The fact that two seminal studies of school effects on children's learning (Barr & Dreeben, 1983, in elementary schools and Rutter et al., 1979, in secondary schools) reported important effects for ability context motivated us to consider the ability context of the classroom. We used the aggregate of pretest scores on the PSI. On average, classroom means match children's means. Only a small proportion of children in classrooms were recent immigrants, although variability is high. Given the high proportion of children qualified for free lunch, the rates of poverty and public assistance in the school districts where the classrooms were located were surprisingly low. Classroom compositions of African American and Hispanic children reflect the distribution of racial/ethnic minority status among children, although variability in these compositional measures is considerable. On average, classes enrolled 19 children (although actual samples were smaller). More classrooms were sampled in Texas and Michigan than in New Jersey, California, or Florida.

Multivariate and Multilevel Analyses

Statistical properties of the outcome. We started our analyses by partitioning the variance in the Spring PSI into its between and within-classroom components. This is a one-way analysis of variance (ANOVA) with random effects, accomplished by running a fully unconditional HLM. Results, in Table 2, show that the PSI score has modest (but acceptable) lambda reliability (.662). The variance between classrooms (tau) is .123; variance between children within classrooms (sigma-squared) is .882. The intraclass correlation is .183. Thus, slightly less than a fifth of the vari-

Table 2 Partitioning Variance in Spring PSI Score: Fully Unconditional HLM Model for Preschool Context Study (*N* = 677 Children in 55 Classes)

Variance between classrooms (tau)	.123
Between-classroom standard deviation	.350
Variance within classrooms, pooled across classrooms (sigma-squared)	.882
Estimated HLM reliability (lambda)	.662
Proportion of variability between classrooms (intraclass correlation)	.122[a]
Proportion of between-classroom variability, adjusted for reliability	.183[b]

[a] The intraclass correlation is computed as follows: tau/(tau + sigma-squared).
[b] The adjusted intraclass correlation is computed by the same formula, except that sigma-squared is multiplied by lambda.

ability in the outcome lies systematically between classrooms. Ignoring (for the moment) a differential effect of context measures on different students within the same classroom, only between-classroom variance can be influenced by classroom context. Although not a large proportion, this level of between-classroom variability is consistent with other school-effects studies with achievement outcomes and sufficient to proceed with our analysis.[7] The intraclass correlation provides preliminary evidence to support the hypotheses posed for Research Question 1, that learning varies between classrooms.

Within-classroom HLM model. The within-classroom HLM model, which allows us to more fully answer Research Question 1, estimates each child's score on the outcome as a function of his or her own cognitive, behavioral, and social background. In this model, an HLM random coefficient regression model, the intercept or classroom average of the spring PSI adjusted for the characteristics of individual children in each class (β_0), is modeled as a random parameter ("free" in HLM). Other variables (with coefficients β_1 through β_{10}) are included as controls, so we constrained their between-classroom variances to zero and centered them around their grand means (i.e., the means for all children in the study). This allows us to include many more controls than would be appropriate in a study in which within-classroom samples are small (12 children). With this centering decision, the intercept may be interpreted as the classroom mean score on the spring PSI, adjusted for individual-level characteristics of children.

Information about this model is displayed in Table 3. Results are presented in two forms: (1) as HLM beta coefficients (col. 1) and (2) as within-classroom effect sizes (ES) in *SD* units (col. 4). Although nominal statistical testing is the most common method for determining the "significance" of effects, consideration of "substantive significance" has been suggested by some specialists (Cohen & Cohen, 1983; Rosenthal & Rosnow, 1984). This convention was also followed in the Head Start Synthesis Project (McKey et al., 1985). By these standards, effects less than .2 *SD* are small, those between .2 and .5 *SD* moderate, and those over .5 *SD* large. Viewed through this lens, the substantive educational significance of effects is determined by their magnitude and direction (i.e., at least .2 *SD*).

Only two characteristics of children are statistically related to the spring PSI: the fall PSI score (ES = .71, $p \leq .001$) and the CBRS (ES = .11, $p \leq .01$). Initially more able children and those with more prosocial behaviors demonstrated significantly higher scores on the PSI at the end of the preschool year. Other child-level controls have point estimates or effect sizes that are small and not statistically related to the outcome. Once cognitive and behavioral status are controlled, Hispanic children score .10 *SD* below (and African American children .03 *SD* below) European American children in Chapter 1 preschools. Children whose families have special needs score .12 *SD* below other children. Results show virtually no effect for mother's education, age, or gender. It is important to remember that each of these effects is computed *net* of others in the model (especially cognitive status); thus, we consider this a "net learning model."

The characteristics of children included in this model have explained 64% of the between-class variance in the outcome, using information drawn from the bottom panel of Table 3.[8] The significant chi-square statistic ($p \leq .001$) suggests that even with the full set of controls for children's cognitive and social characteristics, there is still significant residual variability between classrooms on the outcome. This provides a more definitive confirmation of our hypothesis for Research Question 1: Learning does vary significantly between classrooms even when children's demographic and cognitive status is accounted for.

How do children begin preschool? Because of the large impact of the fall PSI on the spring score, we also explored a level-1 (within-class) demographic HLM model with the fall PSI as the outcome. African American and Hispanic children scored well below European American children, and those with less educated mothers scored lower than those whose mothers had more formal education (all effects $p \leq .001$). Girls scored higher on the pretest than boys, and children needing extra help lower than those not needing help (both ps $\leq .05$). There was no effect for families needing extra help. This model explained 25% of the between-class variance in the fall PSI.[9]

Between-classroom HLM model on learning. The results displayed in Table 4 represent our final between-classroom (level-2) context HLM, which investigates differences between classrooms. This model

7. We also investigated the psychometric properties of the fall PSI, the pretest, with HLM. They were almost identical: The intraclass correlation in the pretest was .183, lambda reliability was .622.

8. Although R^2 figures are not provided by HLM, we may compute them by comparing the variance estimate provided at the bottom of Table 3 with the full between-class variance in the outcome (tau) shown in Table 2.

9. Although we have not included these results in this article, the results of these analyses are available from the authors upon request.

Table 3 Within-Classroom Hierarchical Model for Preschool Context Study: Effects on Spring PSI (N = 677 Children in 55 Classes)

Independent Variables	Estimated Effects: Beta Coefficients	Standard Error	p Value	Effect Size[a]
Classroom average spring PSI (posttest),[b] Mean (β_0)	−.001	.038	.974	. . .
Fall PSI (pretest) slope, Mean (β_1)	.667	.031	.000	.71***
Fall CBRS slope, Mean (β_2)	.100	.032	.003	.11**
Race gap (African American), Mean (β_3)	−.027	.077	.725	−.03
Race gap (Hispanic), Mean (β_4)	−.097	.087	.272	−.10
Age slope, Mean (β_5)	.019	.027	.487	.02
Gender gap (female), Mean (β_6)	.004	.052	.945	.01
Mother's education slope (below high school), Mean (β_7)	−.036	.073	.623	−.04
Mother's education slope (high school degree), Mean (β_8)	−.061	.061	.323	−.06
Child needed help gap, Mean (β_9)	−.041	.098	.681	−.04
Family needed help gap, Mean (β_{10})	−.110	.091	.234	−.12

Random Parameter	Estimated Parameter Variance	Degrees of Freedom	Chi-Square Statistic
Mean posttest	.04450	54	120.50***

[a] Effect sizes computed by dividing each beta coefficient by .939149, which is the within-classroom standard deviation. This is computed by taking the square root of sigma-squared, from Table 2.
[b] In this HLM model, all within-classroom independent variables are "fixed." Only the adjusted intercept term is allowed to vary randomly across classrooms.
** $p \leq .01$; *** $p \leq .001$.

Table 4 Between-Classroom Hierarchical Model for Preschool Context Study: Effects on Cognitive Gains (N = 677 Children in 55 Classes)

Independent Variables	Estimated Effects[a]: Gamma Coefficients	Standard Error	p Value	Effect Size[b]
Classroom average posttest (β_0), Mean	.022	.087	.800	. . .
Average pretest	−.004	.040	.913	−.02
Percent low mothers' education	−.069	.055	.220	−.33
Percent African American children	−.162	.065	.016	−.77*
Percent Hispanic children	−.006	.070	.936	−.03
Percent children needing help	−.069	.038	.076	−.33+
Percent recent immigrants	−.082	.047	.088	−.39+

Random Parameter	Estimated Parameter Variance	Degrees of Freedom	Chi-Square Statistic
Mean posttest	.03114	48	90.71***

[a] This HLM analysis also includes all within-classroom variables shown in Table 3. As the effects of these variables were almost identical to those shown in Table 3, we omitted them here. It is important to realize that the analysis contains a full set of controls at both the within- and between-classroom levels.
[b] Effect sizes computed by dividing each gamma coefficient by .21095, the between-classroom standard deviation of the outcome in the level-1 HLM model described in Table 3.
+ $p \leq .10$; * $p \leq .05$; ** $p \leq .01$; *** $p \leq .001$.

is designed to explore Research Question 2: Do contextual elements of the preschool classroom affect student learning? As the full set of within-class controls is also included in this two-level HLM model (including the pretest score), the outcome may be seen as cognitive *learning* rather than *status*. In this more complete model, the within-classroom effects changed only marginally from those shown in Table 3. Thus, we omitted them here. The number of independent variables we could include in a single between-classroom level-2 model is constrained by the modest size of the classroom sample ($N = 55$); thus, our model building strategy was hierarchical and iterative.[10]

Besides reporting the results in Table 4 as HLM gamma coefficients, we also converted these coefficients into between-school effect sizes (ES), presented in *SD* units (col. 4). Among the several context measures we considered, noteworthy is one nonfinding. The *ability context* (as distinguished between an individual's ability), demonstrated here by the average pretest score, has no effect on classroom differences in learning on the PSI. That is, once children's individual ability is taken into account, the average level of children's ability within each classroom has no additional impact on learning.

However, this is not the case for the *social context* of these preschool classrooms. Among the aggregate measures of social context considered, racial composition demonstrates the strongest effect. As the proportion of African American children in a class goes up, average learning declines (ES = $-.77$ *SD*, $p \leq .05$). The proportion of Hispanic children has no effect. The magnitude of other effects, although nonsignificant by conventional standards, suggests substantive importance. Classrooms with higher proportions of children with special needs have lower learning levels (ES = $-.33$ *SD*, $p \leq .10$), as do classrooms with higher proportions of children whose families have immigrated recently to the United States (ES = $-.39$ *SD*, $p \leq .10$). Similarly, classrooms with many children whose mothers have limited formal education show lower average learning (ES = $-.33$ *SD*).

Our results suggest that classroom counts. However, it is social context of the classroom (i.e., racial composition, average SES, high proportions of special needs and immigrant children) that affects learning in these Chapter 1 preschools; the cognitive context (i.e., average ability) is unimportant. Using the residual variance shown in the lower panel of Table 4, our class-level HLM model explains 30% of the remaining between-class variance in the outcome.

Differential Contextual Effects within Classrooms

Based on the importance of classroom racial composition on learning from the results reported in Table 4, we investigated these effects more fully in another HLM contextual model. Specifically, we explored whether racial composition had different effects on children, depending on their own racial status. Referring to Table 3, we modeled both β_0 (the intercept) and β_3 (the learning gap between European American and African American children) as outcomes in a between-classroom "slopes as outcomes" HLM model.[11] The results of this analysis are shown in Table 5.

Although the proportion of African American children in classrooms has a negative effect on average PSI scores (also shown in Table 4), results in Table 5 suggest that the effects of classroom racial composition are *different* for African American and European American children. Specifically, the difference in learning between African American and European American children is *less* in classrooms with high concentrations of African American children. Our finding of a statistically significant effect (ES = .49 *SD*, $p \leq .05$) for classroom racial composition on the race gap in learning in Chapter 1 pre-K classrooms provides support for our hypothesis of differential contextual effects in the same classroom posed in Research Question 3.

It is also possible that these effects are nonlinear (i.e., that there is a particular cutpoint when racial concentration becomes more detrimental for children from different racial backgrounds). To explore this question we reran the analyses using dummy variables for various racial composition levels; we did not find systematic nonlinear effects. According to our results, racial composition effects are approximately equivalent for African American and European American children when classroom racial compositions are 20%–30% African American (were these re-

10. Considering the full list of class-level variables shown in Table 1 and described in the Appendix, we entered sets of contextual variables in groups (demographic composition, special needs, racial composition, structural characteristics) and retained only those with effects whose probability levels were .15 or below. Exceptions were our decision to retain the ability composition and the proportion of mothers with little education (our only SES measure) in the model, given the substantive importance of these factors in sociological research on schools.

11. Exploratory work showed that β_3 did not vary between classrooms. Thus, our level-2 HLM model investigated the intercept as a random variable and the African American "race gap" as fixed.

Table 5 Between-Classroom Hierarchical Model for Preschool Context Study: Effects on Posttest and African American/Non–African American Race Gap (N = 667 Children in 55 Classes)

Independent Variables	Estimated Effects: Gamma Coefficients	Standard Error	p Value	Effect Size[a]
Intercept model:				
Classroom average posttest (β_0), Mean	−.097	.097	.327	. . .
Average pretest	−.020	.041	.637	.09
Percent low mother's education	−.064	.056	.254	−.30
Percent African American children	−.380	.104	.001	−1.80***
Percent Hispanic children	−.051	.073	.485	.24
Percent children needing help	−.086	.039	.032	−.41*
Percent recent immigrants	−.095	.048	.051	−.47+
Race gap (African American) model (β_3):[b]				
Mean	.131	.090	.148	. . .
Percent African American children	.278	.104	.011	.49*

Random Parameter	Estimated Parameter Variance	Degrees of Freedom	Chi-Square Statistic
Mean PSI Gain	.03257	48	92.45***

[a] Effect sizes computed by dividing each gamma coefficient in the intercept model by .210952, the between-classroom standard deviation of the outcome in the level-1 HLM model described in Table 3.
[b] Race gap is fixed, as it does not vary systematically between schools. Effect size computed by dividing the gamma coefficient in the race-gap model by the standard deviation of the slope, .57104, calculated by multiplying the standard error for that coefficient shown in the within-school HLM model displayed in Table 3 by the square root of the sample size.
+ $p \leq .10$; * $p \leq .05$; *** $p \leq .001$.

sults graphed, the lines would cross at that point). The average proportions of African American children in Chapter 1 pre-K classrooms is, however, considerably higher (43%, see Table 1). There is also much variability in classroom racial composition in these preschools, as evidenced by the high *SD*s of these variables.

DISCUSSION

Expanding the Boundaries of Preschool Research

Advantages of the data. We suggest that this study expands the scope of research on young children's cognitive development in several ways. The strong study design, with data available at both the beginning and end of the school year for a large number of children and Chapter 1 pre-K classrooms allowed us to explore the effects of social context on cognitive development of a large and geographically diverse population of disadvantaged children outside of Head Start. This had not been done before. The data structure allowed us to extend a conceptual and analytic approach now common in sociological research on secondary schools to settings that have heretofore not been examined in this way. Our aims were (1) to demonstrate the importance of the social context of preschool classrooms for disadvantaged children and (2) to identify important context elements that influence learning. We were more successful with the first objective than with the second.

Data limitations. Despite the advantages of the data we used, some limitations also constrained our ability to explore the social context of preschool classrooms. First, outcome data were limited to a single measure of children's cognitive development, whereas it would be useful to examine contextual effects on children's socioemotional development. The absence of information about and from these children's families (data about children and families were provided by teachers) was another disadvantage, particularly the lack of data on family income. Although we considered many variables about classrooms drawn from teachers' reports collected on surveys, none showed significant effects. Lack of effects here could be real or could result from either measurement error or from problems in conceptualization.[12] Thus, we were restricted to focusing on a

12. According to Seppanen et al. (1993), the Chapter 1 Preschool Study also included data drawn from observations of

rather narrow range of context measures: primarily composition of the classroom, defined in terms of aggregate information from individuals.

The Context of Learning

An interest in context. Ideally, a study of how social context influences young children's learning would explore family processes around cognitive development, including the family conditions that influence context. Increasingly, researchers are focusing on contextual factors in understanding real-world development, especially for young children growing up in impoverished settings. Typically, the contexts to be considered are historical, cultural (including family), social, and physical (Fisher & Tryon, 1990). In this article, we centered our exploration of context on the classroom and school. The research on social context effects in early education is sparse. Developmental psychologists have begun to recognize that the effects of social interventions may vary as a function of children's personal characteristics. For this reason, "applied developmental research and intervention strives to be multilevel" (Fisher & Tryon, 1990, p. 5). Although such contentions are reassuring to those of us who advocate a multilevel approach, we are not so sanguine about the existing studies of young children's development before they enter formal school settings.

We were struck by the lack of an effect for the ability context. However, our ability to uncover other contextual effects undercuts alternative explanations that relate to data constraints and low power. The major contextual effects we identified concern the demographic composition of preschool classrooms targeted at disadvantaged children. The fact that the ability context of preschool contexts *does not* appear to affect learning is an important substantive finding.

Is local always best? Our findings raise questions about the wisdom of a policy that maintains the local nature of preschool programs for poor children. The concentration in preschool classrooms of children by race, recent immigrant status, and special needs (often associated with SES) appears to impede children's learning. Although the United States made substantial strides several decades ago in desegregating our public schools, the increasing residential segregation of the nation's large cities (Farley, 1991; Farley & Frey, 1994) has led to educational resegregation. Conclusions drawn from our evidence would argue that, from a policy perspective, such social and educational concentration (and surely economic concentration, if we had the data to measure it) is detrimental to children's learning—even (or maybe particularly) the learning of very young children. In our opinion, segregation of educational settings by race and poverty *is* an appropriate issue for social policy and not just a situation to be accepted as inevitable.

Our findings have relevance for policies about educational settings where children with special needs and those who have immigrated recently are concentrated. Clearly, immigrant families most often choose (or are forced) to settle near other recent immigrants. Virtually every teacher would report problems in dealing with large numbers of immigrant children (as well as children and families with special needs) in a single classroom. Our empirical evidence provides some suggestion of these difficulties in terms of children's cognitive learning. These findings suggest a policy effort toward integrating such children into classrooms with others *unlike* themselves.

Cross-level interactions. Our analyses that investigated the possibility of differential effects of classroom context on children within each class focused on race. Because we found strong effects of racial concentration on learning between classrooms, it made sense to focus on cross-level interactions by race. In preschool classrooms with racially mixed enrollments, higher proportions of African American children had differential effects on children of different races. Both racial groups learned less on average in these classrooms than in classrooms with higher proportions of Hispanic or European American children, although race differences in learning were less in classrooms with higher proportions of African American children.

How should we interpret this finding? Two different (and contradictory) interpretations seem plausible. Above we argued that educational policies that allow disadvantaged children to be concentrated in particular neighborhoods and schools should be questioned; however, these results suggest some countervailing advantages (for African American children). One interpretation is cultural. Recent experiments in implementing cultural awareness and pride through Afrocentric curricula in a few inner-city elementary schools would favor an explanation that there are some advantages of isolating African American children in school settings. Another interpretation suggests that the finding of more racial equity in classrooms with high concentrations of African American students is illusory. European American children do worse in these settings. Our major findings indicate that in classrooms where mi-

classrooms. However, there appeared to be barriers to confidentiality, so that linking the classroom observation data to the public-use data we received from RMC was impossible.

nority children are concentrated, everyone does worse. We suggest that social equity in learning is meaningless unless it is accompanied by generally high levels of overall learning. The doubly multivariate nature of such analytic models means that researchers need to be careful not to draw meaning from particular coefficients in isolation from others.

A final comment. This study indicates that disadvantaged children are learning in Chapter 1 classrooms. In addition, we found that performance on the PSI is influenced by both the cognitive and demographic status of children (an expected finding) and also by classroom context. The demographic makeup of the classroom makes a difference in how much children learn, above and beyond their personal status.

That classroom context *matters* certainly extends beyond preschool. But will the disadvantaged children fortunate enough to experience publicly financed pre-K in Chapter 1 classrooms continue their positive educational experiences thereafter? Probably not. The boost children receive by their participation in Chapter 1 might not be sustained. In discussing ethical issues involved in applied developmental research, Fisher and Tryon (1990) alert us to the "temporality of individual change." The fact that any intervention at some point in a child's developmental trajectory may be effective does not rule out important environmental factors that can interfere thereafter. Other research has shown that disadvantaged children who experienced Head Start were actually in lower-quality middle schools than their disadvantaged peers who had no preschool experience, and particularly compared to children who attended non–Head Start preschools (Lee & Loeb, 1995). Lewis (1990) suggests that expectations of permanent change from educational (or other social) interventions place an unfair burden on these programs. This is particularly true when cost effectiveness and universality are added to the list of criteria for the "success" of publicly supported programs (which they almost always are).

ACKNOWLEDGMENTS

We are grateful to Christine Dwyer of the RMC Research Corporation for supplying us with the data and codebooks for this study. We also appreciate the advice of Patricia S. Seppanen, who directed the original study, "Chapter 1–Funded Early Childhood Programs" (funded by the U.S. Department of Education). Any errors of analysis or interpretation, however, are entirely ours. An early version of this study was presented at the annual meeting of the American Sociological Association, Washington, DC, August 1995.

ADDRESSES AND AFFILIATIONS

Corresponding author: Valerie E. Lee, School of Education, University of Michigan, Ann Arbor, MI 48109; e-mail: velee@umich.edu. Susanna Loeb and Sally Lubeck are also at the University of Michigan.

APPENDIX

DESCRIPTION OF VARIABLES CONSIDERED IN PRESCHOOL CONTEXT STUDY: DEPENDENT VARIABLE

PSI Posttest

The Preschool Inventory (PSI) is a 32-item assessment designed to measure "achievement in areas regarded as necessary for success in school" (Caldwell, 1970, p. 4). The test developers intended that the PSI be used with disadvantaged children, and it was specifically designed for evaluating Head Start. Rather than predicting future scholastic aptitude, as an IQ test might, the intention was that the PSI would evaluate whether children had the requisite skills to allow them to fulfill their intellectual potentials in the context of the classroom. Compared to disadvantaged children, "it was expected that middle-class children would arrive at their first formal school experience with skills measured on the PSI firmly in place" (B. Caldwell, personal communication, May 29, 1987, cited in Lee et al., 1988, p. 213).

The PSI was administered by trained testers in either English or Spanish. The posttest PSI score was measured in spring 1991, the end of the preschool year. Our sample included only children with both pretest and posttest PSI scores (86% of the total of 783 sampled children) converted to a z score ($M = 0$, $SD = 1$) for multivariate analyses.

Independent Variables Measured on Children

PSI pretest. The PSI (described above) was individually measured in fall 1990, early in the preschool year. It was converted to a z score ($M = 0$, $SD = 1$) for multivariate analyses. The correlation with spring PSI = .73.

Child Behavior Rating Scale (CBRS). The CBRS is a 32-item behavior rating scale, designed to evaluate a child's social behavior (with peers and adults) and task behavior. The scale, adapted from the Bronson Executive Skill Profile (Bronson, 1985) by the original author, has been used in an earlier evaluation of Head Start by RMC. The score is the mean of the 32 behaviors, each measured on a 5 point scale (Sepannen et al., 1993). Ratings were completed by the child's teacher early in the preschool year. The correlation with fall PSI = .38, with spring PSI = .34.

Age. Children's age at the beginning of the preschool

year, measured in months. Variable was converted to a z score ($M = 0$, $SD = 1$) when used in multilevel analyses.

Gender. Dummy-coded variable, with girls coded 1, boys coded 0.

Race. The three-level race variable was converted to two dummy-coded variables. One variable, *Afro-American,* coded African American children 1, others 0. A second dummy variable, *Hispanic,* coded Hispanic children 1, others 0. This coding results in the comparison group for both dummy variables composed of European American children (i.e., European Americans are coded 0 throughout).

Mother's education. The variable measuring mother's education in years was a decidedly nonnormal distribution. Therefore, we created two dummy-coded variables: *MEDLTHS* (children whose mothers had less than high school coded 1, others coded 0), and *MEDHS* (children whose mothers had a high school degree but no more coded 1, others coded 0). This resulted in the comparison group for both dummy variables composed of children whose mothers had some education beyond high school (i.e., these children coded 0 throughout).

Free or reduced lunch. This variable indicates whether children are eligible for free or reduced lunch, which is a function of family income. We coded it 1 for yes, 0 for no. A substantial proportion of the sample (20%) was missing data on this variable. We eliminated it from our multivariate analyses due to nonsignificance.

Child needed help. This dummy-coded variable was drawn from teachers' response to the query: "During the current school year, have you felt, or has anyone suggested, that this child needed help for any behavioral, emotional, or mental problem?" Variable coded 1 for yes, 0 for no.

Family needed help. This dummy-coded variable was drawn from teachers' response to the query: "During the current school year, have you felt, or has anyone suggested, that this child's family needed help due to stress related to substance abuse, economic problems or behavioral, emotional, or mental problems?" Variable coded 1 for yes, 0 for no.

Independent Variables Measured on Classrooms

Average pretest. Classroom average of children's scores of the fall PSI. In multilevel analyses, the variable was used as a z score ($M = 0$, $SD = 1$) on the sample of classrooms.

Percent low mothers' education. Classroom aggregate of dummy variable indicating the educational level of each child's mother. The aggregate thus indicates the percent of children in the classroom whose mothers have less than a high school education. In multilevel analyses, the variable was used as a z score ($M = 0$, $SD = 1$) on the sample of classrooms.

Proportion recent immigrants. The teacher indicated whether the child was a recent immigrant (coded 1 on student data file) or not (coded 0). The aggregate thus indicates the percent of children in each classroom who were recent immigrants. In multilevel analyses, the variable was used as a z score ($M = 0$, $SD = 1$) on the sample of classrooms.

Percent poverty. Using data from the 1990 U.S. Census, this variable indicates the proportion of residents in the school district in which each school was located with income below the poverty line. In multilevel analyses, the variable was used as a z score ($M = 0$, $SD = 1$) on the sample of classrooms. It was subsequently deleted from analysis due to nonsignificance.

Percent welfare. Using data from the 1990 U.S. Census, this variable indicates the proportion of residents in the school district in which each school was located who receive public assistance. In multilevel analyses, the variable was used as a z score ($M = 0$, $SD = 1$) on the sample of classrooms. It was subsequently deleted from the analysis due to nonsignificance.

Proportion African American. A classroom aggregate of the dummy variable measuring the child's race (i.e., African American versus non–African American). The aggregate indicates the percent of children in each classroom who were African American. In multilevel analyses, the variable was used as a z score ($M = 0$, $SD = 1$) on the sample of classrooms.

Proportion Hispanic. A classroom aggregate of the dummy variable measuring the child's race (i.e., Hispanic or not Hispanic). The aggregate indicates the percent of children in each classroom who were Hispanic. In multilevel analyses, the variable was used as a z score ($M = 0$, $SD = 1$) on the sample of classrooms. It was subsequently deleted from analysis due to nonsignificance.

Principal leadership. A factor score-weighted composite of eight variables drawn from teachers' reports about the classroom principal. Each item was scored from 1 (Not at all like this) to 5 (Very much like this). In multilevel analyses, the variable was used as a z score ($M = 0$, $SD = 1$) on the sample of classrooms. It was subsequently deleted due to nonsignificance. Included items, listed in order of factor weights within the composite, are as follows: (1) provides strong leadership, (2) helps staff work toward program goals with support and training, (3) solves administrative problems well, (4) is an expert resource in the area of child development, (5) shares ideas and suggestions with staff, (6) sets appropriate goals for the program, (7) delegates authority appropriately, (8) resolves conflict between staff members effectively.

Class size. Teacher report of the number of children in her class at the beginning of the preschool year. In multilevel analyses, the variable was used as a z score ($M = 0$, $SD = 1$) on the sample of classrooms. It was subsequently deleted from analysis due to nonsignificance.

State location. We created four dummy variables representing the five states from which the school district and classroom samples were drawn. The excluded category was *New Jersey* (4 districts, 10 schools, 10 classes). Dummy variables represented *California* (2 districts, 9 schools, 10 classes), *Florida* (2 districts, 9 schools, 9 classes), *Texas* (3 districts, 12 schools, 13 classes), and *Michigan* (5 districts, 9 schools, 13 classes). The variables were subsequently deleted from analysis due to nonsignificance.

REFERENCES

Barr, R., & Dreeben, R. (1983). *How schools work.* Chicago: University of Chicago Press.

Bereiter, C. (1963). Some persisting dilemmas in the measurement of change. In C. W. Harris (Ed.), *Problems in measuring change* (pp. 35–52). Madison: University of Wisconsin Press.

Bickel, D. D., Zigmond, N., & Strayhorn, J. (1991). Chronological age at entrance to first grade: Effects on elementary school success. *Early Childhood Research Quarterly, 6,* 105–117.

Bidwell, C. E. (1972). Schooling and socialization for moral commitment. *Interchange, 3,* 1–26.

Bidwell, C. E., & Kasarda, J. (1980). Conceptualizing and measuring the effects of schools and schooling. *American Journal of Education, 88,* 401–430.

Bronson, M. (1985). *Manual for the Bronson social and task skill profile.* Chestnut Hill, MA: Boston College.

Bronson, M., Tivnan, T., & Seppanen, P. (1995). Relations between teacher and classroom activity variables and the classroom behaviors of pre-kindergarten children in Chapter 1 funded programs. *Journal of Applied Developmental Psychology, 16,* 253–282.

Brown, B. W., & Saks, D. H. (1985). The microeconomics of schooling. *Review of Research in Education, 9,* 217–254.

Bryk, A. S., Lee, V. E., & Holland, P. B. (1993). *Catholic schools and the common good.* Cambridge, MA: Harvard University Press.

Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical linear models: Applications and date analysis methods.* Newbury Park, CA: Sage.

Caldwell, B. M. (1970). *Cooperative Preschool Inventory: Revised edition.* Menlo Park, CA: Addison-Wesley.

Caldwell, B. M., & Freund, J. H. (1980). *The Preschool Inventory.* Little Rock: University of Arkansas, Department of Rehabilitation and Special Education, Center for Child Development and Education.

Claus, R N., & Quimper, B. E. (1991). *State bilingual and ECIA Chapter 1 migrant product report, 1990–1991.* Saginaw, MI: Saginaw Public Schools, Department of Evaluation Services. (ERIC Document Reproduction Service No. ED 338 461).

Claus, R N., & Quimper, B. E. (1992). *State bilingual and ECIA Chapter 1 migrant product report, 1992–1993.* Saginaw, MI: Saginaw Public Schools, Department of Evaluation Services. (ERIC Document Reproduction Service No. ED 345 910).

Cohen, J., & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences* (2d ed.). Hillsdale, NJ: Erlbaum.

Coleman, J. S., Campbell, E., Hobson, C., McPartland, J., Mood, A., Weinfeld, F., & York, R. (1966). *Equality of educational opportunity.* Washington, DC: Government Printing Office.

Cusick, P. A. (1983). *The egalitarian ideal and the American high school.* New York: Longman.

Eyler, J., Cook, V. J., & Ward, L. E. (1983). Resegregation: Segregation within desegregated schools. In C. H. Rossell & W. D. Hawley (Eds.), *The consequences of school desegregation* (pp. 126–162). Philadelphia: Temple University Press.

Farley, R. (1991). Residential segregation of social and economic groups among Blacks. In C. Jencks & P. E. Peterson (Eds.), *The urban underclass* (pp. 274–298). Washington, DC: Brookings.

Farley, R., & Frey, W. H. (1994). Changes in the segregation of Blacks and Whites. *American Sociological Review, 59,* 23–45.

Fisher, C. B., & Tryon, W. W. (1990). Emerging ethical issues in an emerging field. In C. B. Fisher & W. W. Tryon (1976). *Preschool inventory—revised.* Princeton, NJ: Educational Testing Service.

Frede, E., & Barnett, W. S. (1992). Developmentally appropriate public school preschool: A study of implementation of the High/Scope curriculum and its effects on disadvantaged children's skills in the first grade. *Early Childhood Research Quarterly, 7,* 483–499.

Garet, M. S., & Delaney, B. (1988). Students' courses and stratification. *Sociology of Education, 61,* 61–77.

Grant, G. (1988). *The world we created at Hamilton High.* Cambridge, MA: Harvard University Press.

Hallinan, M. T., & Sorenson, A. B. (1983). The formation and stability of instructional groups. *American Sociological Review, 48,* 838–851.

Lee, V. E., Brooks-Gunn, J., & Schnur, E. (1988). Does Head Start work? A 1-year follow-up comparison. *Developmental Psychology, 24,* 210–222.

Lee, V. E., & Bryk, A. S. (1988). Curriculum tracking as mediating the social distribution of high school achievement. *Sociology of Education, 61,* 78–94.

Lee, V. E., & Bryk, A. S. (1989). A multilevel model of the social distribution of high school achievement. *Sociology of Education, 62,* 172–192.

Lee, V. E., Bryk, A. S., & Smith, J. B. (1993). The organization of effective secondary schools. In L. Darling-Hammond (Ed.), *Review of Research in Education, 19* (pp. 171–267). Washington, DC: American Educational Research Association.

Lee, V. E., & Loeb, S. (1995). Where do Head Start attendees end up? One reason why preschool effects fade out. *Educational Evaluation and Policy Analysis, 17,* 62–82.

LeTendre, M. J. (1991). The continuing evolution of a federal role in compensatory education. *Educational Evaluation and Policy Analysis, 13,* 328–334.

Lewis, M. (1990). Assessing social intervention: Scientific and social implications. In C. B. Fisher & W. W. Tryon (Eds.), Ethics in applied developmental psychology, *Annual advances in applied developmental psychology* (Vol. 4, pp. 81–91). Norwood, NJ: Ablex.

Linn, R. L., & Slinde, J. A. (1977). The determination of the significance of change between pre- and posttesting periods. *Review of Educational Research, 47,* 121–150.

Madaus, G. (1993, October). *Assessment issues around the reauthorization of Chapter 1.* Paper presented to the National Academy of Education, Ann Arbor, MI.

Mahard, R. E., & Crain, R. L. (1983). Research on minority achievement in desegregated schools. In C. H. Rossell & W. D. Hawley (Eds.), *The consequences of school desegregation* (pp. 103–125). Philadelphia: Temple University Press.

May, C., & Farha, J. (1989). *A longitudinal study of the Chapter 1 prekindergarten program in the Wichita Public Schools.* Pa-

per presented at the annual meeting of the American Educational Research Association, San Francisco.

McKey, R. H., Condelli, L., Granson, H., Barrett, B., McConkey, C., & Plantz, M. (1985, June). *The impact of Head Start on children, families, and communities.* Final Report of the Head Start Evaluation, Synthesis, and Utilization Project (Contract No. 105-81-C-026). Washington, DC: Administration for Children, Youth, and Families.

Meisels, S. J. (1987). Uses and abuses of developmental screening and school readiness testing. *Young Children, 42,* 4–6, 68–73.

New York City Board of Education. (1989). *Project Giant Step Chapter 1 program, 1987–88* (OREA Report). New York: New York City Board of Education. (ERIC Document Reproduction Service No. ED 317 653).

Reynolds, A. J. (1989). A structural model of first-grade outcomes for an urban, low socioeconomic status, minority population. *Journal of Educational Psychology, 81,* 594–603.

Reynolds, A. J. (1991). Early schooling of children at risk. *American Educational Research Journal, 28,* 392–422.

Rogosa, D. R., & Willett, J. B. (1985). Understanding correlates of change by modeling individual differences in growth. *Psychometrika, 50,* 203–228.

Rosenthal, R., & Rosnow, R. L. (1984). *Essentials of behavioral research: Methods and data analysis.* New York: McGraw-Hill.

Rutter, M., Maughan, B., Mortimore, P., Outson, J., & Smith, A. (1979). *Fifteen thousand hours: Secondary schools and their effects on children.* Cambridge, MA: Harvard University Press.

Schofield, J. W., & Sagar, H. A. (1983). Desegregation, school practices, and student race relations. In C. H. Rossell & W. D. Hawley (Eds.), *The consequences of school desegregation* (pp. 58–102). Philadelphia: Temple University Press.

Seawell, T., & Ross, G. (1992, April). *Relationships between a prekindergarten intervention program and school success.* Paper presented at the annual meeting of the American Educational Research Association. San Francisco.

Seppanen, P., Godin, K., Metzger, J., Bronson, M., & Cichon, D. (1993). *Observational study of early childhood programs: Chapter 1-funded early childhood programs* (Final report, Vol. 2). Washington, DC: U.S. Department of Education.

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

Shiminski, J. (1992). *"Unusually successful": Pittsfield Chapter 1.* Pittsfield MA: Pittsfield Public Schools. (ERIC Reproduction Service No. ED 360 079)

Sinclair, B., & Gutmann, B. (1990, August). *A summary of state Chapter 1 participation and achievement information for 1987–88* (Contract No. LC89015001). Washington, DC: U.S. Department of Education.

Sorensen, A. B. (1987). The organizational differentiation of students in schools as an opportunity structure. In M. T. Hallinan (Ed.), *The social organization of schools* (pp. 103–130). New York: Plenum.

Stringfield, S. (1991). Introduction to the special issue on Chapter 1 policy and evaluation. *Educational Evaluation and Policy Analysis, 13,* 325–327.

Trangmoe, J. (1988). *Glendive migrant program dedicated to meeting the needs of migrant children and their families.* East Lansing, MI: National Center for Research on Teacher Learning. (ERIC Reproduction Service No. ED 317 333)

Willett, J. B. (1988). Questions and answers in the measurement of change. In E. Z. Rothkopf (Ed.), *Review of research in education* (Vol. 15). Washington, DC: American Educational Research Association.

Willett, J. B. (1994). Measuring change more effectively by modeling individual growth over time. In T. Husen & N. Postelthwaite (Eds.), *The international encyclopedia of education* (2d ed.). Oxford: Pergamon.