Journal of Educational Psychology
1994, Vol. 86, No. 4, 542-555

0022-0663/94/$3.00

Outcomes of an Emergent Literacy Intervention in Head Start

G. J. Whitehurst, J. N. Epstein, A. L. Angell, A. C. Payne, D. A. Crone, and J. E. Fischel

Classrooms of 4-year-olds attending Head Start were randomly assigned to an intervention condition, involving an add-on emergent literacy curriculum, or a control condition, involving the regular Head Start curriculum. Children in the intervention condition experienced interactive book reading at home and in the classroom as well as a classroom-based sound and letter awareness program. Children were pretested and posttested on standardized tests of language, writing, linguistic awareness, and print concepts. Effects of the intervention were significant across all children in the domains of writing and print concepts. Effects on language were large but only for those children whose primary caregivers had been actively involved in the at-home component of the program. One linguistic awareness subtest, involving the ability to identify the first letter and first sound of words, showed significant effects.

> The single most important activity for building the knowledge required for eventual success in reading is reading aloud to children (National Academy of Education, 1985, p. 23).

> Preschooler reading . . . may make a weaker or more indirect contribution to literacy acquisition than is generally believed. . . . Educators may have overemphasized shared reading (Scarborough & Dobrich, 1994, p. 295).

Are shared reading and other emergent literacy experiences in the preschool period important for later literacy? This question may seem akin to challenging the value of motherhood or apple pie, but the research base for causal conclusions about the effects of shared reading on later development is weaker than most individuals suppose. Scarborough and Dobrich (1994) have suggested that correlations between frequency of shared reading and emergent literacy or reading achievement can largely be accounted for either by other variables that covary with social class and frequency of shared reading or by the effects on parents of individual differences in children's inherent interest in books. Scarborough and Dobrich estimate that only about 8% of the variance in emergent literacy or literacy outcomes in children can be attributed to preschool children's exposure to books. However, given the many statistical and methodological problems in the literature reviewed by Scarborough and Dobrich, such firm and pessimistic conclusions may be unwarranted (Lonigan, 1994; Payne, Whitehurst, & Angell, in press).

One significant impediment to interpreting the current research literature is that early literacy-related experiences and later literacy outcomes have been treated holistically. However, there are many forms of preschool emergent literacy activity, several domains of emergent literacy ability, and at least two critical dimensions of early reading ability (Gough, 1991). Multivariate problems often include complex relationships that do not resolve into single questions or answers.

Figure 1 is our attempt to capture some of this complexity while still allowing a simple conceptual model. Here preschool literacy activities are divided into those that focus on shared book reading and those that focus more directly on the sounds and letters that form the basis of the reading code. Examples of the former are prototypical instances of shared reading of picture books by parents and children. Examples of the latter include preschoolers' learning of the letters of the alphabet through a puzzle board, or children engaging in rhyming games that teach sound structure. In Figure 1, individual differences in the frequency or form of preschool literacy activities are assumed to influence individual differences in preschoolers' emergent literacy skills, which consist of four components: language (e.g., vocabulary size), writing (e.g., the ability to print one's first name), linguistic awareness (e.g., awareness of the phonemic segments of speech), and print concepts (e.g., naming letters). Emergent literacy abilities, in turn, are assumed to influence individual differences in the components of early reading (i.e., decoding and word recognition or comprehension).

We believe that the model represented by Figure 1 is consistent with most of the theorizing about the role of early

G. J. Whitehurst, J. N. Epstein, A. L. Angell, A. C. Payne, and D. A. Crone, Department of Psychology, State University of New York at Stony Brook; J. E. Fischel, Department of Pediatrics, State University of New York at Stony Brook.

This research was supported by Grant 91-01249-000 to G. J. Whitehurst from the Pew Charitable Trusts. It was also supported by Grant 90CD095701 to G. J. Whitehurst and J. E. Fischel and by Grant 90CD096201 to G. J. Whitehurst, both of which were from the U.S. Administration for Children and Families.

Views expressed herein are those of the authors, and these views have not been cleared by the grantors.

Portions of the data reported herein formed the basis of a doctoral dissertation submitted to the State University of New York at Stony Brook by J. N. Epstein.

We are deeply appreciative of the support of the administration, staff, and families of both Long Island Head Start and Children's Community Head Start, without whom this work could not have been accomplished.

Correspondence concerning this article should be addressed to G. J. Whitehurst, Department of Psychology, State University of New York, Stony Brook, New York 11794-2500. Electronic mail may be sent via Internet to russ@psych1.psy.sunysb.edu.

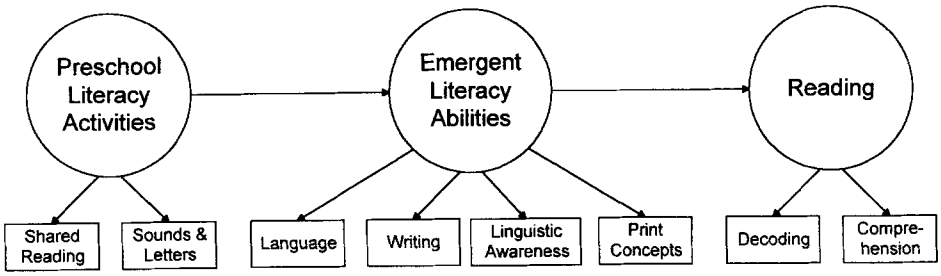

Figure 1. Prevalent model of the relations between emergent literacy experience and the development of literacy skills.

experience in the development of literacy (e.g., Adams, 1990; Mason, 1993). It is also broadly consistent with studies in which batteries of phonological, linguistic, and cognitive tests have been given to young children and subjected to factor analysis (Olson, Wise, Conners, & Rack, 1990; Stanovich, Cunningham, & Cramer, 1984; Wagner, Torgesen, Laughon, Simmons, & Rashotte, 1993).

Research on the linkages suggested by Figure 1 has focused for the most part on individual bivariate relationships that are embedded within the model rather than on the causal implications of the model as a whole. For example, research has shown that at least some forms of shared book reading increase vocabulary size in preschoolers (Valdez-Menchaca & Whitehurst, 1992; Whitehurst et al., 1988, 1994), but researchers have not examined whether this in turn has any effect on reading decoding or comprehension. Evaluation of the model in Figure 1 is further complicated by the fact that most research has been based on correlational approaches with small sample sizes and methods that invite competing causal explanations (Lonigan, 1994). For example, although the positive association between reading aloud to children in the preschool period and later reading ability has been repeatedly demonstrated (e.g., Raz & Bryant, 1990; J. Stevenson & Fredman, 1990; Wells 1985), it is quite possible that reading aloud in the home is simply a marker variable for other characteristics of homes in which educational achievement is valued. It could be these other characteristics rather than the emergent literacy experience per se that are causal. Alternatively, it may be, as argued by Scarborough and Dobrich (1994), that frequency of book reading primarily reflects the child's inherent interest in print and learning, rather than the parents' interest in reading to the child. We believe that theoretical progress in understanding the relations between preschool literacy experience and later reading is likely to be furthered by research approaches that are multivariate and that advance causal models through either experimental or appropriate statistical methodologies.

The model that is sketched in Figure 1 also has important implications for social and educational policy. Substantial numbers of children enter kindergarten with low levels of the skills that are prerequisites to school success. These children are drawn disproportionately from families with low incomes (Alexander & Entwisle, 1988; Carnegie Foundation for the Advancement of Teaching, 1991; Walker, Greenwood, Hart, & Carta, 1994). Furthermore, individual differences in emergent numeracy and literacy at entry into kindergarten tend to be maintained or magnified over the school years (Morrison, McMahon, & Williamson, 1993; H. W. Stevenson & Newman, 1986). For instance, H. W. Stevenson and Newman found a correlation of .52 between the ability to name the letters of the alphabet as a child entered kindergarten and performance on the Gates–MacGinitie Reading Test (Gates & MacGinitie, 1965) in Grade 10.

One antidote to infrequent experience with books in the home might be exposure to shared book reading and other language and literacy enhancing experiences in preschool settings. However, language and emergent literacy interactions may be among the weakest areas of current preschool practice. For example, in a recent study of a randomly selected sample of 50 Chapter 1 public preschools in North Carolina Bryant, Peisner-Feinberg, and Clifford (1993) found, on the basis of a standardized observational measure of classroom quality, that the mean score on the language and reasoning subscale was "the lowest of all subscales and considerably below the target score" (p. 5).

Head Start is the U.S. government's preschool program for children from families who are living at the poverty level. Head Start has served about 12 million families and children since its inception in 1965 and has academic readiness as a principal aim: "The first [of Head Start's major goals] is to provide disadvantaged children with the skills necessary for success in school" (U.S. Department of Health and Human Services, 1992, p. 1). At least some Head Start programs appear to have weaknesses in curriculum practices that mirror those reported for public preschools. A survey of 32 Head Start classrooms in the Southeastern United States led Bryant, Lau, Burchinal, and Sparling (in press) to conclude that it was "particularly distressing that the ratings of language use in the classroom were so low.... This seems to be an important area to target for improvement by these Head Start programs" (ms. p. 16). Edgar, Jenkins, Hayes, Sampson, and Vadasy (1992) examined pretest to posttest changes on various preacademic and language measures over 1 year of Head Start for children from 12 program sites distributed over three states in the Pacific Northwestern United States. Edgar et al. reported that "gain scores were of small magnitude" (p. 17). Bryant, et al. (in press) reported that children exiting Head Start in

their sample from the Southeastern United States were about one standard deviation below national averages on a standardized test of cognitive skills. In our own research, we have found no appreciable difference between the language and emergent literacy developmental quotients of a cohort of children graduating from Head Start and a cohort entering Head Start the following year at the same centers (Whitehurst, et al., 1993). Furthermore, scores at graduation from Head Start at these sites were low in absolute terms; they ranged from a mean at the 20th percentile for prereading concepts such as phonemic awareness to a mean at the 10th percentile for vocabulary. If the model depicted in Figure 1 is valid, then children with such low levels of preacademic skills can be expected to have difficulties succeeding in school.

The data reported above may not apply to Head Start in general: The researchers conducting the studies reported have not sampled Head Start programs systematically. We expect that there is considerable variability across programs in the degree that the curriculum supports the enhancement of preacademic skills. The data are also not necessarily an indictment of the value of Head Start programs that do not emphasize preacademic skills: Head Start provides valuable services in areas such as health, nutrition, and the development of children's social competence that justify public expenditures. Data on low levels of performance on standardized tests by children in Head Start programs should not be interpreted as demonstrating pathology in these children or their families: The skills that are assessed are products of experiences that have been rare for many children, that reflect values that may not be part of the cultural tradition of some parents, and that depend on patterns of interaction that may be difficult in the context of the stresses of poverty. Nevertheless, success in school depends on these skills. Children who do not acquire them are at educational risk in U.S. culture, regardless of their ethnic, racial, linguistic, or class backgrounds.

For the present project, we had two goals. First, we hoped to provide stronger support of the model depicted in Figure 1 than is currently available by using an experimental design to vary elements of preschool literacy experience, by examining outcomes longitudinally, and by scrutinizing the interrelationships of several components of preschool literacy experience, emergent literacy abilities, and reading. Second, we hoped to make it easier for Head Start and other preschool programs to enhance children's language and literacy skills by developing a practical, effective emergent literacy program that could be added to existing preschool curricula.

In the present study, 4-year-olds attending Head Start were randomly assigned by classroom to experience the add-on emergent literacy curriculum or to experience an unaltered Head Start curriculum. Children were pretested and posttested on various standardized tests of language and emergent literacy skills. In this article, we summarize the outcomes from the project through the end of the Head Start year. In subsequent articles we will examine the outcomes in kindergarten and elementary school and in replications. Thus, this article is relevant only to the linkages between the first two constructs in Figure 1.

The emergent literacy curriculum was based on two components. The first was centered on an interactive style of adult–child shared picture book reading called *dialogic reading*. There is now a relatively large literature demonstrating that dialogic reading increases the rate of language development in preschoolers with normal development or with developmental disabilities; from families with low incomes and from families with upper middle incomes; in settings as varied as homes, preschools, daycare centers, and libraries; and for children who speak Spanish as well as for children who speak English (D. H. Arnold, Lonigan, Whitehurst, & Epstein, 1994; D. S. Arnold & Whitehurst, 1994; Dale, Notari, Crain-Thoreson, & Cole, 1993; Lonigan, 1993; Morriset, 1993; Novak, Staggs, & Jones, 1993; Valdez-Menchaca & Whitehurst, 1992; Whitehurst et al., 1988, 1994). These previous studies have not involved children in Head Start and have had shorter intervention periods than what we used in the present research. The second component of the curriculum was a program to teach children about the phonemic structure of language and the relation between phonemes and letters (e.g., that the words *sun* and *sand* begin with the same sound, that the sound is /s/, and that the letter *s* represents that sound). There is a large body of research suggesting that phonemic awareness is an important predictor of individual differences in early reading achievement (e.g., Bradley & Bryant, 1983; Lundberg, Frost, & Petersen, 1988; Mann, 1993; Stanovich & Siegel, 1994). We adapted to our use a commercially available curriculum called *sound foundations* (Byrne & Fielding-Barnsley, 1992) that has been demonstrated to produce long-term effects in Australian children on literacy measures such as spelling and pseudoword reading (Byrne & Fielding-Barnsley, 1991, 1993).

Our hypothesis was that the combination of dialogic reading and sound foundations would enhance language, linguistic awareness, and print knowledge when introduced into intervention classrooms and compared with the performance of children in control classrooms. We hypothesized that the effect of the intervention might be modulated by individual differences in characteristics of children, their families, or aspects of their classroom experience.

Method

Subjects

The sample consisted of 167 4-year-olds who attended classrooms in selected Head Start centers in Suffolk County, New York, during 1992–1993, whose enrollment began between September 1992 and December 1992, who continued to attend Head Start through the end of May of 1993, and whose birthdates made them eligible for entry into public kindergarten for the 1993–1994 academic year. The racial composition of the sample of children was 46% Caucasian, 45% African American, 8% Latin American, and 1% Asian. The racial or ethnic composition of Head Start nationally is 33% Caucasian, 38% African American, 22% Latin

American, and 3% Asian (U.S. Department of Health and Human Services, 1992), so the current sample underrepresented Latin Americans. Boys constituted 56% of the sample.

Suffolk County, the easternmost of the two counties that constitute Long Island, has a population of 1.3 million. Although sections of the north shore of the county and the east end are among the most affluent in the nation, the center of the county and some areas of the south shore include large populations living in conditions of poverty. Living costs, particularly housing costs, are very high. The median reported family income of children attending Head Start in Suffolk County, New York during the 1992–1993 year was $10,500, as determined by Head Start records.

Other demographic information on the sample was gathered from a self-report instrument, the Stony Brook Family Reading Survey (Whitehurst, 1992b), that was administered to each child's primary caregiver at the beginning of the Head Start year. Parents or primary caregivers received $10 for completing the survey, which was obtained for 92% of the children in the sample. Information from the survey indicated that 90% of the primary caretakers were mothers; 5% were fathers; 3% were grandparents, and 2% were individuals not biologically related to the child. Fifty-three percent of the primary caretakers were in the home full time; 28% were out of the home from 4–8 hr per day, and 19% were out of the home more than 8 hr per day. The mean number of children per family, including the target child, was 2.7 (SD = 1.5). The mean number of adults per family was 2.0 (SD = 1.0). Of the primary caregivers, 29% had not completed high school; 23% had a high school degree; 46% had some college or trade school coursework, and 2% had completed college. The language usually spoken in the home was English for 94% of the families. We know of no national data on Head Start families that would allow comparison with these demographics, but we expect that the present sample may be above the national mean on educational level.

Another 40 children were included in the sample at the outset but left Head Start prior to the end of the school year and could not be posttested. These children scored significantly lower on the battery of pretest assessments than the 167 children who completed the entire year. Differences between the dropouts of the intervention condition (N = 15) and control condition (N = 25) on a principal component weighted average of all the pretest assessments (described in more detail subsequently) did not approach statistical significance (p = .938).

Six additional children who otherwise met the criteria for participation in the study were excluded from the research design because they were unable to produce any expressive responses in English on the pretest battery. Five of these children were monolingual in a language other than English (3 Spanish, 1 Arabic, 1 Hindi), and one child from an English-speaking home was electively mute.

Procedure

Design. Four Head Start centers were selected to participate in the study on the basis of their geographical proximity to Stony Brook University. Three of these centers were large, long-standing operations that included multiple classrooms of 3- and 4-year-olds. The fourth center consisted of two geographically separated single classrooms that were new branches of existing centers. For purposes of research design and data analysis, the two outlying classrooms were combined into one center unit. Within each center, all classrooms that included a majority of 4-year-olds participated in the study.

Classrooms were randomly assigned to intervention and control conditions. We used a pairing procedure to ensure that the number of classrooms in each condition was as equivalent as possible, to ensure that each center had both intervention and control classrooms, and to ensure balance across conditions in classrooms in which students attended for half of the day versus classrooms in which students attended for a full day. Within each center, pairs of teachers of full-day classrooms were formed randomly, and then one member of each pair was assigned randomly to the intervention condition. Teachers of full-day classrooms that remained after this procedure because of an odd number of such classrooms at a given center were paired randomly with the remaining teachers of full-day classrooms from other centers, and then one member of each of those pairs was assigned randomly to the intervention condition. There were not enough teachers of half-day classrooms to pair within centers, so pairs of such teachers were formed randomly across centers and then one member of each of those pairs was assigned randomly to the intervention condition. Because there was an odd number of half-day classrooms in all, one teacher remained after all pairs were formed; this teacher was assigned randomly to the control condition. These procedures resulted in seven classrooms (four full-day, three half-day) in the intervention condition, and eight classrooms (four full-day, four half-day) in the control condition. At initial assignment, there were 110 children in the intervention condition and 97 in the control condition. Because of the higher frequency of dropout from the control classrooms, the final Ns were more disparate (intervention N = 94; control N = 73). To the extent that classroom size should be negatively correlated with children's learning, this should represent a conservative bias with respect to the intervention.

Each classroom was rated on the Early Childhood Environment Rating Scale (ECERS; Harms & Clifford, 1980), which produces scores in seven areas of preschool practice on a scale ranging from *low* (1) to *high* (7). There were no differences approaching significance in any of these seven areas as a function of assignment to intervention versus control conditions of the study (all ps > .25). The summary score across all the subscales of the ECERS ranged across classrooms from 3.9 to 5.4 (M = 4.62). According to Harms and Clifford, a summary score of at least 5.0 is the criterion for developmentally appropriate practice. The areas of practice with the greatest variability across classrooms and the lowest scores for particular classrooms were in the provision for adult needs (e.g., having a separate well-furnished lounge area for teachers) and social development (e.g., having ample and varied toys, games, and equipment for free play).

Assessment. Children were pretested at their Head Start centers between September 14, 1992 and October 23, 1992, at which point the classroom intervention began. Children who were latecomers to participating Head Start classrooms continued to be pretested and added to the sample cohort through December 21, 1992. To maintain the logic of a random assignment design, we maintained children who switched classrooms after December 21, 1992 (five in all) in their initial condition assignment for purposes of data analysis. This is a conservative procedure for tests of effects of the intervention.

At pretest, children were administered the Peabody Picture Vocabulary Test—Revised, Form M (PPVT–R; Dunn & Dunn, 1981), a test of receptive vocabulary, the Expressive One Word Picture Vocabulary Test (One Word; Gardner, 1981), a test of expressive vocabulary, the expressive subscale of the Illinois Test of Psycholinguistic Abilities (ITPA; Kirk, McCarthy, & Kirk, 1968), a test of verbal fluency in describing common objects, and 18 subscales from the Developing Skills Checklist (DSC; CTB, 1990) that measure emergent literacy skills such as the abilities to

name letters, segment words into constituent sounds, and identify the function of words and numbers (see Tables 1 and 2). Despite its name, the DSC is not a checklist: It consists of various questions in each domain of assessment that are administered to children individually in the typical style of a standardized test of language or intelligence. Age-corrected language quotients were used for the language tests, and raw scores were used for the DSC subtests: the DSC does not provide standardized scores for subtests; it provides norms by time of school year (e.g., spring of prekindergarten) rather than by the age of the child.

Each of the assessment instruments is standardized and normed on a nationally representative sample. Split-half reliability is high for each measure (PPVT–R = .80, One Word = .94, ITPA = .86, and DSC = .84). Other measures of reliability and appropriate validity studies are reported in the manual for each instrument. Together they assess a broad range of skills and knowledge in the areas of language and print.

During the period in which children were being pretested, each child's principal caregiver was administered the Stony Brook Family Reading Survey (Whitehurst, 1992b), which has a multiple-choice format and assesses home demographic and literacy related variables. Items from this instrument have been shown to correlate strongly with children's language and emergent literacy skills (Payne et al., in press). At the same time, the principal caregiver completed a pencil-and-paper adaptation of the Quick Test (Ammons & Ammons, 1962), which is a test of adult IQ that correlates highly with scores from individually administered IQ tests such as the Stanford–Binet. In the standard administration of the Quick Test, the examinee selects from a plate of four pictures the item that matches the word that is presented both in printed form and orally by an examiner. In our adaptation, the examinee received the words on a printed list but did not hear them pronounced by the examiner. Although this may have had some effect on the mean scores of the primary caretakers, our concern was not with the mean or with any examinee's absolute score but only with relative differences among the examinees that could be used in regression analyses to estimate the effects of caregivers' intelligence on children's language skills and emergent literacy abilities. The split-half reliability for our version of the Quick Test is .90, as determined through administration to 236 Head Start primary caregivers (a superset of the current sample).

Table 1
Factor Loadings for Posttest Outcome Variables

Variables	Factor loadings			
	Language	Writing	Linguistic awareness	Print concepts
One word	0.64	—	—	—
PPVT–R	0.71	—	—	—
ITPA	0.70	—	—	—
Tell story in sequence	0.49	—	—	—
Name letters	—	—	—	0.58
Identify sounds and letters	—	—	0.53	—
Blend CVC words	—	−0.47	—	0.45
Identify same–different sounds	—	—	0.63	—
Segment sentences	—	—	0.74	—
Segment words	—	—	0.64	—
Rhyming	—	—	—	0.47
Hold book–turn pages	—	—	—	—
Identify people reading	—	—	—	0.40
Distinguish words–pictures–numbers	—	—	—	0.57
Identify function of words–numbers	0.40	—	—	0.50
Identify components of writing	—	—	—	0.48
Print in left–right progression	—	0.77	—	—
Print first name	—	0.56	—	—
Draw a person	—	—	—	—
Write message mechanics	—	0.56	—	—
Write message quality	—	—	—	—
Proportion total variance	0.11	0.09	0.10	0.11

Note. Dashes indicate factor loadings of less than 0.40. PPVT–R = Peabody Picture Vocabulary Test—Revised (Dunn & Dunn, 1981); ITPA = Illinois Test of Psycholinguistic Abilities (Kirk, McCarthy, & Kirk, 1968); CVC = consonant–vowel–consonant.

Children were posttested between May 19, 1993 and June 21, 1993 with the same instruments with which they had been pretested, with the exception that alternate forms were used whenever they were available: Form L of the PPVT–R was substituted for Form M, and the revised version of the One Word (Gardner, 1990) was substituted for the original version (Gardner, 1981). Examiners at both pretest and posttest were doctoral students in clinical psychology with extensive experience in assessing children. Testing for each child occurred across two sessions at both pretest and posttest in a quiet area of the Head Start center away from other children. Examiners were unaware of the assignment of children to condition.

Intervention. The first element of the intervention, as presented in Table 3, was an interactive book reading program for children at home and at school called dialogic reading. The program featured small group reading in the classroom (e.g., four children to one adult, with a frequency of three to five times per week) and involved one-on-one reading at home with the same books that were used in the classroom. Both parents and teachers were trained in how to read dialogically by means of a 20-min video (Whitehurst, 1992a), which was coupled with brief role-playing and discussion following exposure to the video. We have shown previously that video training is an even more effective mechanism for imparting the principles of dialogic reading to parents than is direct one-on-one training by a professional (D. H. Arnold et al., 1994).

Parent and teacher training occurred only once at the beginning of the school year. The first wave of parents was trained at parent meetings in their children's Head Start centers. We organized and conducted the training sessions, which included a free meal and child care, were well advertised by the director of each center, and were scheduled on at least two different occasions at each center. Of parents who were eventually trained, about half (52%) attended these meetings. Training of the remaining half of the parents or caregivers who were eventually trained was conducted by a member of the center's Head Start staff, who was paid $25 per family to conduct training. Typically, this involved the trainer making an individual appointment with the parent or caregiver for training at the Head Start center, but on occasion it necessitated a home visit. Ultimately, 89% of primary caregivers received training. Children of untrained parents were retained in the intervention condition for all analyses. As with the previously described treatment of children who switched classrooms late in the year, our aim was to retain the logic of a random assignment design and to maintain a conservative bias with respect to the evaluation of treatment effects.

In the dialogic reading program we used 30 books over the course of the school year, one book per week. The books are commercially available, but we altered them by adding hints for

Table 2
Means, Standard Errors of the Mean, and Norms for Test Battery

Variables	Pretest means			Posttest means			
	Control	Intervention	SE_m	Control	Intervention	SE_m	Norms[a]
One word	89.07	91.06	1.07	84.26	87.33	1.14	100.00
PPVT–R	85.88	88.23	1.13	87.90	89.74	1.02	100.00
ITPA	106.32	104.09	1.31	104.73	103.88	1.35	100.00
Tell story in sequence	0.00	0.03	0.01	0.02	0.09	0.02	0.49
Name letters	1.14	1.71	0.18	3.82	5.45	0.32	7.32
Identify sounds and letters	0.14	0.06	0.03	0.37	0.86	0.12	2.64
Blend CVC words	0.00	0.00	0.00	0.00	0.01	0.01	0.39
Identify same–different sounds	5.16	5.15	0.16	5.31	5.33	0.17	6.30
Segment sentences	0.12	0.06	0.03	0.64	0.59	0.09	1.48
Segment words	2.00	1.85	0.11	2.72	2.80	0.11	3.05
Rhyming	1.16	1.01	0.07	1.14	1.50	0.08	1.98
Hold book–turn pages	0.84	0.74	0.03	0.87	0.91	0.02	0.95
Identify people reading	1.99	1.98	0.09	2.59	2.49	0.05	2.70
Distinguish words–pictures–numbers	2.95	3.11	0.08	3.73	4.01	0.09	4.92
Identify function of words–numbers	1.33	1.02	0.08	2.25	2.57	0.10	3.00
Identify components of writing	1.21	1.37	0.05	1.60	1.76	0.07	2.70
Print in left–right progression	0.67	0.74	0.04	0.86	0.91	0.02	0.90
Print first name	1.12	1.20	0.05	1.89	2.16	0.09	2.38
Draw a person	1.51	1.66	0.05	2.26	2.39	0.07	2.33
Write message mechanics	1.51	1.55	0.07	2.14	2.33	0.08	2.09
Write message quality	2.23	1.89	0.12	2.41	1.82	0.12	2.80

Note. All values are uncorrected for the pretest covariate and are not conditionalized on center or sex; as such, they are different from the analysis of covariance results reported in the text. PPVT–R = Peabody Picture Vocabulary Test—Revised (Dunn & Dunn, 1981); ITPA = Illinois Test of Psycholinguistic Abilities (Kirk, McCarthy, & Kirk, 1968); CVC = consonant–vowel–consonant.

[a] Norms are from tabled values provided by the test publisher. Norms for the Developing Skills Checklist (CTB, 1990) subtests are available only for the spring of the pre-kindergarten year, which is the time at which posttesting occurred.

wh- prompts on each page and hints for recall prompts at the back of the book (*wh-* prompts and recall prompts are defined below). Each book was accompanied by a book guide that explained the story or purpose of the book and offered hints for how to introduce and read it, which were modeled on materials developed by Karweit (1989). There were versions of the book guides for parents and teachers. The two versions differed primarily in the inclusion of extra material for teachers related to classroom activities involving the book.

The at-home reading program was facilitated by a lending library that was administered through each Head Start classroom. On the first day of each week, children were sent home with a copy of the book that was being used in their classroom reading group for that week, along with the related book guide. Parents or caregivers were encouraged through the book guide to try to read the book with their child at least three times during the week and to return the book to school with the child on the last day of the week. Teachers sent home reminders for books that were not returned, and recalcitrant parents or caregivers eventually received a notice that they would be billed for the costs of unreturned books. Nearly all books were eventually returned, and threatened fines or penalties were never actually invoked. By rotating books across classrooms and Head Start centers, it was possible to cover the needs of approximately 100 children in seven classrooms with 25–30 copies of each book. Although no class size exceeded 18 children, the additional 7–12 copies of each book were necessary to fill the gaps left when books needed in a particular classroom had not yet been returned by parents of children in another classroom.

A major goal of dialogic reading is to make children active participants in shared picture book reading rather than passive listeners to stories being read by adults (D. S. Arnold & Whitehurst, 1994). The principal mechanism for achieving this is questioning by the adult reader. The acronym CROWD, as presented in the training video, was designed to help adult readers remember five types of questions they could pose to children while sharing a picture book with them: *Completion prompts* were fill-in-the-blank questions, such as "Something went bump, and that made us _____?" *Recall prompts* were questions that required the children to remember aspects of a book (e.g., "Can you remember some things that happened to Lena when she went to school?"). *Open-ended prompts* were statements that encouraged the child to respond in his or her own words (e.g., "I told about the last page, now it's your turn. You tell me about this page."). *Wh-prompts* were *what, where,* and *why* questions (e.g., "What's this called?"). *Distancing prompts* were questions that required the child to relate

Table 3
Emergent Literacy Intervention in Head Start

Intervention components	Location	Scheduling–Number	Episode duration	Program duration
		Dialogic reading		
Small group reading	Classroom	2–5 per week	10–15 min	30 weeks
One-on-one reading	Home	? per week	?	30 weeks
Video training for teachers	School	Once	1 hr	1 hr
Video training for parents	School or home	Once	1 hr	1 hr
Altered books	School and home	One per week/30	—	30 weeks
Book guides	School and home	One per week/30	—	30 weeks
		Sound foundations		
Sound posters	Classroom	Once per week/16	10–15 min	16 weeks
Worksheets	Classroom	Once per week/16	10–15 min	16 weeks
Extension activities	Classroom	1–2 per week/32	10–15 min	16 weeks

Note. For dialogic reading, the targets were language and print skills. For sound foundations, the targets were the phonemic structure of words and the alphabetic principle of spelling. Dashes indicate that the data are not applicable. Question marks indicate that the frequency and duration of these activities were not assessed.

the content of the book to aspects of life outside the book (e.g., "Did you ever play in the snow like Peter did? What did it feel like?").

The acronym PEER, as presented in the training video, was intended to help adult readers remember to embed the five types of questions outlined above into interaction sequences in which the adult *prompted* the child to respond to the book, *evaluated* the child's response, *expanded* the child's response by repeating and adding information to it, and encouraged the child to *repeat* the expanded utterance. As an example of such a sequence, the adult reader might prompt the child by saying "What does it feel like to play in the snow?" The child might respond by saying "It's cold." The adult might evaluate this response as being correct but shorter and less detailed than desirable and respond with an expansion such as, "Yes, it's cold when your feet get wet or when someone hits you with a snowball." The adult might encourage the child to repeat some of this new information by waiting until the end of the book and then asking the child a recall question: "Can you remember some things we talked about that make people cold when they play in the snow?"

The second element of the intervention was our adaptation of sound foundations (Byrne & Fielding-Barnsley, 1992), a phonemic awareness curriculum that was developed and tested in Australia (Byrne & Fielding-Barnsley, 1991, 1993). The sound foundations curriculum began in February in each classroom and continued through June. Teachers introduced children to seven consonant sounds *(s, m, p, g, l, t, sh)* in the beginning and ending positions in words and to two vowel sounds (*a* and *e*) in the beginning position only. As an example of how this element of the curriculum worked, on Monday of beginning /s/ week, children were invited individually to walk forward from the group and to find objects with names beginning with the /s/ sound (e.g., seal or scissors) in a large colorful poster. On Wednesday, the students were asked to find objects on their own worksheets that began with the /s/ sound and to color the objects with crayons. On Friday, the students were asked to play a version of the game hot potato in which the children sat in a circle and passed a bean bag to each other while music was playing. The child who was holding the bean bag when *the music stopped had to find something in* the classroom the name of which began with /s/. Our adaptation of sound foundations included introducing the children to the manuscript letters that correspond to the curriculum sounds. We incorporated letters into the curriculum by placing a large poster with the letter for the week in the classroom and having the teacher give completion prompts to the children in the form, "This is the letter *S;* it says ____." We dropped from the standard sound foundations curriculum audio tapes containing rhymes and stories because the accents of the readers and much of the tapes' content were distinctly Australian. We developed games and extension activities as substitutes for the card games recommended for use with sound foundations. These card games were used early in the intervention but were then dropped as they were judged by teachers to be too difficult for the children in their classrooms. Please note that sound foundations is not a beginning reading program. It simply prepares the child for later reading instruction by introducing the basic phonemic and alphabetic principles of reading.

Compliance. Teachers and aides in intervention classrooms were given a daily log for recording to whom they read and the nature of the reading activity (e.g., small group vs. large group; interactive vs. straight reading).

One of us visited classrooms at least once every 2 weeks to check compliance and to answer questions or give feedback as needed. In addition, we met with parent trainers from each center, at a Head Start location every 4 to 6 weeks to discuss the project and any problems related to it.

Each classroom teacher and assistant was videotaped during a shared book reading session during the late spring of the school year. These were regular dialogic reading sessions for intervention classrooms. Teachers in control classrooms understood that the videotaping was to compare reading styles across classrooms. Teachers in control classrooms were asked to read in their typical manner. The videotapes were coded for major categories of teacher behavior that are relevant to dialogic reading and were compared across the intervention and control conditions. So that we could control for book effects, we had teachers in control classrooms read unaltered versions of the books that were used in the corresponding intervention classrooms during the week when videotaping occurred.

To assess for the effects of compliance by caregivers with the at-home dialogic reading program, we sent a follow-up survey in

May of the school year to caregivers whose children were in the intervention condition. The survey was sent home to caregivers with children in their school bags on two separate occasions and included the offer of a free children's book for responding. The survey was subsequently mailed to nonresponders, and we also attempted to phone caretakers who had not responded. Return rate on the survey was 63%. The follow-up survey contained some direct questions about compliance with the dialogic reading program (e.g., "How frequently did you or other family members read with your child?"). The survey also included a list of 28 picture books by title and author and asked for ratings on a Likert-type scale, ranging from *didn't like at all* (1) to *enjoyed very much* (5), as to how much the caregiver liked using each book. Each rating scale had a place for the caregiver to check if she did not remember the book, and the lead paragraph indicated that every book on the list had not been received by every child. Half of the book titles on the list had actually been sent home to the parents, and half had not. The book identification rate (encountered books rated minus unencountered books rated) served as an unobtrusive measure of each caregiver's involvement in the at-home reading program.

Results

Data Reduction

Given the number of assessments that we conducted at pretest and posttest on each child, we used a method of data reduction to decrease the total number of statistical tests conducted and to enhance the comprehensibility of outcomes. A principal components analysis was conducted on the 21 language and literacy measures that were obtained on each child at posttest. On the basis of the scree test (Cattell, 1966) for determining the number of factors to retain and the relationship of the factors to those that have emerged in previous research we extracted four factors from the posttest measures. We rotated these factors with the varimax normalized method, which accounted for 41% of the total variance in the outcome scores. Table 1 indicates that the loadings on these four factors were relatively clean and interpretable. The highest loadings on the Language factor were on the three language tests and the story telling task. The highest loadings on the Writing factor were on the child's ability to print from left to right, the child's ability to print his or her first name, and the mechanics of writing. The highest loadings on the Linguistic Awareness factor were on identifying sounds and letters, identifying sounds as same or different, segmenting sentences into words, and segmenting words into syllables. The Print Concepts factor had high loadings on typical measures of concepts of print (naming letters, identifying people reading, distinguishing between words, pictures, and numbers, and identifying components of writing). Less neatly, the Print Concepts factor also had loadings on a measure of rhyming and on a measure of sound blending, which are usually categorized as measures of linguistic awareness. Of these, sound blending was subject to a floor effect (see Table 1), thus, its loading can be discounted. Note that the loadings for the four factors and their interpretation are reasonably consistent with the model of emergent literacy ability that is pictured in Figure 1.

We also applied a principal components analysis to the pretest measures. Here our purpose was to extract a covariate that would reduce error variance in subsequent analyses and correct for any imbalances across groups at pretest that might have occurred despite random assignment. Accordingly, we extracted a single unrotated factor, which accounted for 18% of the total variance in pretest scores. This single pretest covariate had the desirable feature of correlating significantly with each of the four outcome factors (r = .43, .22, .33, and .42 for Language, Writing, Linguistic Awareness, and Print Concepts, respectively).

Outcomes Analysis

To assess the effects of the emergent literacy intervention on outcome scores, we conducted a 2 (intervention vs. control) × 2 (sex) × 4 (Head Start centers) multivariate analysis of covariance (MANCOVA), using the pretest factor as a covariate and the four posttest factors as dependent variables. Sex was included as a variable because of the inherent interest in its effects and because preliminary analyses had shown it to control significant outcome variance in Writing. Head Start center was included as a variable to increase the precision of the analysis: The four centers differed in ways that were observationally obvious to casual observers, and preliminary analyses had shown center to control significant variance in the outcome measures. The MANCOVA analysis resulted in significant main effects for intervention versus control, Wilks's λ = .86, df = 4, 147, p = .0002, for Head Start center, Wilks's λ = .80, df = 12,389, p = .001, and for sex, Wilks's λ = .91, df = 4,147, p = .006. None of the interaction terms approached significance (all ps > .15).

In the context of the preceding analysis, we followed up the sex and intervention effects by examining specific effects on the four posttest factors. The center effect was not followed up because we had no general interest in the relative performance of the various centers. The only significant effect for sex was that girls performed at a higher level than did boys on the Writing factor, $F(1, 150)$ = 11.77, p = .0008. Boys performed better than did girls on the Language factor, $F(1, 150)$ = 3.48, p = .064. The effects of sex on the Linguistic Awareness and Print Concepts factors did not approach significance (both ps > .30). Analysis of the intervention effect indicated that children in the intervention condition performed at a significantly higher level than did children in the control condition on the Writing factor, $F(1, 150)$ = 7.98, p = .005, and the Print Concepts factor, $F(1, 150)$ = 10.35, p = .002. Differences on the Language and Linguistic Awareness factors did not approach significance (both ps > .70). We qualify the latter conclusion by noting that specific effects on linguistic awareness may have been masked by the use of broad factors as outcome measures. In particular, there was a significant effect of the intervention, $F(1, 150)$ = 5.70, p = .018, on the Identify Sounds and Letters subtest (see Tables 1 and 2), although there is no overall effect on the Linguistic Awareness factor on which this subtest loads most highly.

Of all the subtests, this one is the most directly relevant to the content of sound foundations because it requires that the child first indicate the beginning sound of the name of a pictured object and then indicate the letter that corresponds to that sound.

The intervention effect on the four outcome factors is displayed in Figure 2, which displays the marginal means, uncorrected for the pretest covariate. Because the factor scores are standardized, Cohen's (1988) effect-size index *d* can be read directly from the figure as the difference between intervention and control means. The effect sizes of .516 for the Writing factor and .624 for the Print Concepts factor fall into Cohen's medium effect size category and have been interpreted in the context of previous Head Start research as indicating educationally meaningful effects (e.g., Lee, Brooks-Gunn, Schnur, & Liaw, 1990).

With the individual as the unit of analysis, as in the previous analyses, it is problematic when an intervention has occurred in a group setting: If individuals are responding in concert, then error terms will be correlated and the probability of Type I error will increase beyond the stated alpha level (Kenny, 1988, Levin & Serlin, 1993). The appropriate unit of analysis for the present study is not obvious because some components of the intervention were conducted on an individual basis (at-home dialogue reading), others on a small group basis (in-school dialogic reading), and others on a classroom basis (sound foundations). Nevertheless, it is important to protect against the possibility that group effects were constraining error terms in the previous analyses and leading us to inappropriately reject the null hypothesis. Accordingly, we used classroom means as the unit of analysis and conducted a 2 (intervention vs. control) × 2 (sex) × 4 (Head Start centers) analysis of variance (ANOVA). The intervention and center variables were between-classroom factors, and the sex variable was a within-classroom factor. The pattern of significant effects was identical to that reported previously with the individual child as the unit of analysis: The only significant effect for sex was that girls performed at a higher level than did boys on the Writing factor, $F(1, 7) = 18.32$, $p = .004$. The effects of sex on the Language, Linguistic Awareness, and Print Concepts were not significant (all $ps > .40$). Analysis of the intervention effect indicated that children in the intervention condition performed at a significantly higher level than did children in the control condition on the Writing factor, $F(1, 7) = 13.77$, $p = .008$, and the Print Concepts factor, $F(1, 7) = 5.10$, $p = .058$. Differences on the Language and Linguistic Awareness factors were not significant (both $ps > .15$). The similarity of the present results with those from the previous analysis with the individual child as the unit demonstrates that the conclusions from the previous analysis were not compromised by restricted error terms arising from group processes at the classroom level.

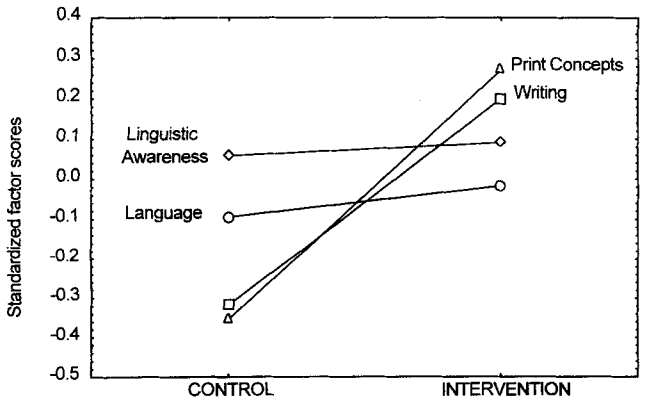

Figure 2. Standardized factor outcomes scores as a function of membership in the intervention or control condition (marginal means from Condition × Sex × Center analyses of covariance reported in the text using the individual child as the unit of analysis, uncorrected for pretest covariate).

A difficulty with factor scores such as those represented in Figure 2 is that the absolute scale of performance is lost. Thus, although we can determine from Figure 2 that children in the intervention condition performed about a half a standard deviation better at posttest than did children in the control condition on the Writing factor, it would be useful for us to anchor those differences to a meaningful standard. This cannot be done precisely: First, the DSC test publisher does not provide standard deviations for the normative sample at the level of individual subtests. Second, and more important, factor scores are weighted composites that reflect shared variance in the outcome measures and that capture only a portion of the total variance in those measures. The factor scores are only imperfectly related to any of their individual constituents. Nevertheless, a general sense of the absolute level of performance of children can be derived from examination of the means for the intervention and control conditions for each individual subtest and comparison of these scores with the means from the national norms. This information, which can be found in Table 2, indicates that although children in the intervention condition performed better than did children in the control condition on nearly every subtest, children in both conditions still fell substantially below the U.S. national averages on nearly every subtest at exit from Head Start.

Compliance

In the home. We hypothesized that the effectiveness of the intervention might be influenced by the degree to which children actually experienced it in the form in which it was designed. With respect to the at-home component of the intervention, we analyzed two responses to the follow-up survey that we obtained from the primary caregivers at the end of the Head Start year ($N = 59$). The first (identification rate) was the *z*-score of the hit rate by the primary caregiver in rating books that had been sent home as part of the shared reading program minus the *z*-score for the number of false alarms by the caregiver in rating books that had not been sent home. The second (reading frequency) was the primary caregiver's rating, on a 4-point scale ranging from *hardly ever* (1) to *almost daily* (4), of how often she or another family read to the child.

A preliminary examination of variable correlations indicated that identification rate and reading frequency were correlated with the Language factor ($r = .38$ and $r = .39$, respectively) but not with the other three outcome factors (all $rs < .14$). Thus, we limited subsequent analyses to the relationship between the two compliance measures and Language outcome.

Because the two measures of compliance, identification rate and reading frequency, might mark family characteristics that existed before the intervention rather than the family's compliance with the intervention, we examined correlations between these two variables and three family measures that were obtained at pretest: primary caregiver IQ, primary caregiver education, and reported frequency of shared reading. Reading frequency at pretest was correlated with reading frequency at follow-up ($r = .60$), but no other correlation approached significance (all $rs < .14$). Accordingly, we did not consider IQ and education any further.

To control for the effects of pretest reading frequency and each child's entering abilities, we conducted a hierarchical linear regression analysis with the Language factor score as the dependent variable. Each child's pretest factor score and the primary caregiver's reported frequency of shared reading at pretest were entered first into the equation. At the next step, identification rate, and frequency of reading at follow-up were entered. Pretest reading frequency did not contribute significantly to the prediction equation ($\beta = .088$, $p = .442$), but pretest factor score did ($\beta = .463$, $p < .001$). Identification rate and reading frequency added significantly to the prediction equation (R^2 change $= .115$, $p = .006$), and each had a significant beta weight (identification rate: $\beta = .232$, $p = .031$; reading frequency: $\beta = .289$, $p = .029$). Thus it appears that compliance with the at-home component of the intervention was significantly related to language outcome after controlling for other variables.

Because this is a potentially important finding, we explored it further by creating a single compliance score for each caregiver and relating it to language outcome. We obtained the composite compliance variable by weighting the identification rate and reading frequency scores for each caregiver and then summing the two variables, with weights derived from the linear regression of Language factor outcome score onto identification rate and reading frequency ($R = .507$, $p = .0002$). On the basis of natural breaks in the distribution of the composite compliance scores, we recoded compliance on an ordinal scale with five levels. Figure 3 plots these levels of compliance against mean Language factor outcome scores (corrected for differences in the pretest factor covariate). Note that the function that relates these two variables is a steep and virtually straight line.

At school. The frequency with which teachers and aides in the intervention classrooms reported engaging in reading interactions with children appeared to be high, ranging from 61% to 95% of available days. Because weekly reading logs sometimes were not completed by teachers or sometimes were filled out retrospectively, we chose not to conduct analyses using variables from that source. Instead, we examined the effects of each child's daily attendance, as

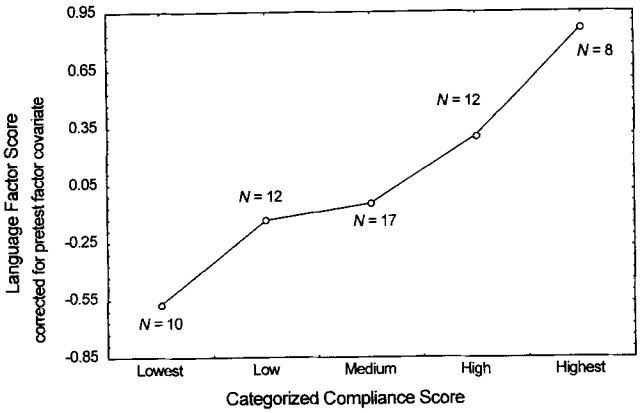

Figure 3. Standardized language factor outcome scores as a function of compliance by primary caregiver with at-home reading program (using the individual child as the unit of analysis and corrected for pretest factor covariate).

determined by daily attendance sheets in each Head Start center's central records. There was a significant correlation between the number of days that children in the intervention classrooms attended school and the Writing outcome factor ($r = .32$, $p = .001$). Attendance was not correlated with the other three outcome factors (all $rs < .17$; all $ps > .10$). However, this cannot be viewed as a reflection of an effect on the intervention because the same pattern of correlations was found among children in the control condition (e.g., $r = .30$, $p = .011$ for attendance and Writing factor score among children in the control condition). Thus, the more children attended Head Start, the more they learned about writing. This was true in both intervention and control classrooms, even though the mean level of performance was higher in the intervention than in the control classrooms.

Figure 4 represents the frequency of various categories of interactive teacher behavior that we observed during book reading by 22 teachers or assistants, half of whom were in the intervention classrooms and half of whom were in the

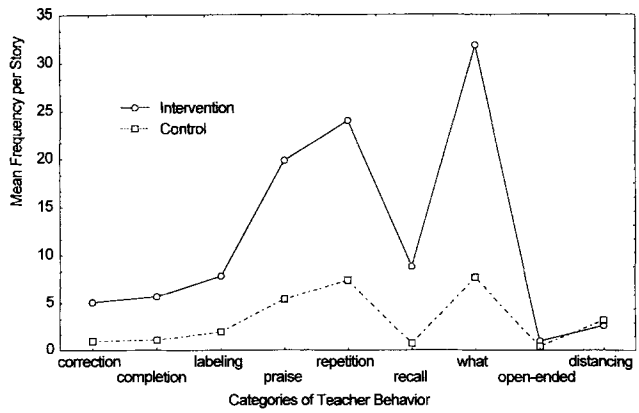

Figure 4. Mean frequency per story of nine categories of teacher behavior as a function of membership in the intervention or control condition.

control classrooms. The categories of behavior include correction (e.g., "I think that's a dog, not a cow."), completion prompts (e.g., "Donna is giving them pats on the _____."), labeling (e.g., "That's a vole."), praise (e.g., "That's right."), repetition (e.g., Child says, "That's snow," and the teacher says, "Snow?"), recall questions (e.g., "Do you remember what happens after Peter goes inside?"), what questions (e.g., "What are these called?"), open-ended prompts (e.g., "Tell me what's going on here."), and distancing prompts (e.g., "Has something like this ever happened to you?"). For most categories in Figure 4, there are very low frequencies in the control classrooms and substantially higher frequencies in the intervention classrooms. By and large, the teachers in the control classrooms only engaged in straight reading to the children in their classrooms, whereas the teachers in the intervention classrooms used a much more interactive reading style. The more complicated forms of dialogic question asking, such as distancing prompts and open-ended questions, occurred with very low frequencies in both the intervention and control classrooms. Because teachers in the intervention classrooms did not go beyond simple questions, the effects of the intervention may have been reduced.

We have only anecdotal and informal data on the degree of teacher compliance with the sound foundations component of the curriculum, that are based on the periodic meetings that were held between the research staff and Head Start staff and on the unsystematically scheduled classroom visits. It appeared that participation by teachers for the first day of each sound foundations week was nearly universal. On the first day of each week the children were involved in finding objects on a large colorful poster the names of which began or ended with the key sound for the week, and the corresponding letter for the week was introduced. Participation on the worksheet day seemed to be highly variable across classrooms but consistent within classrooms. Most teachers reacted very positively to the worksheets and indicated that their children enjoyed coloring them. A staff member at one center viewed them as developmentally inappropriate and resisted their use. Participation on the extension activity day of each week (e.g., a game of hot potato) appeared to be highly variable both within and across classrooms. Several options were given for extensive activities for each week. Teachers appeared to take that as a license to skip the activity or to create one of their own.

Discussion

Overall, these data demonstrate that educationally and statistically significant increments in children's emergent literacy abilities can result from relatively modest additions to the typical Head Start curriculum. The additions can be labeled as modest for three reasons. First, training teachers for the program took very little time and depended on a videotape rather than direct training by an expert. Likewise, parent training depended on one exposure to a brief videotape. Second, the materials were inexpensive, with the exception of multiple copies of books for children to take home. The lending library book costs were within the range of $4,000 to $6,000 for about 100 children in the 1st year of the program. Costs would decrease in subsequent years if book purchases were limited to replacements and additions to the original list. Third, in terms of classroom time, the dialogic reading program alone took no more than 2 hr of individual teacher time per week, and no more than 1 hr of individual child classroom time per week, spread over 7 months. The sound foundations component added no more than 45 min or so of classroom time per week for teachers and children over 4 months. A child who participated maximally over the course of the school year would have invested about 42 hr of time in the classroom program, or about 8% of the Head Start day for children attending half-day programs and 4% for children attending full-day programs.

The present study adds important information to the theoretical and empirical controversy about the role of shared book reading and other emergent literacy experiences in the development of language and emergent literacy skills (Dunning, Mason, & Stewart, 1994; Lonigan, 1994; Payne et al., in press; Scarborough & Dobrich, 1994). Recall that the measure of compliance with the at-home book reading program by primary caregivers in the intervention condition was correlated with language outcome scores at .51. Furthermore, the relationship between compliance with the dialogic reading program and language outcomes remained after being pitted against the effects of children's pretest skills, primary caregiver intelligence, primary caregiver education, and frequency of shared reading reported at pretest. These data demonstrate that Head Start children's language abilities can be increased by means of a conscientiously applied program of interactive book reading in the home. At the same time, effects on language were limited to children whose parents were actively involved in the at-home reading program. Given the extraordinary effort that was involved in reaching parents for training, it may prove difficult for typical Head Start centers to do more to increase the involvement of parents at the lower end of the dimension of participation.

It is important to note that classroom-based interactive reading did not, by itself, generate increases in children's language skills. We believe these results, when coupled with data from other studies, support a hypothesis with important implications for social and educational policy: Children from low-income families need frequent one-on-one language interactions with an adult to enhance their language skills. Group-based interactions may not be sufficient in the late preschool years, even if the groups are small and the forms of interaction are optimized.

Other studies support this hypothesis. Bryant et al. (in press) examined Head Start children's scores on cognitive, achievement, and language measures as a function of classroom quality and quality of the home environment. Bryant et al. concluded that "classroom quality seems more strongly related to the intelligence and achievement measures and home quality seems more strongly related to ratings of children's language" (ms. p. 18). Whitehurst et al. (1994) compared a center-based dialogic reading interven-

tion with a combined center- and home-based intervention for children from families with low incomes who were attending subsidized daycare. Children in the combined intervention were superior to those in the center-based approach at posttest. Lonigan (1993) replicated and extended the Whitehurst et al. (1994) study with a similar population but added a home-based intervention condition to both the center-based and combined center- and home-based conditions. Lonigan found that children in the home-based condition performed better on some measures than did children in either the center-based or combined intervention conditions. Finally, we can compare the variance accounted for by interventions such as the present one with the variance that is attributable to home-based literacy environment. Payne et al. (in press) bracketed the variance accounted for in language scores of Head Start children by literacy activities in the home as between 12.5% and 18%. Variance accounted for by successful shared-book reading interventions, including the present one, is consistently less (D. H. Arnold et al., 1994; Dale et al., 1993; Lonigan, 1993; McCormick & Mason, 1986; Morriset, 1993; Novak et al., 1993; Valdez-Menchaca & Whitehurst, 1992; Whitehurst et al., 1988, 1994).

Frequency of exposure may be the most straightforward explanation of why the home is a more powerful venue for language development than school. Within the population at large, there is the well-demonstrated negative relationship between family size and children's verbal and cognitive skills (Zajonc, 1976). Within the language acquisition literature there is a similarly well-established relationship between frequency of maternal speech to children and rate of children's language acquisition (Huttenlocher, Haight, Bryk, Seltzer, & Lyons, 1991). We think it is a small leap to visualize preschool or Head Start classrooms with 18 children as being similar to large families in which the relative frequency with which any individual child receives language input is necessarily diluted. Small group interactive shared book reading improves a child's chances for language input from an adult in a preschool setting. However, a child who participated maximally in the present center-based dialogic reading program still spent only about 30 hr in shared book reading over the course of the school year, and that occurred within a group with a minimum of three other children. We believe that interventions that have as their aim a meaningful increase in the language skills of children in the late preschool period who are from backgrounds of poverty may need to be focused on the home environment, with all of the difficulties that entails, or may need to have substantially increased opportunities for one-on-one interaction in the classroom (e.g., by using volunteers to read and talk with children individually).

We emphasize the late preschool period as one in which group interactions may be insufficient because we have shown previously that center-based approaches can have substantial effects on younger children (Valdez-Menchaca & Whitehurst, 1992; Whitehurst et al., 1994). If one imagines an inverted triangle as a container to be filled with language skills during each child's preschool years and a given group-based intervention as adding a constant amount to the contents of the container, then a given intervention will produce a larger relative increase in language skills when the child is younger and when there is less already in the container than when the child is older. Standardized tests of language ability are correspondingly more likely to detect an effect of the intervention in younger children because the items on the test are sampled from a relatively small pool of knowledge at the bottom of the container and are thus likely to include items that reflect what the child has acquired from the intervention. In contrast, standardized language tests of older preschoolers, which contain roughly the same number of items at the 4-year-old level as they do at the 2-year-old level are less likely to include items that test the particular vocabulary or language skill acquired from an intervention. We stress that this is not simply a measurement problem, though it is also that. More important, it reflects the positive correlation between age and how much children have to learn to catch up to or not fall further behind their peers.

The picture with respect to preacademic skills such as concepts of print, letter recognition, and writing is quite different from that for language. Regarding preacademic skills, children in the late preschool period are starting at the bottom of the inverted triangle of knowledge. A modest group-based intervention can add substantially to their preacademic skills in both absolute and relative terms. Correspondingly, standardized tests are likely to be sensitive to these changes. Consistent with that model, the present intervention had substantial effects on standardized measures of writing and concepts of print. The effect sizes of .52 for the Writing factor score and .62 for the Concepts of Print factor score in the present study compare favorably with the average effect size of .31 for readiness skills and .54 for achievement skills reported in the McKey et al. (1985) synthesis of studies in which children in Head Start were compared with children with no Head Start experience. In other words, in selected areas, the present emergent literacy add-on curriculum contributes as much to children's preacademic skills as the entire Head Start experience has been shown to contribute in earlier studies.

Given that promising outcome, it will be important for future researchers to address issues that the present study raises but does not answer. One issue is the relative contribution of the dialogic reading component versus the sound foundations component of the curriculum package to various components of emergent literacy. It appears that language skills are uniquely affected by at-home dialogic reading. However, the effects on the Writing and Print Concepts factor scores and the isolated effects in the area of linguistic awareness could conceivably have been a result of either or both components of the program. At some point, these two components of the curriculum need to be separated and analyzed.

Another issue to be addressed by future researchers is the long-term effects of the intervention (i.e., the effects of heightened emergent literacy skills on word decoding and reading comprehension in elementary school—see Figure 1). Follow-up of the present sample to address that issue is underway. A third important issue has to do with teacher

training and individual differences in teachers as reflected in the performance of the children in their classrooms. Although training of teachers through a brief videotape and simple written materials has the advantage of efficiency, it is our impression that a more systematic monitoring and feedback program, embedded within Head Start, would generate higher and more consistent levels of teacher performance.

Although the present intervention produced impressive gains on important groups of emergent literacy skills, it is important to note that this add-on curriculum fell short of bringing children in Head Start up to the typical level of performance of children of their age (see Table 2). It is interesting that some studies of children attending Head Start 20–25 years ago showed them close to the national mean on cognitive measures after exit (e.g., Lee et al., 1990). Perhaps children are entering Head Start with lower levels of skill than they did 25 years ago. Perhaps their families are in greater distress. Perhaps Head Start's curriculum is less focused and effective than it used to be. It is impossible to answer these questions because systematic collection of data on Head Start's effectiveness ceased over 2 decades ago. In any case, it is clear that children who begin Head Start, on average, a standard deviation behind other children of their age will need more than an add-on emergent literacy curriculum to close the gap completely. We believe focused interventions such as the present one can have a significant impact, but these interventions need to be part of a multifaceted effort to improve the quality and scope of Head Start (Zigler & Styfco, 1993).

References

Adams, M. J. (1990). *Learning to read: Thinking and learning about print.* Cambridge, MA: MIT Press.

Alexander, K. L., & Entwisle, D. R. (1988). Achievement in the first 2 years of school: Patterns and processes. *Monographs of the Society for Research in Child Development, 53* (2, Serial No. 218).

Ammons, R. B., & Ammons, C. H. (1962). *The Quick Test.* Psychological Test Specialists.

Arnold, D. H., Lonigan, C. J., Whitehurst, G. J., & Epstein, J. N. (1994). Accelerating language development through picture book reading: Replication and extension to a videotape training format. *Journal of Educational Psychology, 86,* 235–243.

Arnold, D. S., & Whitehurst, G. J. (1994). Accelerating language development through picture book reading. In D. Dickinson (Ed.), *Bridges to literacy: Approaches to supporting child and family literacy.* Cambridge, MA: Basil Blackwell.

Bradley, L., & Bryant, P. E. (1983). Categorizing sounds and learning to read—A causal connection. *Nature, 301,* 419–421.

Bryant, D. M., Lau, L. B., Burchinal, M., & Sparling, J. J. (in press). Family, teacher, and classroom correlates of Head Start children's developmental outcomes. *Early Childhood Research Quarterly.*

Bryant, D. M., Peisner-Feinberg, E. S., & Clifford, R. M. (1993, November). *Developmental outcomes of participation in public preschool programs.* Paper presented at the 2nd National Head Start Research Conference, Washington, DC.

Byrne, B., & Fielding-Barnsley, R. F. (1991). Evaluation of a program to teach phonemic awareness to young children. *Journal of Educational Psychology, 82,* 805–812.

Byrne, B., & Fielding-Barnsley, R. F. (1992). *Sound foundations: A scientifically researched method of introducing pre-reading skills.* Artarmon, Australia: Leyden.

Byrne, B., & Fielding-Barnsley, R. F. (1993). Evaluation of a program to teach phonemic awareness to young children: A one year follow-up. *Journal of Educational Psychology, 85,* 104–111.

Carnegie Foundation for the Advancement of Teaching. (1991). *Ready to learn: A mandate for the nation.* New York: NY: Carnegie Foundation for the Advancement of Teaching.

Cattell, R. F. (1966). The scree test for the number of factors. *Multivariate Behavioral Research, 1,* 245–276.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). New York: Academic Press.

CTB. (1990). *Developing skills checklist.* Monterey, CA: CTB/McGraw–Hill.

Dale, P., Notari, A., Crain-Thoreson, C., & Cole, K. (1993). Parent–child storybook reading as an intervention technique for young children with language delays. *Society for Research in Child Development Abstracts, 9,* 310.

Dunn, L. M., & Dunn, L. M. (1981). *Peabody picture vocabulary test—Revised.* Circle Pines, MN: American Guidance Service.

Dunning, D., Mason, J., & Stewart, J. (1994). Half empty or half full? A response to Scarborough & Dobrich. *Developmental Review, 14,* 324–339.

Edgar, E., Jenkins, J. R., Hayes, M., Sampson, P., & Vadasy, P. (1992). *Head Start Research Project: Variation in service delivery models in Region X.* Seattle, WA: Washington Research Institute.

Gardner, M. F. (1981). *Expressive one-word picture vocabulary test.* Novato, CA: Academic Therapy Publications.

Gardner, M. F. (1990). *Expressive one-word picture vocabulary test—Revised.* Novato, CA: Academic Therapy Publications.

Gates, A., & MacGinitie, W. (1965). *Gates–MacGinitie Reading Tests.* New York: Teachers College Press.

Gough, P. B. (1991). The complexity of reading. In R. R. Hoffman & D. S. Palermo (Eds.), *Cognition and the symbolic processes* (pp. 141–149). Hillsdale, NJ: Erlbaum.

Harms, T., & Clifford, R. M. (1980). *Early childhood environment rating scale.* New York: Teachers College Press.

Huttenlocher, J., Haight, W., Bryk, A., Seltzer, M., & Lyons, T. (1991). Early vocabulary growth: Relation to language input and gender. *Developmental Psychology, 27,* 236–248.

Karweit, N. (1989). The effects of a story-reading program on the vocabulary and story comprehension skills of disadvantaged prekindergarten and kindergarten students. *Early Education and Development, 1,* 105–114.

Kenny, D. A. (1988). The analysis of data from two-person relationships. In S. Duck (Ed.), *Handbook of personal relationships: Theory, research and interventions* (pp. 57–78). New York: Wiley.

Kirk, S. A., McCarthy, J. J., & Kirk, W. D. (1968). *Illinois test of psycholinguistic abilities.* Urbana, IL: University of Illinois Press.

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

Levin, J. R., & Serlin, R. C. (1993, April). *No way to treat a classroom: Alternative units—Appropriate statistical strategies.*

Paper presented at the annual meeting of the American Educational Research Association, Atlanta.

Lonigan, C. J. (1993). Somebody read me a story: Evaluation of a shared reading program in low-income daycare. *Society for Research in Child Development Abstracts, 9,* 219.

Lonigan, C. J. (1994). Reading to preschoolers exposed: Is the emperor really naked? *Developmental Review, 14,* 303–323.

Lundberg, I., Frost, J., & Petersen, O. (1988). Effects of an extensive program for stimulating phonological awareness in preschool children. *Reading Research Quarterly, 23,* 263–284.

Mann, V. A. (1993). Phoneme awareness and future reading ability. *Journal of Learning Disabilities, 26,* 259–269.

Mason, J. M. (1993). Reading stories to preliterate children: A proposed connection to reading. In P. B. Gough, L. C. Ehri, & R. Treiman (Eds.), *Reading acquisition* (pp. 215–241). Hillsdale, NJ: Erlbaum.

McCormick, C. E., & Mason, J. M. (1986). Intervention procedures for increasing preschool children's interest in and knowledge about reading. In W. H. Teale & E. Sulzby (Eds.), *Emergent literacy: Writing and reading* (pp. 90–115). Norwood, NJ: Ablex.

McKey, R. H., Condelli, L., Ganson, H., Barrett, B. J., McConkey, C., & Plantz, M. C. (1985). *The impact of Head Start on children, families and communities. Final report of the Head Start evaluation, synthesis and utilization project* (DHHS Publication No. (OH25) 85–31193). Washington, DC: U.S. Government Printing Office.

Morisset, C. E. (1993). SPARK: Seattle's parents are reading to kids. *Society for Research in Child Development Abstracts, 9,* 219.

Morrison, F., McMahon, E. H., & Williamson, G. A. (1993). Two strikes from the start: Individual differences in early literacy. *Society for Research in Child Development Abstracts, 9,* 221.

National Academy of Education (1985). *Becoming a nation of readers.* Washington, DC: National Academy of Education.

Novak, G., Staggs, D., & Jones, S. (1993). Multilogic reading with low English proficiency students. *Society for Research in Child Development Abstracts, 9,* 522.

Olson, R. K., Wise, B., Conners, F., & Rack, J. P. (1990). Organization, heritability, and remediation of component word recognition and language skills in disabled readers. In T. H. Carr & B. A. Levy (Eds.), *Reading and its development: Component skills approaches* (pp. 261–322). San Diego, CA: Academic Press.

Payne, A. C., Whitehurst, G. J., & Angell, A. L. (in press). The role of home literacy environment in the development of language ability in preschool children from low-income families. *Early Childhood Research Quarterly.*

Raz, I. S., & Bryant, P. (1990). Social background, phonological awareness and children's reading. *British Journal of Developmental Psychology, 8,* 209–225.

Scarborough, H. S., & Dobrich, W. (1994). On the efficacy of reading to preschoolers. *Developmental Review, 14,* 245–302.

Stanovich, K. E., Cunningham, A. E., & Cramer, B. (1984). Assessing phonological awareness in kindergarten children: Issues of task comparability. *Journal of Experimental Child Psychology, 38,* 175–190.

Stanovich, K. E., & Siegel, L. S. (1994). Phenotypic performance profile of children with reading disabilities: A regression-based test of the phonological-core variable-difference model. *Journal of Educational Psychology, 86,* 24–53.

Stevenson, H. W., & Newman, R. S. (1986). Long-term prediction of achievement and attitudes in mathematics and reading. *Child Development, 57,* 646–659.

Stevenson, J., & Fredman, G. (1990). The social environmental correlates of reading ability. *Journal of Child Psychology and Psychiatry, 31,* 681–698.

U.S. Department of Health and Human Services. (1992). *Head Start research grantees meeting: Background material.* Washington, DC: Ellsworth Associates.

Valdez-Menchaca, M. C., & Whitehurst, G. J. (1992). Accelerating language development through picture book reading: A systematic extension to Mexican day-care. *Developmental Psychology, 28,* 1106–1114.

Wagner, R. K., Torgesen, J. K., Laughon, P., Simmons, K., & Rashotte, C. A. (1993). Development of young readers' phonological processing abilities. *Journal of Educational Psychology, 85,* 83–103.

Walker, D., Greenwood, C., Hart, B., & Carta, J. (1994). Prediction of school outcomes based on early language production and socioeconomic factors. *Child Development, 65,* 606–621.

Wells, G. (1985). *Language development in the preschool years.* Cambridge: Cambridge University Press.

Whitehurst, G. J. (1992a). *Dialogic reading in Head Start, K, and pre-K.* Stony Brook, NY: Author.

Whitehurst, G. J. (1992b). *The Stony Brook family reading survey.* Stony Brook, NY: Author.

Whitehurst, G. J., Arnold, D. S., Epstein, J. N., Angell, A. L., Smith, M., & Fischel, J. E. (1994). A picture book reading intervention in daycare and home for children from low-income families. *Developmental Psychology, 30,* 679–689.

Whitehurst, G. J., Epstein, J. N., Angell, A. L., Payne, A., Crone, D., & Fischel, J. E. (1993, November). *An emergent literacy curriculum for Head Start: First year outcomes.* Paper presented at the second National Head Start Research Conference, Washington, D.C.

Whitehurst, G. J., Falco, F. L., Lonigan, C., Fischel, J. E., DeBaryshe, B. D., Valdez-Menchaca, M. C., & Caulfield, M. (1988). Accelerating language development through picture book reading. *Developmental Psychology, 24,* 552–558.

Zajonc, R. B. (1976). Family configuration and intelligence. *Science, 192,* 227–236.

Zigler, E., & Styfco, S. J. (1993). Using research and theory to justify and inform Head Start expansion. *Social Policy Report: Society for Research in Child Development, 12,* 1–21.

Received February 2, 1994
Revision received April 19, 1994
Accepted May 18, 1994 ■