NCEE 2007-4007

U. S. DEPARTMENT OF EDUCATION

National Evaluation of Early Reading First

Final Report to Congress

National Evaluation of Early Reading First

Final Report

MAY 2007

Russell Jackson, Project Director
Ann McCoy
Carol Pistorino
Anna Wilkinson
Decision Information Resources, Inc.

John Burghardt, Principal Investigator
Melissa Clark
Christine Ross
Peter Schochet
Mathematica Policy Research

Paul Swank
Center for Improving the Readiness of Children for Learning and Education

Stefanie R. Schmidt, Project Officer
Institute of Education Sciences

NCEE 2007-4007
U.S. DEPARTMENT OF EDUCATION

U.S. Department of Education
Margaret Spellings
Secretary

Institute of Education Sciences
Grover J. Whitehurst
Director

National Center for Education Evaluation and Regional Assistance
Phoebe Cottingham
Commissioner

May 2007

This report was prepared for the Institute of Education Sciences under Contract No. ED-01-CO-0027/0002. The project officer was Stefanie R. Schmidt in the National Center for Education Evaluation and Regional Assistance. The National Institute for Literacy also contributed funds for this evaluation.

IES evaluation reports present objective information on the conditions of implementation and impacts of the programs being evaluated. IES evaluation reports do not include conclusions or recommendations or views with regard to actions policymakers or practitioners should take in light of the findings in the reports.

To order copies of this report,

- Write to ED Pubs, Education Publications Center, U.S. Department of Education, P.O. Box 1398, Jessup, MD 20794-1398.
- Call in your request toll free to 1-877-4ED-Pubs. If 877 service is not yet available in your area, call 800-872-5327 (800-USA-LEARN). Those who use a telecommunications device for the deaf (TDD) or a teletypewriter (TTY) should call 800-437-0833.
- Fax your request to 301-470-1244.
- Order online at www.edpubs.org.

This report also is available on the Department's website at http://www.ed.gov/ies.

Upon request, this report is available in alternate formats such as Braille, large print, audiotape, or computer diskette. For more information, please contact the Department's Alternate Format Center at 202-260-9895 or 202-205-8113.

Acknowledgments

This study represents a collaborative effort of ERF grantees, unfunded ERF applicants, preschools, teachers and other preschool staff, parents, researchers, data collection experts, and IES staff. We appreciate the time and effort of the many preschool staff in making their classrooms and students available and in responding to requests for data from the evaluation team.

The National Institute for Literacy provided crucial financial support for the data collection phase of this study.

A technical working group provided invaluable and critical guidance to us in designing the evaluation and analyzing the data. We thank Margaret Burchinal, Janet Currie, John Guthrie, and Mark Lipsey for their insights. Their participation immeasurably enhanced the quality of this effort. Important substantive expertise was also provided to our team by Chris Lonigan.

We also acknowledge and thank Audrey Pendleton, Ricky Takai, Tracy Rimdzius, and Phoebe Cottingham of IES for their interest, support, and thoughtful feedback throughout the project. In addition, Patricia McKee and her Early Reading First program staff in the Office of Elementary and Secondary Education assisted in ensuring that we obtained maximum cooperation from grantees throughout the evaluation.

The listed authors of this report represent only a small part of the team involved in this project across three organizations. At DIR, we thank Nancy Dawson, Pam Wells, Pia White, Irvin Williams, Sherri Provost, Cynthia Melgoza, Brenda Johnson, Nancy McArthur, Cynthia Lewis, Scott Peecksen, Carla Prince, Antwanette Hill, Kenneth Jackson, and other staff who supported the site recruitment, data collection, and analyses activities. Over 40 observers and assessors were critical to the success of the data collection and we thank them for their tireless efforts. We also thank Jamie Diamandopoulos for her skillful editing of the final report and Frenetta Tate for her production assistance.

At MPR, we thank Jennifer Baskwell (production support), Mark Dynarski (technical review), William Garret (production support), Jamila Henderson (programming support), and Steve Williams (technical review of sampling). At Center for Improving the Readiness of Children for Learning and Education (CIRCLE) at the University of Texas-Houston Health Sciences Center, we thank Mike Assel, Lauren Jenkins, April Crawford, Susan Landry, Wei Wu, Pornchai Direkwut, and Yimei Han for the significant contribution that they made to the success of this evaluation.

We thank Ronna Cook at Westat for providing unpublished tabulations of Cronbach's alpha for the pre-CTOPPP Elision subtest using the Head Start Impact Study data, and Tracy Rimdzius and Caroline Ebanks of IES for sharing similar unpublished tabulations using data from the Even Start Classroom Literacy Interventions and Outcomes (CLIO) Study and Preschool Curriculum Evaluation Research (PCER) Study, respectively.

Disclosure of Potential Conflicts of Interest[1]

The research team for this evaluation consists of a prime contractor, Decision Information Resources (DIR), Inc., and two subcontractors, Mathematica Policy Research (MPR), Inc. and the Center for Improving the Readiness of Children for Learning and Education (CIRCLE) at the University of Texas-Houston Health Sciences Center. DIR and MPR have no interests that could be affected by findings from the evaluation of the Early Reading First (ERF) program.

CIRCLE developed one of the study's classroom observation measures and advised on the selection of the child assessments. CIRCLE also trained DIR staff to collect classroom observation and child assessment data, but CIRCLE staff did not collect any data. They scored the classroom observations and child assessments by entering item-level data into computer programs, but they did not know the treatment or control status of the classrooms for which they entered data. CIRCLE carried out descriptive analyses of the classroom observation data that was incorporated into the report, but had no role in the impact analyses in the report. Under a separate contract with an ERF grantee that was part of the 2003 cohort in the study, CIRCLE provided services to and conducted the grantee's local evaluation as required under its grant. Under another separate contract with an ERF grantee that was part of the 2005 cohort and not in the study, CIRCLE provided technical assistance on the use of its classroom observation measure. In addition, CIRCLE has adopted a public position supporting early childhood classroom activities and instructional materials and seeks funding to provide services that are consistent with the goals of the Early Reading First program.

A consultant to DIR, Professor Christopher Lonigan of Florida State University, provided an assessment that was used in a battery of assessments for the evaluation. Dr. Lonigan's role in the DIR project was to review and provide feedback on the preliminary results of the study and to provide information on the psychometric properties of the assessment he developed. Dr. Lonigan had previously developed the assessment and he had no role in the selection of assessments. The assessment was not commercially available at the time it was selected by DIR and during the data collection phase. A revised version of the assessment became commercially available as the Test of Preschool Early Literacy (TOPEL) in January 2007, after the ERF data collection. Dr. Lonigan has a financial interest in the commercial version. Under a separate contract, Dr. Lonigan was commissioned by the Early Reading First Program Office to provide lectures on components of effective instruction in phonemic awareness at meetings of Early Reading First grantees in April 2006 and April 2007 after the preliminary descriptive findings of the report were shared with the grantees. Dr. Lonigan was also commissioned by the Early Reading First Program Office to provide a lecture on the preliminary findings of the National Early Literacy Panel in October 2004.

[1] Contractors carrying out research and evaluation projects for IES frequently need to obtain expert advice and technical assistance from individuals and entities whose other professional work may not be entirely independent of or separable from the particular tasks they are carrying out for the IES contractor. Contractors endeavor not to put such individuals or entities in positions in which they could bias the analysis and reporting of results, and their potential conflicts of interest are disclosed.

Contents

Tables

Table **Page**

Figures

Figures **Page**

Exhibits

Exhibits **Page**

Executive Summary

The No Child Left Behind (NCLB) Act of 2001 created the Early Reading First (ERF) program to enhance teacher practices, instructional content, and classroom environments in preschools and to help ensure that young children start school with the skills needed for academic success. This discretionary grant program provides funding to preschools that particularly serve children from low-income families so that the preschools can support age-appropriate development of children's language and literacy skills. The program, which was authorized under Title I, Part B, Subpart 2 of the Elementary and Secondary Education Act (ESEA) as reauthorized by NCLB, reflects the research of the last several years about the kinds of skills that young children must have to become successful readers. These skills include oral language (expressive and receptive language and vocabulary development), phonological awareness (rhyming, blending, segmenting), awareness of the print conventions, and alphabet knowledge (letter recognition) (Whitehurst and Lonigan 2001; Pullen and Justice 2003).

The NCLB Act also mandated an independent national evaluation of the ERF program and required a final report to Congress. This final report presents the impacts of the program on the language and literacy skills of children and on the instructional content and practices in preschool classrooms.

The main findings of the national evaluation of ERF are that the program had positive, statistically significant impacts on several classroom and teacher outcomes and on one of four child outcomes measured. Specifically, ERF had positive impacts on

- the number of hours of professional development that teachers received and on the use of mentoring as a mode of training
- aspects of classroom environments and teacher practices that were major focuses of the ERF program, including
    - language environment of the classroom
    - book-reading practices
    - the variety of phonological-awareness activities and children's engagement in them
    - materials and teaching practices to support print and letter knowledge and writing
    - the extensiveness and recency of child-assessment practices
- other, more general aspects of classroom quality, including the quality of teacher-child interactions, the organization of the classroom, and the planning of activities for children.

With regard to child outcomes, ERF had a positive impact on children's print and letter knowledge but not on phonological awareness or oral language.

ERF neither enhanced nor diminished children's social-emotional development during the preschool year. Patterns of results that were observed for the overall sample were also observed for most subgroups examined.

Study Background

Preventing Reading Difficulties in Young Children (National Research Council 1998) shows that a high percentage of children from low-income families attend preschools that may successfully address other developmental domains but often fail to provide the language, cognitive, and early-reading instruction and activities necessary to develop skills to become successful readers. Improving the instructional program to support the age-appropriate development of these skills is the central focus of ERF.

ERF provides grants to school districts, other public, nonprofit, and private organizations, and collaborations of the same entities that serve 3- to 5-year-olds, especially those from low-income families. The grants must be used to provide services that will better prepare children to enter kindergarten with the necessary language, cognitive, and literacy skills that can avert reading difficulties. ERF grants are intended to support the following items:

- A high-quality oral language and print-rich classroom environment
- Activities and instructional materials developed according to scientifically based reading research that will help develop children's oral language, phonological awareness, print awareness, and alphabet knowledge
- Screening and assessments to monitor children's acquisition of skills and to guide instruction
- Professional development formulated according to scientifically based reading research that will help teachers to enhance children's language, cognitive, and early literacy skills
- Integration of the instructional materials, activities, tools, and measures into the grantee's existing programs

Two key elements of ERF are the use of scientifically based methods and the goal of enhanced professional development. Scientifically based reading research is defined as that which applies rigorous, systematic, and objective procedures to obtain valid and reliable knowledge relevant to reading development, reading instruction, and reading difficulties. Consistent with the statutory definition of "professional development," ERF professional development was expected to be continuous, intensive, and classroom focused.

Five rounds of ERF grants have been awarded since the program began in 2002. These awards ranged from $750,000 to $4.5 million per site for a 3-year period. The national evaluation of ERF focused on the second cohort of grantees from FY 2003, in which the grants totaled approximately $75 million; the average award was $2.5 million, and individual awards ranged from $1,074,846 to $4,358,750 to be spent over three years.

The national evaluation of ERF was intended to investigate the effects on children's language development and emergent literacy when:

- preschools receive funding to adopt scientifically based methods and materials

- teachers are provided with focused professional development that supports the use of these materials and methods

The following research questions were addressed by the evaluation:

- What is the impact of ERF on the language and literacy skills of children enrolled in preschools that receive ERF support?

- What is the impact of ERF on the quality of language and literacy instruction, practice, and materials that preschools provide?

- To what extent are variations in ERF program quality and implementation associated with differences in the language and literacy skills of the children served?

Study Design

The study uses a regression-discontinuity (RD) design to assess the impact of ERF funding and program support for preschools on the language and literacy preparedness of preschool children. This study design takes advantage of the fact that the U.S. Department of Education (ED) is required to follow a formal, structured process for selecting grantees to receive ERF funding. In its published announcement of the availability of ERF grants for FY 2003 (*Federal Register* of March 11, 2003), ED established criteria for scoring each application received. Independent reviewers used these criteria to review and score applications. ED then awarded ERF grants to the grant applicants with the highest application scores, progressing down the score distribution until all funding available for the fiscal year had been allocated. In this way, 30 grants were awarded to the grant applicants with scores of at least 74; applicants with scores below 74 were not awarded grants. Impact estimates were obtained by comparing child outcomes and teacher practices in funded sites to those in unfunded sites, controlling for a smooth function of the application score.

The final evaluation sample was composed of a treatment group, which consisted of 4-year-olds attending preschool in 28 of 30 ERF grantee sites, whereas the comparison group consisted of children attending preschool in 37 of the 67 unfunded applicant sites that had the highest application scores and that agreed to participate in the study. Approximately three classrooms were selected from each participating site with probabilities proportional to the number of eligible students in each class (see Table 1). The study team randomly selected approximately 11 4-year-old students per classroom whose parents had provided written consent for participation in the study.

Table 1. Sample sizes for National Evaluation of ERF

Unit of Analysis	Funded sample size	Unfunded sample size	Total
ERF grantees/unfunded applicants	28	37	65
Preschools	86	75	161
Classrooms observed	78	91	169
Teachers surveyed	92	102	205
Children assessed	803	855	1,658

The study team collected data for the evaluation from several sources. Trained staff directly assessed the language and literacy skills of children participating in the study. Trained observers measured classroom practice in a subsample of study classrooms. The teachers of all children in the sample and the director or principal of each preschool participating in the study completed a self-administered questionnaire. Teachers of the sampled children were also asked to rate each child's social-emotional behavior. The study team also obtained data from the preschools about children's school attendance for the 2004–2005 year. Finally, parents of the sampled children were interviewed by telephone.

Data were collected at two times: fall 2004 and spring 2005. The same data-collection instruments and procedures were used in the funded and unfunded sites.

Child Assessments. Table 2 shows the instruments that were used to measure children's language and literacy skills in three domains (print and letter knowledge, phonological awareness, and oral language) and their social-emotional behavior.

Table 2. Data-collection instruments: child assessments

Instrument name	Domain measured	Psychometric information from published sources
(Pre-LAS)[1]	English proficiency screening	Internal consistency reliability = .86–.90
Preschool Comprehensive Test of Phonological and Print Processing (Pre-CTOPPP)[2]	Print and letter knowledge	Test of Preschool Early Literacy (TOPEL): • Internal consistency reliability = .95 • Test-retest reliability = .89
	Elision[3]	Internal consistency reliability = .71–.88
Expressive One-Word Picture Vocabulary Test (EOWPVT) [4]	Expressive vocabulary	• Internal consistency reliability coefficients = .96–.98 • Test-retest reliability = .95
Preschool Language Scale (PLS-4)[5]	Auditory comprehension	• Test-retest reliability = .83–.91 • Internal consistency reliability coefficients = .83–.90
Social Competence & Behavior Evaluation (30-item)—Teacher Rating[6]	• Social competence • Anger-aggression • Anxiety-withdrawal	Internal consistency reliability coefficients = .85–.92

[1] Duncan, S. E., and DeAvila, E. A. (1998). *Pre-LAS 2000*. Monterey, CA: CTB/McGraw-Hill.
[2] Lonigan, C., Wagner, R., Torgesen, J., and Rashotte, C. (2007). *The Test of Preschool Early Literacy (TOPEL).* Austin, TX: PRO-ED.
[3] Internal-consistency reliability coefficients of Elision subtest from unpublished tabulations using data from the Head Start Impact Study (U.S. Department of Health and Human Services 2005), and the forthcoming Even Start Classroom Observations and Interventions and Preschool Curriculum Evaluation Research studies, both being conducted by IES.
[4] Brownell, R. (2000). *Expressive One-Word Picture Vocabulary Test Manual*. Novato, CA: Academic Therapy Publications.
[5] Zimmerman, I. L., Steiner, V.G., and Pond, R.E. (2002). *Preschool Language Scale-4th Edition, Examiner's Manual*. San Antonio, TX: The Psychological Corporation.
[6] La Freniere, P. J., and Dumas, J. E. (1996). "Social competence and behavior evaluation in children ages 3 to 6 years: The short form (SCBE-30)," *Psychological Assessment, 8,* 369–377.

Classroom observations and surveys. Classroom practice and overall quality of the preschool classrooms were measured by two observation instruments—the Teacher Behavior Rating Scale (TBRS)[2] and 11 items from the Early Childhood Environment Rating Scale-Revised (ECERS-R) that form the Teaching and Interactions Subscale.[3] Trained members of the study team conducted the classroom observations.

[2] Landry et al. (2004). "Teacher Behavior Rating Scale (TBRS)," unpublished research instrument.
[3] Harms, T., Clifford, R.M., and Cryer, D. (1998). *Early Childhood Environment Rating Scale: Revised Edition*. NY: Teachers College Press, and Clifford, R.M., Barbarin, O., Chang, F., Early, D., Bryant, D., Howes, C., Burchinal, M., and Pianta, R. (2005). "What Is Pre-Kindergarten? Characteristics of Public Pre-Kindergarten Programs." *Applied Developmental Science*, vol. 9, no. 3, pp. 126–143.

The evaluation team also developed self-administered surveys that the teachers and preschool principals or directors completed in the fall of 2004 and spring 2005. Parents of children in the study were interviewed through computer-assisted telephone interviewing. The team conducted in-depth telephone interviews with grantee directors for each of the 28 funded grantees to learn about their use of ERF funds, including challenges encountered and notable successes.

Impact estimation and hypothesis testing. Impact estimates were obtained by comparing child outcomes and teacher practices in funded sites to those in unfunded sites, controlling for a smooth function of the application score. If the application score fully reflects the selection rule used to award ERF grants and we control for the correct function of the score, this approach produces unbiased estimates of the effect of ERF.

We adopted a 2-tailed hypothesis test because it was unclear before the evaluation whether ERF funding would improve all outcomes. For each outcome, the findings indicate the statistical significance of the impact estimates at the 5-percent level. The analysis methods accounted for the fact that some outcome domains contained multiple measures. The tables presented include checkmarks for domains in which impacts are jointly statistically significant once the adjustment for multiple comparisons is made. The tables also include p-values for tests of statistical significance of individual outcomes that do not reflect adjustments for multiple comparisons. The conclusions are unaffected when adjustments for multiple comparisons are applied.

The following sections contain findings about

- characteristics of ERF children and preschools
- ERF impacts on teachers and classroom practices
- ERF impacts on children's language and literacy skills and social-emotional outcomes

The evaluation also estimated ERF impacts for several subgroups defined by key characteristics of children, preschools, and teachers.

Characteristics of ERF Children and Preschools

Characteristics of children. ERF participants appeared to be more disadvantaged than the national average. A relatively large proportion of children served by ERF grantees had some characteristics associated with disadvantage. More than one-third of the ERF sample reported monthly income of less than $1,500, compared to 17 percent of households with 3- to 5-year-olds nationally. Children in this cohort were also more likely than children nationally to come from single-parent households (40 percent compared to 28 percent), be Hispanic (46 percent compared to 21 percent), and have foreign-born parents (39 percent compared to 23 percent). About 4 out of 10 ERF parents (41 percent) reported that the primary language spoken in the home was something other than English. Initial scores on three standardized assessments suggest that children were functioning below national norms (which were standardized to be 100 on all three tests) when they entered the ERF program. ERF participants scored an average of 94 on test of print and letter knowledge, 91 on a test of auditory comprehension (an oral language measure), and 83 on a test of expressive vocabulary (another oral language measure).

Characteristics of preschools. The vast majority of ERF preschools (95 percent) combined ERF funding with other government funding sources, which was consistent with the goal of the program to enhance the quality of existing programs that particularly serve children from low-income families. The most common funding sources were state and local education agencies, state child-care funds, and Head Start, which were received by 56 percent, 38 percent, and 36 percent of ERF preschools, respectively. Just over half of ERF preschools received funding from only one of these sources, while over 40 percent received funding from two or more sources. The schedule on which ERF preschools operate and the characteristics of their teachers provide useful context for examining study findings. Three-quarters are full-day programs (operating for an average of 8 hours per day), 62 percent have a class size of 20 children or fewer, and almost 70 percent have a staff-to-child ratio of 1:10 or better. Seventy-five percent of the ERF teachers have bachelor's degrees, 67 percent have teaching certificates or licenses. Among teachers in ERF classrooms, 87 percent had completed college-level courses in early-childhood education or development, 67 percent had completed courses in teaching reading to elementary-school children, and 79 percent had completed courses in teaching language and literacy skills to children in a preschool setting.

ERF funding in the preschools. Based on the reported number of preschool children expected to be served by the FY 2003 grantees, the median ERF allocation across the 28 grantees evaluated in the FY 2003 cohort was $3,549 per preschool child per year.[4] These funds are in addition to the other government funding sources received by the preschools. To provide perspective, annual average Head Start funding per child in Fiscal Year 2003 was $7,092.[5]

Professional development through ERF. ERF teachers reported receiving an average of 72 hours of professional development during the previous year—the equivalent of 9 days. One hundred percent of teachers in ERF-funded classrooms reported receiving professional development in phonemic and phonological awareness (see Table 3). The vast majority of ERF teachers received training in six other language-development and early literacy topics, including literacy-rich print environments (97.8 percent), concepts of print writing and prewriting (96.7 percent), oral language (96.7 percent), facilitating emergent literacy (95.7 percent), alphabetic knowledge (92.4 percent), and oral comprehension and cognition (88.0 percent). Nine out of 10 ERF teachers reported receiving training in child assessment. Three-fourths of ERF teachers reported receiving training in traditional early-childhood topics, including children's development and ways to manage children's behavior in the classroom.

[4] The methodology used to compute the ERF allocation per child is described in Appendix B, "Data Collection Methods."

[5] U.S. Department of Health and Human Services (April 2004), *Head Start Program Fact Sheet Fiscal Year 2003,* Administration for Children and Families. http://www.acf.hhs.gov/programs/hsb/research/2004.htm.

Table 3. Topics in which ERF teachers received professional development in the past 12 months

Topic Areas	% ERF teachers who received training in topic
Language Development and Early Literacy	
Phonemic & phonological awareness	100.0
Literacy-rich environments	97.8
Concepts of print writing & prewriting	96.7
Oral language	96.7
Facilitating emergent literacy	95.7
Alphabetic knowledge	92.4
Oral comprehension & cognition	88.0
Child Assessment	
Child Development and Behavior	90.2
Early childhood growth & development	76.1
Classroom management	76.1
Other Topics	56.5
Number of Topics	**% ERF teachers who received training in number of topics**
0	0.0
1 to 4	1.1
5 to 8	21.7
9 or 10	77.2
Mean # of topics (SD)	9.6 (1.7)
Sample Size	92

SOURCE: Spring teacher surveys.

Curriculum and assessment. The statute requires ERF grantees to identify and provide activities and instructional materials that are designed according to scientifically based reading research for developing children's oral language, phonological awareness, print awareness, and alphabet knowledge.[6] ERF programs are also expected to integrate assessments of child progress with teaching so that instruction can build on what children already know and bring them to the next level (U.S. Department of Education 2003).

In ERF preschool classrooms, 39 percent of the teachers reported following one curriculum, and 61 percent reported using a combination of curricula. The most commonly reported curricula in ERF classrooms are Creative Curriculum (reported by 46 percent of teachers) and High/Scope (Educating Young Children) curriculum (reported by 24 percent of teachers).

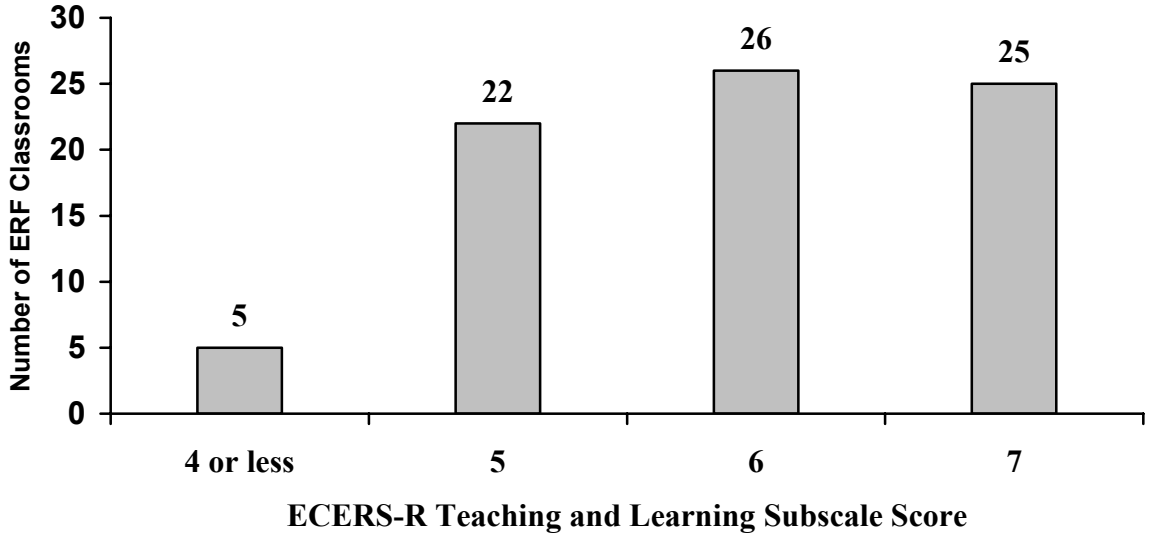
Nearly all ERF teachers (98 percent) reported using at least one assessment tool for children in their classes. A majority of ERF teachers (64 percent) reported using more than one assessment instrument with children in their classes.

Classroom environments and teacher practices. The Early Childhood Environment Rating Scale-Revised (ECERS-R) provided a measure of the general quality of the preschool

[6] U.S. Department of Education. *Guidance for the Early Reading First Program*. Washington, DC, March 2003, p. 5.

environment. The quality of teacher-child interactions refers to the teacher's responsiveness to children; sensitivity to children's needs; consistent, positive guidance; and encouragement. As one measure of teacher-child interactions, we used the Teaching and Interactions subscale of the ECERS-R (Clifford et al. 2005). The average score on the ECERS-R Teaching and Interactions subscale in the spring was 5.8 for ERF classrooms (slightly higher than 5.7 average score in the fall), with all but 5 classrooms scoring at least a "good" or 5 on the subscale (see Figure 1).[7]

Figure 1. Number of ERF classrooms by ECERS-R Teaching and Interactions Subscale, spring 2005



The TBRS measures the general quality of preschool classrooms (including teacher sensitivity) as well the language and early literacy aspects of teacher instructional practices and the available classroom materials. The TBRS items are scaled so that higher values represent greater frequency or quality or both, using Likert ratings that range from 1 (low or none) to 4 (high frequency/high quality) for virtually all of the items. Because of a high correlation between quantity and quality item scores, we have averaged them to create a single-item score and created subscales from these composite items.[8]

[7] Scores on the Teaching and Interactions subscale tend to be higher than scores on the full ECERS-R scale. In a sample of Head Start classrooms, the ECERS-R score was 4.9, and the Teaching and Interactions subscale score was 5.5.

[8] Appendix C contains additional information about the TBRS subscales used in the ERF evaluation.

The total TBRS score summarizes all of the TBRS general quality and language, literacy, and assessment subscales. The subscales measured

- oral-language use
- book-reading practices
- phonological-awareness activity
- print and letter knowledge
- written expression
- portfolios
- dynamic assessment

The average TBRS total score was 2.7 for ERF classrooms in the fall and 2.6 in the spring.

ERF Impacts on Teachers and Classroom Practices

In assessing the impact of ERF on teachers and classroom practices, we examined the following outcomes:

- teacher knowledge and skills
- the general quality of the preschool environment
- the quality of language, early literacy, and child-assessment practices and environments

Within each of these outcome areas, we examined measures for several domains. We also examined impacts on selected subgroups of teachers and classrooms.

Teacher knowledge and skills. We expected that ERF preschools would enhance teachers' knowledge and skills through professional development. Overall, we find that ERF had positive impacts on the hours of teachers' professional development during the 12 months preceding the spring 2005 survey and that it increased the proportion of teachers receiving professional development through mentoring.

- ERF increased the number of hours of professional development that focused on language and early literacy topics by 48 hours (6 days) over the 12 months preceding the survey.

- A higher proportion of ERF teachers than teachers in unfunded programs reported receiving professional development on language or literacy topics and on curriculum topics through mentoring or tutoring. The program's impact on the proportion of teachers receiving mentoring or tutoring on language and literacy topics was 41 percentage points.

- A larger proportion of ERF teachers than teachers in unfunded programs reported receiving workshop training on language and literacy topics. The program's impact on the proportion of teachers receiving workshop training on language and literacy topics was 41 percentage points.

ERF did not induce centers to raise the wages of their teachers who had received additional professional development through the program.

General quality of the preschool environment. This study examines teacher behaviors and environmental factors that relate to the general quality of the preschool classroom environment. We selected general quality measures, including teacher behaviors and classroom environment, that previous research has found to be positively correlated with young children's cognitive skills and emotional development (Vandell and Wolfe 2000; NICHD Early Childhood Research Network 2002, 2003, and 2006). However, given its correlational nature, this research is not conclusive. Further, the study examines the measures of teacher instructional practices and classroom environment that are closely related to ERF's emphasis on language and emerging literacy skills.

In the spring, ERF had *pervasive* impacts on the general quality of the preschool classroom—the classroom language environment, materials, and teaching practices that support early literacy, and child-assessment practices. In particular, ERF

- Increased the lead teachers' sensitivity and the quality of interactions toward children by approximately one standard deviation relative to what we would have expected in the absence of the program.
- Improved the quality of the assistant teachers' interactions with children by 0.79 standard deviations.
- Had positive impacts on measures of the organization of the classroom environment; effect sizes exceed one standard deviation.
- Significantly improved lesson planning.
- Increased the overall quality of the classroom-learning environment, measured by the total TBRS score (the average across subscales measuring general classroom quality and the language and early literacy environment).
- Increased the general quality of teacher-child interactions as measured by the ECERS-R teaching and learning subscale.

Quality of language, early literacy, and child-assessment practices and environments. In the spring, ERF had impacts on all domains of classroom language, early literacy, and assessment practices. Specifically

- Oral language use by both the lead and assistant teachers
- Book-reading practices that include introducing new vocabulary, using expressive voice, and asking open-ended questions during the book-reading session
- Phonological awareness activities that promote knowledge of letter and word sounds
- Print and letter knowledge materials and activities to promote letter recognition and the association between sounds and letters

- Written expression and early writing activities
- Child screening and progress assessments on a regular basis to plan instruction

ERF Impacts on Children's Language and Literacy Skills and Social-Emotional Outcomes

Ultimately, through its effects on classroom practices, the ERF Program is intended to provide young children with the necessary language, cognitive and early-reading skills to prevent reading difficulties and ensure school success as they enter kindergarten. We obtained the outcome measures for the child analyses from assessments that were given to children in spring of the school year on their literacy and language skills and behavior. The assessments measured *print and letter knowledge, phonological awareness, and oral language.* We also estimated ERF's impacts on children's social-emotional development.

Impact findings. Overall, we find that ERF had a statistically significant positive effect on children's print and letter knowledge but no statistically discernable impact on phonological awareness or oral language. We find no evidence of negative impacts on children's social-emotional skills. Specifically:

- ERF increased children's standard scores on Pre-CTOPPP print awareness by 5.78 points relative to what we would have expected in the absence of the program. This increase indicates that ERF improved children's ability to recognize letters of the alphabet and associate letters with their sounds. The impact estimate translates into an effect size of 0.34 standard deviations. Comparison of the regression-adjusted standard scores for children in the unfunded sites to the national norms for this subtest indicates that in the absence of ERF, children in the ERF sites would have scored about 3 percentage points below the national average of 100; with exposure to ERF, their average score of 102.69 was slightly above the national average for this subtest.
- We find no evidence that ERF improved children's phonological awareness.
- We find no evidence that ERF improved children's oral language skills.
- ERF did not affect children's social-emotional skills, as measured by the SCBE-30 anger-aggression, social-competence, and anxiety-withdrawal scales. The lack of program effects in this domain is noteworthy in light of concerns that ERF might *adversely* impact these skills by compelling teachers to focus on improving language and literacy at the expense of developing other skills.

Analysis of Mediators of ERF's Impacts on Classroom Instructional Practice and Children's Language and Literacy Skills

As a final part of the analysis of ERF, we explored potential channels, or mediators, through which ERF generated its positive impacts on classroom and child outcomes. Unlike the impact analyses, this analysis is correlational, rather than quasi-experimental, because we could not use the regression-discontinuity design to identify the causal effects of particular mediators.

Consequently, any observed effect of mediators on child or classroom outcomes might be due to the effects of unobserved factors that happen to be correlated with these mediators, rather than to the mediators themselves.

For our analysis of the channels through which ERF generated positive impacts on classroom and child outcomes, we hypothesized that the additional hours of professional development attributable to ERF and the increased proportion of teachers receiving professional development through intensive, individualized mentoring account for at least some of ERF's impact on the classroom language and early literacy environment. The impacts on classroom environments, in turn, might account for at least some of the program's impacts on children's language and literacy skills.

To investigate this hypothesis, we first examined the extent to which hours of professional development and the use of mentoring as a mode of training were associated with the classroom outcomes affected by ERF. We then examined the associations between classroom outcomes and the child outcome on which ERF had a positive impact—print and letter knowledge. Thus, our model of print awareness includes as mediators the number of phonological awareness activities, print- and letter-knowledge learning opportunities, written-expression learning opportunities, the classroom print environment, opportunities and materials for writing, book-reading practices, child portfolios, and teacher sensitivity.

The estimated marginal effect of hours of professional development is generally small and not statistically significant on each of the 10 measures with the exceptions of classroom print environment and teacher sensitivity; we estimated positive and statistically significant effects of professional development on those two measures. Similarly, the estimated marginal effect of mentoring on each of the 10 outcomes is generally small and not statistically significant, with the exceptions of child portfolios and teacher sensitivity; the estimated marginal effects of mentoring are negative and statistically significant on those two outcomes. The mediators are jointly statistically significant only for child portfolios and teacher sensitivity.

The estimated marginal effects on print and letter knowledge are not statistically significant for any of the potential mediators except print and letter-knowledge learning opportunities, which account for 27 percent of the total implied impact on print-awareness scores. Together, all eight mediators account for 60 percent of the total implied impact on print and letter knowledge and are jointly statistically significant at the 5-percent level.

Chapter 1. Introduction and Study Background

The No Child Left Behind (NCLB) Act of 2001 created the Early Reading First (ERF) program to enhance teacher practices, instructional content, and classroom environments in preschools and help ensure that young children start school with the skills needed for academic success. This discretionary grant program provides funding to preschools that particularly serve children from low-income families so that the preschools can support age-appropriate development of children's language and literacy skills. The program, which was authorized under Title I, Part B, Subpart 2 of the Elementary and Secondary Education Act (ESEA) as reauthorized by NCLB, reflects the research of the last several years about the kinds of skills that young children must have to become successful readers. These skills include oral language (expressive and receptive language and vocabulary development), phonological awareness (rhyming, blending, segmenting), awareness of print conventions, and alphabet knowledge (letter recognition) (Whitehurst and Lonigan 2001; Pullen and Justice 2003).

The NCLB Act also mandated an independent national evaluation of the ERF program and required a final report to Congress. This final report presents the impacts of the program on the language and literacy skills of children and on the instructional content in preschool classrooms.

Rationale and Goals of ERF

Preventing Reading Difficulties in Young Children (National Research Council 1998) shows that a high percentage of children from low-income families attend preschools that may successfully address other developmental domains but often fail to provide the language, cognitive, and early-reading instruction and activities necessary to develop skills to become successful readers. Improving the instructional program to support the age-appropriate development of these skills is the central focus of ERF.

ERF provides grants to school districts, other public, nonprofit, and private organizations, and collaborations of the same entities that serve 3- to 5-year-olds, especially those from low-income families. The grants must be used to provide services that will better prepare children to enter kindergarten with the necessary language, cognitive, and literacy skills that can avert reading difficulties.

ERF grants are intended to support the following items:

- A high-quality oral language and print-rich classroom environment
- Activities and instructional materials developed according to scientifically based reading research that will help develop children's oral language, phonological awareness, print awareness, and alphabet knowledge
- Screening and assessments to monitor children's acquisition of skills and to guide instruction

- Professional development developed according to scientifically based reading research that will help teachers to enhance children's language, cognitive, and early literacy skills

- Integration of the instructional materials, activities, tools, and measures into the grantee's existing programs

Grantees were also encouraged to use funds to support parent engagement and to promote continuity in the transition to kindergarten and elementary school. Two key elements of ERF are the use of scientifically based methods and the goal of enhanced professional development.

Focus on Scientifically Based Methods

The statute (sections 1221(b)(2) and 1208(6), ESEA) defines scientifically based reading research as that which applies rigorous, systematic, and objective procedures to obtain valid and reliable knowledge relevant to reading development, reading instruction, and reading difficulties. Specifically, this research:

- Employs systematic, empirical methods that draw on observation or experiment

- Involves rigorous data analyses that are adequate to test the stated hypotheses and justify the general conclusions drawn

- Relies on measurements or observational methods that provide valid data across evaluators and observers and across multiple measurements and observations

- Has been accepted by a peer-reviewed journal or approved by a panel of independent experts through a comparably rigorous, objective, and scientific review

Using scientifically based reading research, as defined by the statute, to develop curricula and design instruction intended to enhance the oral language, phonological awareness, print awareness, and alphabetic knowledge skills of preschool-age children—particularly those from low-income families—through planned interventions is a relatively new phenomenon. Although research has identified skills that children need in order to become proficient readers, research regarding how to refine and design instructional approaches and activities that will improve the reading outcomes of children is ongoing (Whitehurst and Lonigan 2001; Pullen and Justice 2003). The national evaluation of ERF is intended to

- investigate the effects on children's language development and emergent literacy when preschools and teachers are encouraged to adopt scientifically based methods and materials

- provide evidence of the effects on preschools and teachers of focused professional development that supports the use of these materials and methods

Focus on Professional Development

Professional development and training of teachers is envisioned as a key vehicle for implementing the desired objectives of ERF. The statute requires that the professional development be grounded in scientifically based reading research and knowledge of early language and literacy development. Consistent with the statutory definition of "professional development," ERF professional development was expected to be continuous, intensive, and classroom focused. Professional development that included mentoring and coaching was encouraged.

Funding Levels and the Application Process

Five rounds of ERF grants have been awarded since the program began in 2002. These awards ranged from $750,000 to $4.5 million per site for a 3-year period. From FY 2002 through FY 2006, the average ERF award increased from $2.5 million to $3 million. The national evaluation of ERF focused on the second cohort of grantees from FY 2003. For the 2003 cohort, the grants totaled approximately $75 million with an average award of $2.5 million. Individual awards ranged from $1,074,846 to $4,358,750 to be spent over three years.

For FY 2003, the ERF grant competition was conducted through a 2-stage process. First, applicants were invited to submit brief pre-applications. Second, the highest quality pre-applicants were invited to submit full applications. A peer review panel of experts was convened to evaluate and score each pre-application on the basis of specific selection criteria. For FY 2003, ED received approximately 700 ERF pre-applications, and the 125 highest scoring pre-applicants were asked to submit full applications.

ED received full applications from 124 of the 125 pre-applicants that were invited to submit full applications. Each full application was required to include a brief description of the project's context, a narrative addressing the selection criteria (different than the pre-application selection criteria), a budget, and a budget narrative. A separate peer review panel of experts was convened to evaluate and score the full applications on the basis of the selection criteria.[9]

Through the use of two invitational priorities, ED expressed particular interest in (a) applicants that were partnerships between at least state education agencies or local education agencies and preschools not under administrative control of local education agencies, and (b) applicants serving significant numbers of children with special needs, including those with disabilities and limited English proficiency. Applicants that met the invitational priorities did not automatically receive extra points. However, because of ED's interest in invitational priorities, the composition of the 2003 cohorts may have differed from other cohorts. In particular, the 2003 cohort had more grantees and applicants that formed collaborations of different kinds of preschools not under the same administrative umbrella in their community (for example, collaborations of Head Start programs, preschools administered by school districts, and independent child-care centers).

[9] The full application selection criteria included the capacity and significance of the project, the quality of project activities and services, the quality of project personnel, the quality of the management plan, and the quality of project evaluation.

In October 2003, ED made 3-year grants to the 30 highest scoring applicants. Implementation of the ERF activities was expected to begin by January 2004.

The Evaluation

This section describes the congressional mandate and the research questions.

Congressional Mandate

Section 1226 of the legislation authorizing ERF (Title I, Part B, Subpart 2 of the ESEA as reauthorized by the NCLB) includes a set-aside for an independent evaluation of the effectiveness of ERF. According to the legislative requirements, the evaluation reports submitted to Congress must include information about the following items:

- Ways in which the grant recipients are improving the prereading skills of preschool children
- The effectiveness of the professional development program implemented through these grants
- How early childhood teachers are being prepared with scientifically based reading research about early-reading development
- What activities and instructional practices are most effective
- How prereading instructional materials and literacy activities based on scientifically based reading research are being integrated into preschools, child-care agencies and programs, programs carried out under the Head Start Act, and family literacy programs
- Any recommendations about strengthening or modifying this program

This national evaluation report responds to those legislative requirements.

Research Questions

In line with the legislative direction, the national evaluation of ERF addressed the following questions:

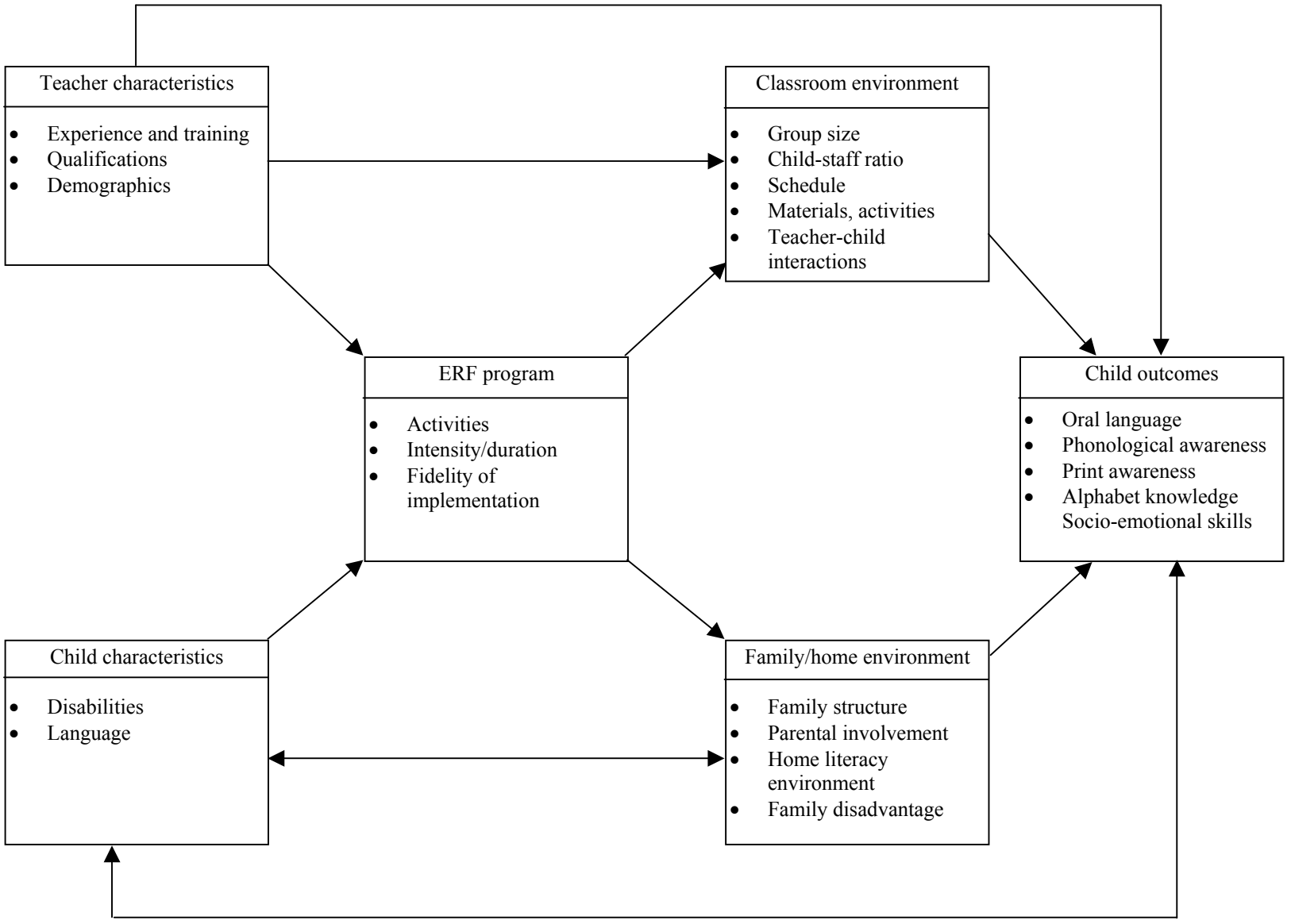
- What is the impact of ERF on the language and literacy skills of children enrolled in preschools that receive ERF support?
- What is the impact of ERF on the quality of language and literacy instruction, practice, and materials that preschools provide?
- To what extent are variations in ERF program quality and implementation associated with differences in the language and literacy skills of the children served?

The conceptual model that informs the research design for this evaluation is depicted in Figure 1.1. The ERF intervention is expected to directly influence teachers' experience and qualifications through professional development and to influence the classroom environment through the materials and activities in the classroom and through teacher-child interactions. As shown in the conceptual model, the quality of teachers' instructional practice and the classroom environment are viewed as central factors in determining the impact of ERF on children's literacy and language outcomes. Another central factor is the relation between ERF participation and children's social-emotional outcomes.

The study uses a regression discontinuity (RD) design to examine the extent to which additional funds and technical assistance given to ERF grantees affected children's outcomes and instructional practice relative to instructional content and outcomes in the absence of ERF. The study assesses the impact of ERF by comparing child outcomes and instructional practice in the treatment and comparison groups drawn from the universe of applicants for the FY 2003 grant competition. The treatment group consisted of 4-year-olds attending preschool in 28 ERF grantee sites, whereas the comparison group consisted of children attending preschool in 37 sites that applied for but did not receive ERF funds.

The remainder of this report presents the findings from the descriptive and impact analyses conducted for this study.

Figure 1.1. ERF conceptual framework



Chapter 2. Study Design

The National Evaluation of Early Reading First (ERF) uses a regression discontinuity design to assess the impact of ERF funding and program support for preschools on the language and literacy preparedness of preschool children. This study design takes advantage of the fact that the U.S. Department of Education (ED) is required to follow a formal, structured process for selecting grantees to receive ERF funding. In its published announcement of the availability of ERF grants for FY 2003 (*Federal Register* of March 11, 2003), ED established criteria for scoring each application received. Applications were reviewed and scored according to these criteria by independent reviewers. ED then awarded ERF grants to the grant applicants with the highest application scores, progressing down the score distribution until all funding available for the fiscal year had been allocated. In this way, 30 grants were awarded to the grant applicants with scores equal to or greater than 74; applicants with scores below 74 were not awarded grants.

Impact estimates were obtained by comparing child outcomes and teacher practices in funded sites to those in unfunded sites, controlling for a smooth function of the application score. Because the application scores fully reflected the selection rule used to award ERF grants, this approach will produce unbiased estimates of the effect of ERF if we control for the correct function of application score.

This chapter provides an overview of the sample, data sources, and analytic methods that are the foundation of the findings presented in Chapters 3 through 8. A fuller description of these issues is presented in Appendix A.

Sample Size and Sample Selection Process

The preschools that received FY 2003 ERF grants serve children as young as three years old. However, because of limited study resources, the study focuses on 4-year-old children who were attending ERF preschools in school year 2004–2005 and who were expected to enter kindergarten in the following school year.

The sample of ERF applicants for the study includes 28 of the 30 applicants who received an ERF grant and 37 of the 67 unfunded applicants with the highest application scores who agreed to participate in the study.

Approximately three classrooms were randomly selected from each participating site (see Table 2.1). The study team randomly selected approximately 11 4-year-old students per classroom whose parents had provided written consent for participation in the study. This section describes the final sample of sites, preschool teachers surveyed, classrooms observed, and students assessed.

Table 2.1 Sample Sizes for National Evaluation of ERF

Unit of Analysis	Funded sample size	Unfunded sample size	Total
ERF grantees/unfunded applicants	28	37	65
Preschools	86	75	161
Classrooms observed	78	91	169
Teachers surveyed	92	102	205
Children assessed	803	855	1,658

The site-selection process began with the 124 sites that submitted full applications to the 2003 grant competition. Figure 2.1 graphically displays the site-level sampling process. The treatment group consists of 28 of the 30 sites that were awarded ERF grants in October 2003. Two successful applicants were excluded from the study because they voluntarily left the program and were no longer ERF sites by spring 2005. All of the remaining 28 grantees agreed to participate in the study.

The comparison group sample began with the 94 sites that applied for but did not receive an ERF grant. Thirty-two unfunded sites were eliminated and not asked to participate for several reasons. Since the regression-discontinuity design makes use of comparison sites with scores close to those of the funded sites, the lowest-scoring 23 applicants—those that scored below 42.4—were not contacted during the recruiting process. Five additional unfunded sites and their associated 25 preschools were removed from the sample because they received a grant in a subsequent round of ERF funding.[10] In addition, three unfunded sites were excluded because they did not meet the criteria for participation in the study.[11] Of the 63 remaining unfunded sites that were contacted for inclusion in the study, 37 sites (59 percent) participated. (see Appendix B for additional information about the site and preschool selection and recruiting process.)

Once we arrived at the final sample of 28 funded sites and 37 unfunded sites, we continued the selection and recruitment process with preschools in those sites. Applicants typically consisted of collaborations of 5–7 preschools. We eliminated 32 preschools in these sites from the sample: 25 unfunded preschools because they were funded by ERF in the 2004 competition and 8 unfunded preschools that served children in special circumstances—for instance, migrant children only (see Figure 2.2).

Once we arrived at the sample of 157 funded and 246 unfunded preschools eligible for the study, the recruiting process continued. Because ED encouraged collaborations of diverse types of preschools to apply for 2003 ERF grants (for example, school-district-administered preschools, Head Start centers, and independent child-care centers), in many unfunded sites the original applying agency did not exercise management control of some of preschools that had been part

[10]Some ERF applicants listed different preschools in their 2003 and 2004 applications. The five unfunded sites that were removed because they were awarded 2004 ERF grants had substantial overlap between the preschools in their successful 2004 applications and the preschools in their unfunded 2003 application. Another four unfunded sites that later received grants in 2004 were included in our sample of sites because there was little to no overlap between the preschools listed in their 2003 and 2004 applications.

[11] Of the three unfunded sites that were excluded because they did not meet the criteria for participation in the study (one served only deaf children; one proposed to provide only wraparound care consisting mainly of lunch and nap; and one proposed to select preschools only after the ERF grant was awarded).

of the 2003 grant application. Thus, eligible preschools in unfunded sites were recruited individually. Only 121 (49 percent) of eligible unfunded preschools agreed to participate in the study. In the funded sites, the process of recruiting preschools was less challenging because the fiscal agent for the grant exercised some administrative control over the preschools. Only one of the 157 eligible funded preschool refused to participate.

After the sites and preschools in the study were recruited, approximately three classrooms were selected across all the participating preschools in each site with probabilities proportional to the number of 4-year-old children in each class.[12] From the preschools that agreed to participate, a total of 229 classrooms were randomly selected—103 ERF classrooms and 126 non-ERF classrooms (379 ERF classrooms and 186 unfunded classrooms were randomly excluded from the sample).

The study team randomly selected approximately 11 4-year-old students per classroom whose parents had provided written consent for participation in the study. Of the 1,914 selected 4-year old children, 803 ERF children and 855 non-ERF children were assessed in spring 2006 and included in the final analysis sample, which represents a response rate of 87 percent.

Surveys were sent to lead teachers in the ERF classrooms and non-ERF classrooms selected for the study and 92 ERF teachers and 102 non-ERF teachers completed the survey.[13]

In sites where child and teacher data was collected from 4 or 5 classrooms, 3 of those classrooms were randomly selected for the classroom observations; 78 ERF classrooms and 91 non-ERF classrooms were observed.

[12] The number of classrooms selected depended on the enrollment in each class and the number of participating classes. If a sample of 33 children could not be attained with 3 classrooms, then additional randomly selected classrooms were added. If only 1–2 eligible classrooms existed in a particular site, then only 1–2 classrooms were selected for the study.

[13] Because some teachers taught two classes (e.g., a morning or afternoon session), they were asked to complete a survey referencing only one of their randomly selected classes. For that reason, teacher surveys were sought from 98 teachers in funded classes and 114 teachers in non-funded classes.

Figure 2.1. Flow of applicants from 2003 ERF grant competition into treatment and comparison sites selected for study sample

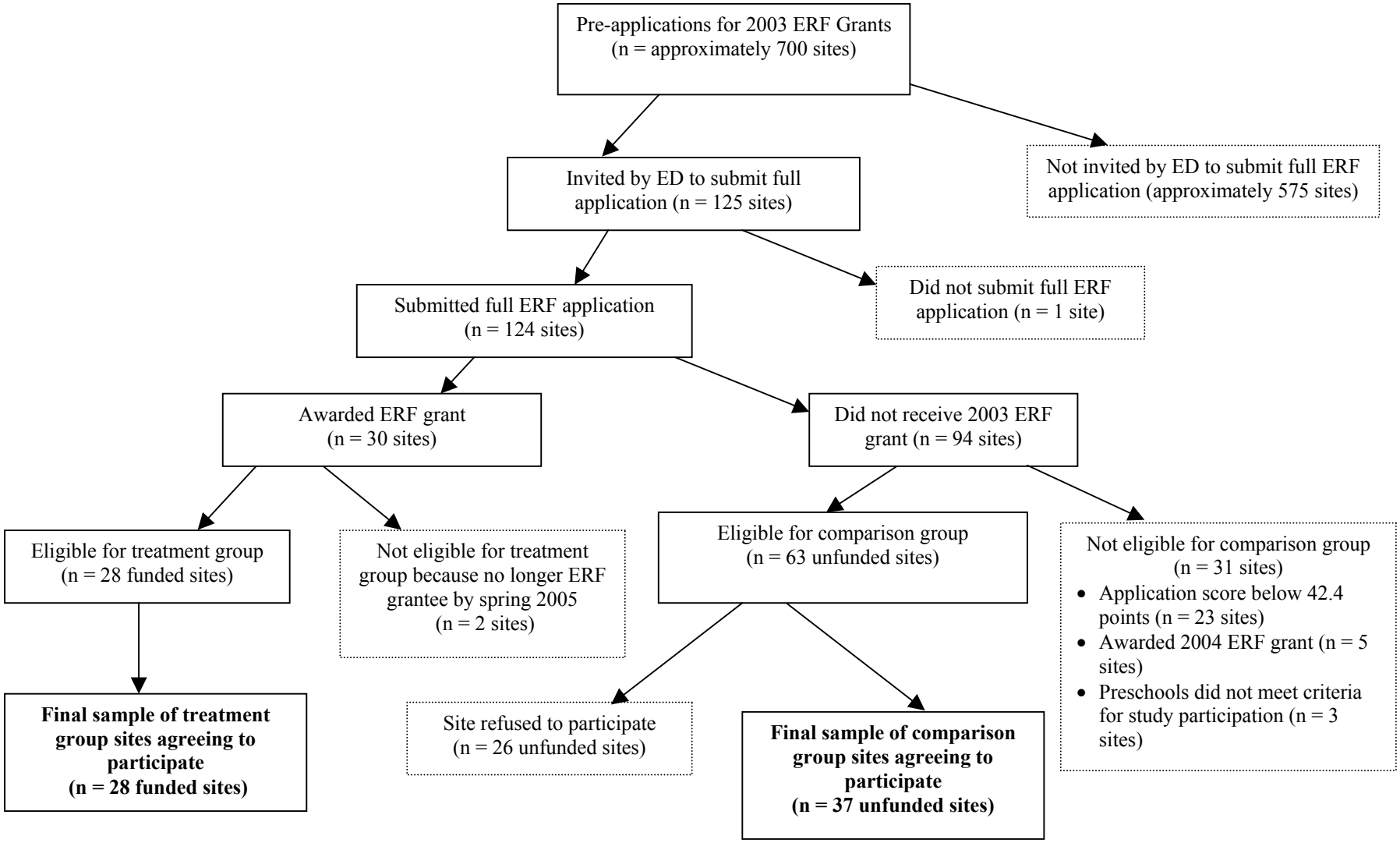
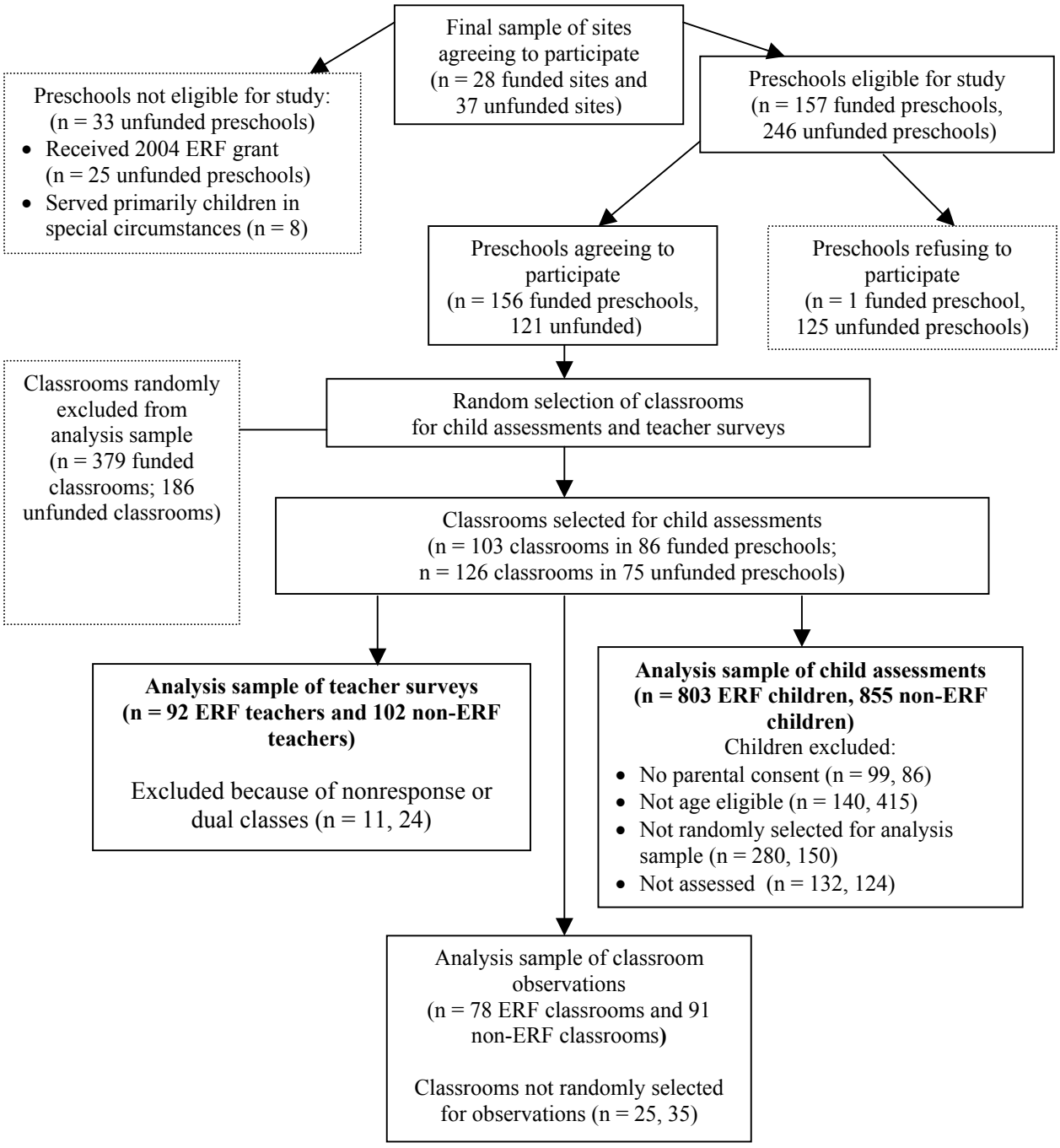


Figure 2.2. Flow of sites selected for study sample into analysis sample of children assessed, teachers surveyed, and classrooms observed



Data

Child outcomes are the primary focus of this evaluation. The study also measured ERF's impacts on key dimensions of teacher qualifications, classroom environment, and classroom practice that ERF sought to affect and that were, in turn, expected to affect children's language and literacy skills (see Figure 1.1 in Chapter 1).

The study team collected data for the evaluation from several sources. Trained staff directly assessed the language and literacy skills of children participating in the study. Trained observers measured classroom practice in a subsample of study classrooms. The teachers of all children in the sample and the director or principal of each preschool participating in the study completed a self-administered questionnaire. Teachers of the sampled children were also asked to rate each child's social-emotional development. The study team also obtained data from the preschools about children's school attendance for the 2004–2005 year. Finally, parents of the sampled children were interviewed by telephone.

Data were collected at two times: fall 2004 and spring 2005. The same data-collection instruments and procedures were used in the funded and unfunded sites.

Child Assessments. Table 2.2 shows the instruments that were used to measure children's language and literacy skills and social-emotional development and gives key data available on the psychometric properties of the instruments.[14] ERF was designed to affect the specific domains of emergent literacy—print and letter knowledge, phonological awareness, and oral language. *Print and letter knowledge* was measured by using the Print Awareness subtest of the Preschool Comprehensive Test of Phonological and Print Processing (Pre-CTOPPP, Lonigan et al. 2002). *Phonological awareness* was measured by using the Elision subtest of the Pre-CTOPPP (Lonigan et al. 2002). *Oral language* was measured by using two separate assessments: the Expressive One-Word Picture Vocabulary Test (EOWPVT, Brownell 2000) and the Auditory Comprehension subtest of the Preschool Language Scale, Fourth Edition (PLS-4, Zimmerman et al. 2002). Higher values for each measure are associated with higher literacy and language skills. All children were assessed in English in the spring. In the fall, Spanish-speaking children who did not pass the English proficiency screener, pre-LAS, were assessed in Spanish.

There were some concerns that an increased focus on literacy activities in preschools might lead teachers to focus less attention on social and emotional development; therefore, teachers were asked to complete a 30-item evaluation of *social-emotional development* for each child—the Social Competence and Behavior Evaluation: SCBE-30 (LaFreniere and Dumas 1996). This social-emotional evaluation was designed to provide measures of children's social competence, anger-aggression, and anxiety-withdrawal.

[14] Greater detail regarding the psychometrics of the child assessment and classroom observation instruments is provided in Appendix C.

Table 2.2. Data-collection instruments: child assessments

Instrument name	Domain measured	Psychometric information from published sources
(Pre-LAS)[1]	English proficiency screening	Internal consistency reliability = .86–.90
Preschool Comprehensive Test of Phonological and Print Processing (Pre-CTOPPP)[2]	Print and letter knowledge	Test of Preschool Early Literacy (TOPEL): • Internal consistency reliability = .95 • Test-retest reliability = .89
	Elision[3]	Internal consistency reliability = .71-.88.
Expressive One-Word Picture Vocabulary Test (EOWPVT) [4]	Expressive vocabulary	• Internal consistency reliability coefficients = .96–.98 • Test-retest reliability = .95
Preschool Language Scale (PLS-4)[5]	Auditory comprehension	• Test-retest reliability = .83–.91 • Internal consistency reliability coefficients = .83–.90
Social Competence & Behavior Evaluation (30-item)—Teacher Rating[6]	• Social competence • Anger-aggression • Anxiety-withdrawal	Internal consistency reliability coefficients = .85–.92

[1] Duncan, S.E., and DeAvila, E.A. (1998). *Pre-LAS 2000*. Monterey, CA: CTB/McGraw-Hill.
[2] Lonigan, C., Wagner, R., Torgesen, J., and Rashotte, C. (2007). *The Test of Preschool Early Literacy (TOPEL)*. Austin, TX: PRO-ED.
[3] Internal-consistency reliability coefficients of Elision subtest from unpublished tabulations using data from the Head Start Impact Study (U.S. Department of Health and Human Services 2005) and the forthcoming Even Start Classroom Observations and Interventions and Preschool Curriculum Evaluation Research studies, both being conducted by IES.
[4] Brownell, R. (2000). *Expressive One-Word Picture Vocabulary Test Manual*. Novato, CA: Academic Therapy Publications.
[5] Zimmerman, I. L., Steiner, V.G., and Pond, R.E. (2002). *Preschool Language Scale-4th Edition, Examiner's Manual*. San Antonio, TX: The Psychological Corporation.
[6] La Freniere, P. J., and Dumas, J. E. (1996). "Social competence and behavior evaluation in children ages 3 to 6 years: The short form (SCBE-30)," *Psychological Assessment, 8,* 369–377.

Classroom Observations. Through direct observations of the preschool classrooms of the assessed children, the ERF evaluation team sought to measure the degree to which ERF grant support changed instructional practice and overall quality of the preschool classrooms. Table 2.3 shows the dimensions of classroom practice and quality measured by the two instruments used for observation—the Teacher Behavior Rating Scale (TBRS)[15] and 11 items from the Early Childhood Environment Rating Scale-Revised (ECERS-R) that form the Teaching and Interactions Subscale.[16] Trained members of the study team conducted the classroom observations.

[15] Landry et al. (2004). "Teacher Behavior Rating Scale (TBRS)," unpublished research instrument.
[16] Harms, T., Clifford, R.M., and Cryer, D. (1998). *Early Childhood Environment Rating Scale: Revised Edition*. NY: Teachers College Press.

Table 2.3. Data-collection instruments: observations

Classroom Observation Instrument name	Primary dimensions, subscales tapped	Psychometric tnformation from ERF sample
Teacher Behavior Rating Scale	Language and Literacy Environment and General Preschool Quality • Book-reading practices • Oral language use by lead teacher • Phonological awareness activities • Print and letter knowledge • Written expression • Child portfolios • Dynamic assessment • General teaching behaviors • Classroom community • Teacher sensitivity • Lesson planning • Quality and organization of activity centers • Quality of team teaching • Math concepts	Internal consistency for subscales = .66–.94 Interrater reliability = .75–1.0
ECERS-R Teaching and Interactions (11 items)	Preschool quality with emphasis on use of language and communication • Interactions among children • Encouraging children to communicate • Discipline • Supervised free play • General supervision of children • Greeting/departing • Group time • Informal use of language • Supervision of gross motor • Reasoning skills • Staff-child interactions	Internal consistency = .85 Interrater reliability = .87–.92

Other Data Sources. The evaluation team also developed self-administered surveys that the teachers and preschool principals or directors completed in the fall of 2004 and spring 2005. Parents of children in the study were interviewed through computer-assisted telephone interviewing (CATI) technology. The major constructs measured by each of these surveys are shown in Table 2.4. The team also conducted in-depth telephone interviews with grantee directors for each of the 28 funded grantees in the sample to learn about their use of ERF funds, and to obtain background information about the context in which ERF grants were implemented. (Appendix B provides additional information on data-collection procedures.)

Table 2.4. Data-collection instruments: surveys and in-depth interviews

Target respondent	Primary dimension(s) tapped
Teachers	• Demographics • Background • Education • Experience • Classroom characteristics • Curricula used & trained on • Assessments used • Professional development methods, hours, and topics
Center directors	• Demographics • Background • Education • Experience • Classroom characteristics • Curricula used & trained on • Assessments used • Professional development methods, hours, and topics • Funding sources
Parents	• Demographics • Child preschool experience • Literacy resources available • Weekly non-school literacy activities

Analytic Methods for the Impact Analysis

The impact analysis uses a regression discontinuity design to address the following research questions:

- What are the impacts of ERF on children's language and literacy and social-emotional indicators?

- What are the impacts of ERF on the quality of language and literacy instruction, practice, and materials?

- Do ERF impacts vary across subgroups defined by key child, teacher, or program characteristics?

The "discontinuity" in grant awards based on the application scores was used to identify ERF impacts. To estimate impacts, we used regression models to compare child and classroom outcomes in the funded sites (the treatment group) to those in the unfunded sites (the comparison group), and we controlled for a smooth function of grant application score. If one assumes that the outcome variables exhibit a stable continuous relationship with the application score and that we have correctly modeled this relationship, the sharp discontinuity in ERF grant receipt at the score cutoff, conditional on this smooth function of application score, identifies ERF's impacts.

Missing values of covariates were imputed using methods described in Appendix A. Sampling weights were used to account for the random selection of classrooms to the analysis sample, and to give equal weight to each site (see Appendix A). Appendix A discusses the statistical models used to estimate impacts, the robustness of our findings for a broad range of analytic decisions, and the statistical power for detecting impacts under the sample design.[17]

[17] The minimum detectable impact in effect size units is 0.30 standard deviations for a typical child outcome and 0.89 standard deviations for a typical classroom outcome.

Chapter 3. Characteristics of Participating Children and Families

The ERF program was designed to serve predominately children in low-income communities. The governing statute contains several requirements, and for FY 2003, the Department of Education (ED) had several preferences about the characteristics of children and families that should be served by the ERF program. Congress required ERF applicants to be located in school districts

- that have the highest numbers or percentages of children in kindergarten through third grade needing reading improvement
- that are generally located in low-income communities

ED also expressed an interest in receiving applications from preschools serving large numbers of children with special needs, including English language learners (ELLs), through an invitational priority in the full application, although such applications were not awarded additional points in scoring.

In this chapter, we summarize the characteristics of children and families in the 2003 cohort of ERF grantees as reported in the spring 2005 survey of parents. When data supports such a comparison, we compare the characteristics of the ERF sample with the characteristics of the general population of children nationally to assess the extent to which the congressional mandate to serve children predominately from low-income families and ED's priority to target students with limited English were achieved.

In order to provide additional context for the study findings and facilitate comparison to other studies, we discuss how children in ERF preschools compare to those in a nationally representative sample of Head Start preschools.[18] Head Start is the largest federally funded preschool program for low-income children and requires that most participants be from households with income below the federal poverty level. Because of the applicant-eligibility requirements for ERF and ED's competitive priority for preschools where at least 75 percent of children are eligible for free or reduced-price lunches (or where at least 75 percent of the children enrolled in the elementary school in the school attendance area in which that preschool is located qualify to receive free or reduced price lunches), most ERF grantees are located in school districts in which a large percentage of children are eligible for free or reduced-price school meals and which have income eligibility cutoffs of 130 percent and 185 percent of the federal poverty level, respectively.[19] Thus, the Head Start program uses a lower income threshold for allocating its services to economically disadvantaged children than ERF uses.

[18] The Head Start Family and Child Experiences Survey (FACES) was first conducted in 1997 with a national probability sample of Head Start children. A 3-stage design was used to sample 3,648 children from 40 Head Start programs across the 50 States, Puerto Rico, and the Territories of the United States. Of those, 3,179 families (87 percent) provided signed consent forms before the fall 1997 data collection. (U.S. Department of Health and Human Services, 2002, *A Descriptive Study of Head Start Families: FACES Technical Report I,* pp. 15–19. http://www.acf.hhs.gov/programs/opre/hs/faces/reports/technical_report/technical_report.pdf)

[19] No income-eligibility requirements are imposed for participation in ERF at the preschool or child level. However, eligibility to receive ERF grants is extended to Local Education Agencies (LEAs) that are eligible to receive a subgrant under the Reading First program or to public and private organizations that are located in one of those LEAs, or to one or more LEAs in applying in collaboration with such an organization or agency. To be eligible for a Reading First state subgrant, an LEA must have large numbers or percentages of students in grades K–3 who read

We compared the characteristics of ERF children to those in unfunded sites to provide some context for interpreting the impact findings presented later in this report. It is important to note that the ERF and non-ERF samples are not designed to be equivalent (which one would expect in a randomized design). Further, the sample of students at preschools that applied for but were not awarded ERF grants is not designed to be representative of all students at unfunded preschools. Because of the regression discontinuity design, we selected a sample of schools in the interval closest to the cutoff point for application scores that were willing to participate in the study. As a result, the funded and unfunded samples may have different characteristics; inclusion of the application score variable in the regression analysis is intended to control for these differences in estimating impacts on child outcomes.

In the following sections, we describe ERF children and families along a series of indicators—household income, national origin and languages spoken, race and ethnicity, and parental marital status—to demonstrate that the ERF program does in fact serve a disadvantaged population, with a higher proportion of Hispanic children, children of immigrants, and English-language learners (ELLs) than occurs in the national population of children in this age group.[20] We also present fall 2004 assessment scores, which show that our sample was functioning below national norms for 4-year-olds on several assessments at the outset of the study. These comparisons demonstrate how different the ERF sample is from the non-ERF sample before controlling for selected covariates, and they provide important context for interpreting the findings presented in this report.[21]

Parent's Household Income

With 35 percent of the households of ERF participants reporting monthly income of less than $1,500 (see Table 3.1), ERF participants are more likely to be low-income than the average child in the U.S. On an annualized basis, this level of monthly income would place the annual income of a family of four at approximately the federal poverty level. Nationally, about 17 percent of children ages 3 to 5 years old live in households with monthly income of less than $1,500.[22] As might be expected, given the different income-eligibility requirement for Head Start, the sample of ERF participants does not appear to be as disadvantaged economically as the Head Start sample, in which 66 percent of parents reported household income of $1,500 or less per month.[23] No differences are apparent in the income levels between sampled households in funded and unfunded sites.

---

below grade level *and* must meet one of the following criteria: (1) has a significant number or percentage of schools identified for school improvement under Title I, Part A (i.e., that fail to meet Annual Yearly Progress goals for two consecutive years), (2) include an empowerment zone or enterprise community as defined by the IRS, or (3) have the highest numbers or percentages of children counted for the purposes of Title I grants to LEAs in comparison to other school districts in the state. In practice, the percentage of students counted under Title I for that purpose is based on the percentage of those who are approved as eligible for free or reduced-price meals.

[20] The data reported for ERF participants are derived from self-reports by parents and are not independently verified. Also, because the survey response rate for parents was about 61 percent, some unmeasured nonresponse bias may exist and should be considered in interpreting these findings.

[21] Our sample selection process eliminated preschools or preschool classrooms that had large percentages of children with learning disabilities because of concerns about conducting assessments with those children. Hence, we are unable to conduct analyses of the extent to which the ERF program served children with learning disabilities.

[22] Calculations from Current Population Survey (U.S. Census Bureau, 2005).

[23] U.S. Department of Health and Human Services (January 2002) *A Descriptive Study of Head Start Families: FACES Technical Report I,* p. 47.

Table 3.1. Parental household income, by ERF funding status

	Overall	ERF participants	Children in non-ERF preschools	P-value	Head Start participants
Percent of participants with monthly household income:				.847[1]	
$500 or less	5.6	5.1	6.0		11.8
$501 to $999	13.6	12.5	14.6		29.6
$1,000 to $1,499	16.7	17.1	16.3		24.8
$1,500 to $1,999	19.0	20.1	18.1		14.4
$2,000 or more	36.3	36.3	36.3		15.7
% refused	8.8	9.0	8.7		unknown
Sample Size	1,146	545	601		2,983

[1] P-value is based on chi-squared test of association.
SOURCE: Spring survey of parents and Head Start FACES technical report (U.S. Department of Health and Human Services, 2002).

National Origin and Language of ERF Families

Table 3.2 shows that the parents of 39 percent of children served by ERF preschools were born in a country other than the United States. Nationally, about 23 percent of 3- to 5-year-olds in 2005 lived in households in which a parent was born in a foreign country.[24] Further, about half (51 percent) of the parents of ERF participants indicated that a language other than English was spoken most often at home. More parents of ERF participants were born outside of the U.S. compared to the FACES Head Start sample (39 percent compared to 19 percent).[25] Similarly, a larger fraction of ERF parents than Head Start parents reported that the primary language spoken at home was other than English (41 percent as compared to 36 percent).[26] Compared to children in the unfunded sites, the sample of children from preschools awarded ERF grants had a higher proportion of children whose parents were foreign born and who lived in households in which the primary language was not English.

[24] Calculations from Current Population Survey (U.S. Census Bureau, 2005).
[25] *A Descriptive Study of Head Start Families: FACES Technical Report I*, January 2002, p. 37.
[26] Ibid., p. 60.

Table 3.2. Parent national origin and language, by ERF funding status

	Overall	ERF participants	Children in non-ERF preschools		Head Start participants
	%	%	%	P-value[1]	%
National origin of parents					
% U.S. born	64.4	60.6	67.9	.022	81.3
% foreign born	35.5	39.3	32.1		18.7
Percent parents with language other than English spoken at home	45.5	50.6	40.8	.001	
Percent parents most frequently speaking language other than English	37.7	41.4	34.3	.025	35.7
Sample Size	1,146	545	601		3,120

[1] P-values are based on chi-squared test of association.
SOURCE: Spring survey of parents and Head Start FACES technical report (U.S. Department of Health and Human Services, 2002).

Race and Ethnicity

The survey results indicate that a majority of the ERF participants were children of color. Table 3.3 shows that Hispanic children composed the largest ethnic group of ERF participants (46 percent). This proportion is more than twice the national proportion of Hispanic children ages 3 to 5, which in 2005 was estimated to be 21 percent.[27] Compared to the 4-year-olds in the Head Start sample, the ERF program served more Hispanic children (46 percent versus 30 percent) and fewer African-American children (24 percent versus 26 percent) and white children (27 percent versus 31 percent).[28] Within the ERF sample, significant differences were found between the funded and unfunded sites, with ERF program sites serving more Hispanic children and fewer white children than sites that did not receive ERF funding.

[27] Current Population Survey, March 2005.
[28] *A Descriptive Study of Head Start Families: FACES Technical Report I,* p. 29.

Table 3.3. Child race and ethnicity, by ERF funding status

	Overall	ERF participants	Children in non-ERF preschools		Head Start Participants Age 4
	%	%	%	P-value[1]	%
Race or ethnicity of child				.010	
% African American	23.8	23.8	23.9		26.1
% Hispanic	42.7	46.2	39.5		30.0
% White	27.2	22.8	31.1		31.4
% Other	6.3	7.2	5.5		11.6
Sample Size	1,145	543	602		1,991

[1] P-value based on chi-squared test of association.
SOURCE: Spring survey of parents and Head Start FACES technical report (U.S. Department of Health and Human Services, 2002).

Parent Marital Status

The parents of almost 40 percent of the ERF participants were unmarried, including 12 percent who were separated, divorced, or widowed and 28 percent who had never been married (see Table 3.4).[29] According to the March 2005 Current Population Survey (CPS), 28 percent of households with 3- to 5-year-olds contain parents who are unmarried, including 19 percent, who had never been married. Compared to households nationally with 3- to 5-year-old children, a larger proportion of parents of ERF children are unmarried. Although the difference is not statistically significant at conventional significance levels, parents in funded sites had a somewhat lower rate of being single parents than parents in the unfunded sites. The proportion of parents who are unmarried in the ERF sample is much lower than in the sample of 4-year-olds in Head Start (58 percent).[30]

[29] The respondent for a family was the person who signed the parent consent form in fall 2004. In the absence of that person, another adult with whom the child lived was interviewed. The birth mother was the respondent for the spring 2005 survey in 80 percent of the cases; the birth father was the respondent in 13 percent of the surveys; the child's grandmother was the respondent for 4 percent of the children.
[30] *A Descriptive Study of Head Start Families: FACES Technical Report I,* p. 37.

Table 3.4. Parent marital status, by ERF funding status

	Overall	ERF Participants	Children in non-ERF Preschools		Head Start Participants
	%	%	%	P-value[1]	%
Parent marital status				.070	
% married	59.9	63.5	56.7		42.1
% unmarried (total)	39.8	36.5	42.9		56.8
% separated/divorced/widowed	11.7	11.0	12.3		23.1
% never married	28.2	25.5	30.6		33.7
Sample Size	1,146	545	601		3,120

[1] P-value based on chi-squared test of association.
SOURCE: Spring survey of parents and Head Start FACES Technical Report, 2002.

Child Standardized Assessment Scores

Table 3.5 shows that children in both funded and unfunded sites scored below national norms (mean score of 100) for 4-year-old children on Print Awareness, Expressive Vocabulary, and Auditory Comprehension in the fall 2004 assessments.[31] Due to the timing of these assessments, some of which did not occur until two to three months into the school year, these scores are not true baseline measures; however, they do provide some indication of the degree to which the ERF sample is disadvantaged relative to other children nationally. Fifteen percent of children in the funded sites and 8 percent of children in the unfunded sites were assessed in Spanish after failing the English language screener. Data for the Head Start sample are not included because the FACES study did not use these child assessments.

Table 3.5. Standard scores on fall 2004 assessments, by ERF funding status

	ERF Participants	Children in non-ERF preschools	
	Mean	Mean	P-value[1]
Standardized Assessment Score			
Print Awareness	93.58	90.83	0.35
Expressive Vocabulary (EOWPVT)	82.90	82.77	0.82
Auditory Comprehension (PLS-IV)	91.71	90.50	0.32
Sample Size	805	864	

[1] P-values (of adjusted difference in means), two-tailed test.
SOURCE: ERF fall child assessments.

[31] Standardized test scores are based on a mean of 100 and a standard deviation of 15.

In summary, ERF participants appeared to be more disadvantaged than the national average. A relatively large proportion of children served by ERF grantees had some characteristics associated with disadvantage. More than one-third of the ERF sample reported monthly income of less than $1,500, compared to 17 percent of households with 3- to 5-year-olds nationally. Children in this cohort were also more likely than children nationally to come from single-parent households (40 percent compared to 28 percent), be Hispanic (46 percent compared to 21 percent), and have foreign-born parents (39 percent compared to 23 percent). About four in 10 ERF parents (41 percent) reported that the primary language spoken in the home was something other than English. Initial scores on standardized assessments suggest that children were functioning below national norms when they entered the ERF program.

While the ERF sample appeared more disadvantaged than the general population of households that had 3- to 5-year-old children, they appeared less disadvantaged economically than the sample of 4-year-olds in the FACES Study. These patterns are consistent with Head Start's participation requirements, which are more tightly focused on disadvantaged children.

Compared to the unfunded preschools in our sample, ERF preschools had more foreign-born parents (40 percent versus 32 percent), more Hispanics (46 percent versus 40 percent), and more children whose parents were married (although the latter was not statistically significant).[32] There were no differences in family income or initial standardized assessment scores between the students at funded preschools and students at unfunded preschools.

[32] The analysis of child outcomes takes account of these differences.

Chapter 4. Characteristics of Programs Receiving ERF Funding

The types of preschools awarded ERF funds varied widely with regard to their sources of funding, their operating schedules, and the characteristics of their teachers. These factors may affect the way that ERF is implemented and the value of the additional resources that ERF provides. In this chapter, we describe the preschools in the national evaluation's sample—both funded and unfunded—and compare them on these characteristics. The data, provided by either the preschool directors or teachers in the spring of 2005, were from preschools drawn from the FY 2003 cohort of ERF applicants.

Overall, the vast majority of ERF preschools (95 percent) combine ERF funding with other government funding sources, which is consistent with the goal of the program to enhance the quality of existing programs that serve particularly children from low-income families. The most common funding sources are state and local education agencies, state child-care funds, and Head Start, which were received by 56 percent, 38 percent, and 36 percent of ERF preschools, respectively. Just over half of ERF preschools received funding from only one of these sources, while over 40 percent received funding from two or more sources. No significant differences in the number or types of funding sources were reported by ERF and non-ERF preschools.

The schedule on which ERF preschools operate and the characteristics of their teachers provide useful context for examining study findings. Three-quarters of ERF preschools are full-day programs (operating for an average of 8 hours per day), 62 percent have a class size of 20 children or fewer, and almost 70 percent have a staff-to-child ratio of 1:10 or better. Three quarters of ERF teachers have bachelor's degrees, 67 percent have teaching certificates or licenses, and most (87 percent) had completed college courses in early-childhood education or development. Many teachers had completed at least 6 college courses in teaching reading to elementary school children (67 percent) and/or teaching language and literacy skills to children in a preschool setting (79 percent).

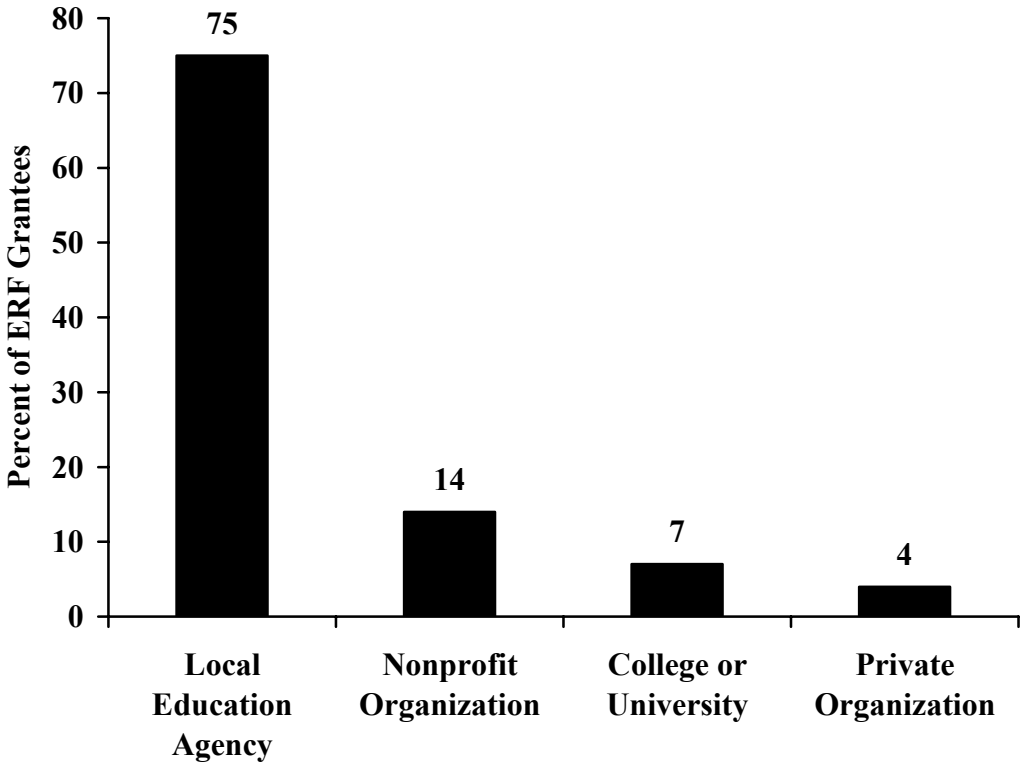
In the following sections, we describe the ERF programs with respect to four major dimensions: funding levels, funding sources, program operations, and teacher characteristics.

Grantee Funding Levels—Overall and by Child

The FY 2003 ERF grants were awarded in October 2003. Sites were expected to begin implementing the program by January 2004. Total funding levels for the 3-year period ranged from a high of $4.36 million to a low of $1.07 million per site. Three-quarters (75.5 percent) of grantee directors reported that their fiscal agent, with responsibility for overseeing the financial aspects of the ERF grant, was their local education agency[33] (see Figure 4.1).

[33] Although just over half of the grantees reported receiving funds from their state or local education agencies, three-quarters reported that their fiscal agent for the ERF grant was their local education agency.

Figure 4.1. Fiscal Agents of ERF Grants



An additional 14 percent of grantee directors indicated that their fiscal agent was a nonprofit organization; 7 percent reported that a college or university fulfilled the role of fiscal agent; and the remaining 3.5 percent reported their fiscal agent to be a private organization.

Based on the reported number of preschool children expected to be served by the FY 2003 grantees, ERF grant amounts ranged from a high of $6,726 per child to a low of $402 per child per year. The median ERF allocation across the 28 grantees evaluated in the FY 2003 cohort was $3,549 per preschool child per year.[34] These funds are in addition to the other government funding sources received by the preschools. To provide perspective, annual average Head Start funding per child in Fiscal Year 2003 was $7,092.[35]

Funding Sources

ERF is designed to enhance instructional practice and classroom environments in existing early-education programs, such as Title I preschools, state pre-kindergarten programs, Head Start centers, child-care centers (including those receiving state child-care funds), and family-literacy programs such as Even Start. The diverse government funding sources of ERF preschools reflect that goal.

[34] The methodology used to compute the ERF allocation per child is described in Appendix B, "Data Collection Methods."

[35] U.S. Department of Health and Human Services (April 2004), *Head Start Program Fact Sheet Fiscal Year 2003,* Administration for Children and Families. http://www.acf.hhs.gov/programs/hsb/research/2004.htm.

The vast majority of ERF preschools received at least one other source of government funding; only 4.7 percent reported no other government funding (see Table 4.1). Just over half of the ERF preschools in the study had a single source of other government funding, and just over 40 percent had two or more other government funding sources. There were no differences in the number of other government-funding sources for ERF and non-ERF preschools: both on average received funds from approximately 1.6 other government sources.

Table 4.1. Number of different sources of other government funding for preschools, by ERF funding status

	All preschools	ERF preschools	Non-ERF preschools	P-value[1]
Number of other government funding sources				
0	3.8%	4.7%	3.0%	
1	53.4%	53.1%	53.7%	
2	26.0%	26.6%	25.4%	
3	15.3%	14.1%	16.4%	
4	1.5%	1.6%	1.5%	
Mean number (standard deviation)	1.57 (0.85)	1.55 (0.85)	1.60 (0.85)	0.74
Sample size	131	64	67	

[1]P-value is based on Student's t-test.
SOURCE: Spring surveys of preschool directors.

According to their directors, many ERF preschools received funding from state and local education agencies (56 percent), state child-care funds (38 percent), or Head Start (36 percent) (see Table 4.2). Federal Even Start and county or city governments were less common sources of funding, accounting for 7.8 percent and 6.3 percent of funded sites, respectively. Unfunded applicant sites did not significantly differ from ERF sites in the sources of funding received.

Table 4.2. Types of other government funding sources received by preschools, by ERF funding status (as percent of preschools receiving each source of funding)

	All preschools	ERF preschools	Non-ERF preschools	P-value[1]
Other government funding source				
State and local education agency[2]	52.7%	56.3%	49.3%	0.42
Child care[3]	39.7%	37.5%	41.8%	0.62
Federal Head Start program	36.6%	35.9%	37.3%	0.87
Other	13.0%	10.9%	14.9%	0.50
County or city government	8.4%	6.3%	10.4%	0.39
Federal Even Start program	6.9%	7.8%	6.0%	0.68
Sample size	131	64	67	

[1] All p-values are based on chi-squared tests of association.
[2] Funds from state and local education agencies include funds from state education agencies, independent school districts, and other sources, channeled through the state education agency.
[3] Child-care funds include state child-care funds and child-care vouchers.
SOURCE: Spring surveys of preschool directors.

Table 4.3 presents data about the extent to which preschools combine funding from Head Start, state or local education agencies, and child-care funds and the manner in which those funds are combined. Of the ERF preschools receiving Head Start funding, approximately one-half relied on Head Start as their only other source; of the ERF preschools receiving funding from state or local education agencies, approximately one-half relied on that as their only other source. However, among the preschools that received funding through child-care subsidies, a much lower percentage—just over 20 percent—relied solely on those subsidies as their only other source of funding. Unfunded applicant sites did not differ significantly from ERF sites in how funding sources were combined.

Table 4.3. Overlap in sources of funding from Head Start, state or local education agencies, and child-care funds for preschools, by ERF funding status

	All preschools	ERF preschools	Non-ERF preschools	P-value[1]
Funding source				
Head Start	36.6%	35.9%	37.3%	0.87
Head Start only	18.3%	17.2%	19.4%	0.74
Head Start & state or local education agency funds	7.6%	3.1%	11.9%	0.06
Head Start & child-care funds	3.0%	4.7%	1.5%	0.28
State or local education agency funds[2]	52.7%	56.3%	49.3%	0.42
State or local education agency funds only	21.4%	26.6%	16.4%	0.16
State or local education agency funds & child-care funds	5.3%	7.8%	2.9%	0.21
Child-care funds[3]	39.7%	37.5%	41.8%	0.62
Child-care funds only	11.5%	7.8%	14.9%	0.20
Sample size	131	64	67	

[1] All p-values are based on chi-squared tests of association.
[2] Funds from state and local education agencies include funds from state education agencies, independent school districts, and other sources, channeled through the state education agency.
[3] Child-care funds include state child-care funds and child-care vouchers.
SOURCE: Spring surveys of preschool directors.

Program Operating Schedules

Data from the Head Start FACES 2000 study indicate that the provision of full-day Head Start services was correlated with greater cognitive gains.[36] Children in full-day Head Start classes showed larger fall-to-spring gains in letter recognition and early-writing skills than those in part-day classes. Although causal inferences cannot be drawn from this correlational study within the context of this research, it is interesting to document the number of operating days per year and hours of operation per day for the schools in our sample as important descriptive characteristics. The survey data indicate that three-quarters of ERF preschools operate for a full day (defined as open 6 or more hours per day) and about half (51 percent) operate for part of a year (see Table

[36] U.S. Department of Health and Human Services (May 2003). *Head Start FACES 2000: A Whole Child Perspective on Program Performance*. (http://www.acf.hhs.gov/programs/opre/hs/faces/reports/faces00_4thprogress/faces00_title.html)

4.4). On average, ERF preschools are open for 8 hours a day. The majority (73 percent) of the ERF-funded preschools are open 5 days a week. The ERF preschools are open for an average of 42 weeks a year, with the number of weeks of operation ranging from 27 to 52.

While we observed no significant differences between funded and unfunded preschools in the average number of hours they were open per day and the weeks they were open per year, a significantly higher proportion of non-ERF preschools were open 5 days a week compared to ERF preschools (88 percent versus 73 percent), and the mean number of operating days per week was correspondingly greater in the non-ERF funded preschools (4.9 days versus 4.7 days).

Table 4.4. Periods of operation of preschools participating in the ERF evaluation, by ERF funding status

	All preschools	ERF preschools	Non-ERF preschools	P-value[1]
Hours of operation per day				
< 3.5 hours	6.2%	1.6%	10.6%	
3.5 to 5.9 hours	13.8%	23.4%	4.5%	
6 to 8.9 hours	41.5%	37.5%	45.5%	
≥ 9 hours	38.5%	37.5%	39.4%	
Median	7.0	7.0	7.5	
Mean (SD)	7.9 (3.0)	7.9 (3.0)	7.9 (3.0)	0.99
Sample size	130	64	66	
Days of operation per week				
3 days	2.3%	3.1%	1.5%	
4 days	16.8%	23.4%	10.4%	
5 days	80.9%	73.4%	88.1%	
Mean (SD)	4.8 (0.5)	4.7 (0.5)	4.9 (0.4)	0.05
Sample size	131	64	67	
Weeks of operation per year:				
< 40	50.4%	50.8%	50.0%	
≥ 40	49.6%	49.2%	50.0%	
Mean (SD)	41.9 (7.9)	41.8 (7.8)	42.0 (8.0)	0.89
Sample size	125	61	64	

[1] P-values are based on Student's t-test.
NOTE: Head Start defines a full-day program as 6 hours or more and a part-time program as at least 3.5 hours.
SOURCE: Spring surveys of preschool directors.

Class Size, Composition, and Adult Supervision

Class size and staff-to-child ratios are important components of the quality standards for early-childhood programs (Barnett, Schulman, and Shore 2004; NICHD Early Child Care Research Network 1999). In this section, we describe the size and composition of classrooms in the study sample. Of the 194 classrooms in the study sample, 92 received ERF funding, and 102 did not. All were preschool classes serving the study's target population of children who were expected to attend kindergarten in the following school year—most, but not all, of whom were 4 years old in fall 2004.

Some research has found that lower group sizes and better staff-to-child ratios in early-childhood settings are positively correlated with children's language, cognitive, and social functioning[37] (Barnett, Schulman, and Shore 2004; NICHD Early Child Care Research Network 1999 and 2002; Vandell and Wolfe 2000). According to the widely used guidelines of the National Association for the Education of Young Children (NAEYC), 4-year-old children should be in groups of 16 to 20 children, with a staff-to-child ratio between 1:8 and 1:10.[38] All groups, regardless of age, should have at least two teachers. Overall, the majority (63.5 percent) of ERF classrooms met or exceeded these criteria. Although causal inferences cannot be drawn from these correlational studies, it is useful to document group sizes and staff-to-child ratios in the context of this literature.

The number of children enrolled in the ERF preschool classes varied from as few as 6 per class to as high as 48 per class (see Table 4.5). The average class size was 23 children, but class size varied tremendously. Sixty-two percent of the children were enrolled in classes of 20 or fewer children (the NAEYC criteria for a high-quality program). On average, there were 3 special needs children per ERF classroom. Because of the criteria used to select classrooms for this study, the overwhelming majority (96 percent) of classes included 4-year-old children. There were no significant differences between ERF and non-ERF classrooms along any of these dimensions.

Table 4.5. Classroom characteristics, by ERF funding status

	All classrooms	ERF classrooms	Non-ERF classrooms	P-value
Number of children enrolled in the class				
Less than 16	15.0%	13.1%	16.7%	
16 to 20	46.9%	48.9%	45.1%	
More than 20	38.1%	38.0%	38.2%	
Mean (SD)	22.6 (8.8)	22.7 (8.9)	22.4 (8.6)	0.81
Range	6 to 48	8 to 44	6 to 48	
Number of special needs children enrolled in the class				
0	28.9%	26.1%	31.4%	
1 or 2	32.0%	33.7%	30.4%	
3 or 4	10.3%	10.9%	9.8%	
5 or 6	10.3%	13.0%	7.8%	
7 to 9	4.1%	5.4%	2.9%	
10 or more	6.2%	4.3%	7.8%	
Mean (SD)	2.8 (3.9)	2.8 (4.2)	2.7 (3.6)	0.82

[37] Several organizations, including the National Association for the Education of Young Children, set standards for a voluntary early childhood program accreditation process. State regulations on teacher-child ratios and class size in early childhood programs vary widely (Vandell and Wolfe, 2000).

[38] The National Institute for Early Education Research uses similar benchmarks in their Quality Standards Checklist for state pre-K programs: maximum class size should be 20 or lower, and staff-to-child ratio should be 1:10 or lower (National Institute for Early Education Research, 2006, p. 32).

Table 4.5. Classroom characteristics, by ERF funding status—*Continued*

	All classrooms	ERF classrooms	Non-ERF classrooms	P-value
Percentage of children enrolled in the class who are special needs				
0 percent	28.9%	26.1%	31.4%	
1 to 10 percent	28.4%	30.4%	26.5%	
11 to 20 percent	14.4%	16.3%	12.7%	
21 percent or more	20.0%	20.7%	19.6%	
Missing	8.3%	6.5%	9.8%	
Mean (SD)	12.4 (15.8)	12.8 (15.3)	12.0 (16.4)	0.75
Ages of children enrolled in the class				0.10[1]
3-year-olds only	0.5%	1.1%	0.0%	
4-year-olds only	6.2%	4.3%	7.8%	
5-year-olds only	2.6%	3.3%	2.0%	
3- and 4-year-olds	7.2%	3.3%	10.8%	
3- and 5-year-olds	0.0%	0.0%	0.0%	
4- and 5-year-olds	48.5%	56.5%	41.2%	
3-, 4-, and 5-year-olds	35.1%	31.5%	38.2%	
Number of paid staff members usually in the class				
1	11.9%	10.9%	12.7%	
2	59.8%	65.2%	54.9%	
3	18.6%	13.0%	23.5%	
4 or more	9.8%	10.9%	8.8%	
Mean (SD)	2.3 (0.9)	2.3 (0.8)	2.3 (0.9)	0.56
Staff-to-child ratio in the class				
1:10 or less	66.0%	68.5%	63.7%	0.49
Mean (SD)	10.9 (5.5)	11.1 (5.8)	10.8 (5.3)	0.74
Number of children absent on a typical day				
0	12.4%	17.4%	7.8%	
1 or 2	71.1%	70.7%	71.6%	
3 or 4	8.2%	6.5%	9.9%	
5 or 6	2.1%	1.1%	2.9%	
Mean (SD)	2.0 (0.6)	1.9 (0.5)	2.1 (0.6)	0.03
Sample Size	194	92	102	

[1] This p-value is based on chi-squared test of association; all other p-values are based on Student's t-tests.
SOURCE: Spring surveys of preschool teachers.

The number of paid staff members per class as reported by teachers varied, although the majority of classes (65 percent) were staffed by two teachers (see Table 4.5). Perhaps a more useful metric is the staff-to-child ratio in a classroom. Just over 68 percent of the ERF-funded classrooms maintained a ratio of one teacher to 10 or fewer children, the professionally accepted upper limit for ratios in preschool classrooms serving 4-year-olds. Differences between ERF and non-ERF classrooms were not statistically significant along any of these dimensions.

The one characteristic for which we observed a statistically significant difference between the ERF-funded and unfunded classrooms was in the area of child absenteeism. On a typical day, the unfunded classrooms reported a higher absentee rate than the funded classrooms. However, in

practical terms, the number of students absent on a typical day was close to two children, regardless of funding status.

Characteristics of Teachers

This section focuses on the teachers in the classrooms of the children selected for the evaluation. Differences that we observed could be due to existing baseline differences, or they could be due to early effects of ERF. A description of the characteristics of the teachers and of significant differences between teachers in ERF-funded and unfunded classrooms is important in determining whether ERF might have influenced any factors that could impact the outcomes for children.

Several correlational studies indicate that higher levels of teacher education are associated with teacher quality and child outcomes.[39] The research linking teachers' level of education to classroom quality is not consistent, and causal inferences cannot be drawn, given the correlational nature of these studies.[40] Within the context of this literature, it is useful to document the educational level of ERF teachers. Three-quarters the teachers in ERF preschools had earned bachelor's degrees, and an additional 12 percent held associate's degrees (see Table 4.6).[41] Teachers in ERF preschools had much more formal education than Head Start teachers in the FACES 2000 sample, in which approximately 25 percent of the staff who provided instruction in the classroom (administrative teachers and classroom teachers) had bachelor's degrees.[42]

The largest percentage of ERF teachers held degrees in early-childhood education (38 percent), followed by elementary education (22 percent), and education (10 percent). Among teachers in ERF classrooms, 87 percent have completed college-level courses in early-childhood education or development, 67 percent have completed courses in teaching reading to elementary-school children, and 79 percent have completed courses in teaching language and literacy skills to children in a preschool setting.

In addition, 30 percent of the teachers in the ERF sites held a child-development associate credential, 42 percent held a state-awarded preschool certificate, 67 percent held a teaching certificate or license, and 24 percent held other types of job-related licenses. Finally, 42 percent of the ERF teachers in the sample were currently enrolled in teacher-related training.

Compared to teachers in non-ERF classrooms, more teachers in ERF classrooms had earned bachelor's degrees, held teaching certificates or licenses, and were currently enrolled in teacher-related training or education. We cannot definitively determine which of these differences preceded ERF funding and which were a direct result of the grant. It is unlikely that ERF

[39] Barnett, W.S. (2004). "Better teachers, better preschools: Student achievement linked to teacher qualifications." In *Preschool Policy Matters (2)*. New Brunswick, NJ: National Institute for Early Education Research.
[40] Early, D., Bryant, D., Pianta, R., Clifford, R., Burchinal, M., Ritchie, S., Howes, C., and Barbarin, O. (2006). "Are teachers education, major, and credentials related to classroom quality and children's academic gains in pre-kindergarten?" *Early Childhood Research Quarterly*, *21*, 175–195.
[41] These results were reported by teachers in a survey and were not independently verified.
[42] U.S. Department of Health and Human Services (January 2002), *A Descriptive Study of Head Start Families: FACES Technical Report I*, January 2002, p. 206.

influenced the attainment of bachelor's degrees or teaching certificates, because the ERF funding had not been available for a sufficiently long period of time for the teachers to have obtained the credentials under the auspices of ERF funding.

Table 4.6. Educational background of teachers and others, by ERF funding status

	All teachers	ERF teachers	Non-ERF teachers	P-value[1]
Highest degree				< 0.01
High-school diploma	4.1%	4.3%	3.9%	
Vocational- or technical-school diploma	1.0%	0.0%	2.0%	
Some college, no degree	13.4%	8.7%	17.6%	
Associate's degree	16.0%	12.0%	19.6%	
Bachelor's degree	37.1%	45.7%	29.4%	
Graduate or professional school, no degree	8.2%	14.1%	2.9%	
Master's or law degree	21.1%	15.2%	24.5%	
Field in which highest degree was earned				0.14
Child development / developmental psychology	6.2%	4.3%	7.8%	
Early-childhood education	33.0%	38.0%	28.4%	
Elementary education	20.1%	21.7%	18.6%	
Education, other	9.3%	9.8%	8.8%	
Psychology, other	2.1%	3.3%	1.0%	
Social sciences, liberal arts, languages	5.7%	7.6%	3.9%	
Business administration, management	4.1%	1.1%	6.9%	
Professional	1.0%	1.1%	1.0%	
No degree	18.6%	13.0%	23.5%	
Completed 6 or more college courses in relevant fields:				
Early childhood education or development	85.6%	87.0%	84.3%	0.60
Teaching reading to elementary school children	65.5%	67.4%	63.7%	0.59
Teaching language and literacy skills to children in a preschool setting	73.7%	79.3%	68.6%	0.09
Earned a credential, certificate, or license				
Child Development Associate (CDA) credential	33.5%	30.4%	36.3%	0.39
State-awarded preschool certificate	43.3%	42.4%	44.1%	0.81
Teaching certificate or license	58.8%	67.4%	51.0%	0.02
Other job-related licenses	20.1%	23.9%	16.7%	0.21
None of the above	16.5%	12.0%	20.6%	0.11
Sample Size	194	92	102	

[1] All p-values are based on chi-squared tests of association.
SOURCE: Spring surveys of preschool teachers.

As shown in Table 4.7, the overwhelming majority (97 percent) of ERF teachers are women. They range in age from 23 to 67 years; the average teacher is 41 years old. The largest percentage of the ERF teachers are white (54 percent), and fewer than a quarter are either Hispanic (23 percent) or black (17 percent). Although the majority of teachers (73 percent) are

monolingual English speakers, a sizeable proportion (21 percent) reported being fluent in both Spanish and English. These numbers are important to keep in mind in light of the findings reported in Chapter 3 that over 43 percent of the overall sample of children are Hispanic. We did not observe any statistically significant differences in demographic characteristics between teachers in the funded sites and those in the unfunded sites.

Table 4.7. Demographic characteristics of teachers, by ERF funding status

Characteristic	All teachers	ERF teachers	Non-ERF teachers	P-value
Gender				
Female	95.9%	96.7%	95.1%	0.57
Age				
20 through 29 years	19.9%	22.2%	17.8%	
30 through 39 years	23.6%	21.1%	25.7%	
40 through 49 years	29.8%	36.7%	23.8%	
50 through 59 years	18.8%	13.3%	23.8%	
60 and older	7.9%	6.7%	8.9%	
Mean (SD)	41.6 (11.3%)	40.8 (10.9%)	42.4 (11.6%)	0.34[1]
Range (years)	23 to 67	23 to 67	23 to 64	
Ethnicity				
American Indian or Alaska Native	3.1%	3.3%	3.0%	
Asian	1.6%	2.2%	1.0%	
Non-Hispanic black or African American	21.8%	17.4%	25.7%	
Native Hawaiian or Pacific Islander	0.0%	0.0%	0.0%	
Non-Hispanic white	51.3%	54.3%	48.5%	
Hispanic	22.3%	22.8%	21.8%	0.68
Missing	0.5%	0.0%	0.9%	
Languages spoken fluently				
English only	74.7%	72.8%	76.5%	
Spanish only	2.1%	3.3%	1.0%	
English and Spanish	20.6%	20.7%	20.6%	
English and other	2.6%	3.3%	2.0%	0.65
Sample Size	194	92	102	

[1] This p-value is based on Student's t-tests; all other p-values are based on chi-squared test of association.
SOURCE: Spring surveys of preschool teachers.

Chapter 5. Professional Development, Instructional Practices, and Classroom Environments in ERF Preschools

To meet the goals of Early Reading First, grantees are expected to create high-quality oral-language and literature-rich classroom environments that offer activities and instructional materials to develop children's oral language, phonological awareness, print awareness, and alphabetic knowledge. ERF funds were awarded in October 2003, and grantees were expected to fully implement programs by January 2004. Accordingly, both the fall 2004 and spring 2005 data collections measure the professional development activities, curriculum and assessment choices, classroom materials, and instructional practices of fully implemented ERF programs.

In this chapter, we describe teachers' professional development activities and the curriculum and assessment choices that are intended to help support the quality of the classroom environments in terms of organization, interactions, language, and early literacy instruction. We also describe the characteristics of ERF preschool classrooms associated with dimensions of interest (classroom organization, variety of activities, and supportive teacher-child interactions) to early-childhood professionals. We describe the preschool classrooms in terms of observed teacher instruction and available classroom materials associated with the goals of ERF: the classroom language environment and the opportunities for developing early literacy skills.[43] The impacts of ERF are presented in Chapters 6 and 7.

Professional Development Experiences

ERF grantees were required by statute to provide professional development. In its guidance to ERF grantees, ED recommended in accordance with the statutory definition of the term (section 9101(34), ESEA) that professional development be ongoing, sustained, intensive, and classroom focused. ED policy guidance lists mentoring or coaching as examples of professional development methods based on scientifically-based reading research (U.S. Department of Education, 2003).

ERF teachers reported receiving an average of 72 hours of professional development during the previous year—the equivalent of 9 days (see Table 5.1).

Table 5.1. Hours of professional development in language and literacy topics received in the past 12 months, by ERF teachers

Hours of training	
Median	55.0
Mean	71.5
Standard deviation	84.7
Sample size	86

SOURCE: Spring teacher surveys.

[43] For the interested reader, Appendix G provides descriptive tables comparing the funded and unfunded classrooms on the variables discussed in this chapter.

One hundred percent of teachers in ERF-funded classrooms reported receiving professional development in phonemic and phonological awareness. The vast majority of ERF teachers received training in six other language-development and early literacy topics, including literacy-rich print environments (97.8 percent), concepts of print writing and prewriting (96.7 percent), oral language (96.7 percent), facilitating emergent literacy (95.7 percent), alphabetic knowledge (92.4 percent), and oral comprehension and cognition (88.0) (see Table 5.2). Nine out of 10 ERF teachers reported receiving training in child assessment. Three-fourths of ERF teachers reported receiving training in traditional early-childhood topics, including children's development and ways to manage children's behavior in the classroom. Most ERF teachers (77 percent) reported receiving training on 9 or 10 professional development topics that were included in the list.

Table 5.2. Topics in which ERF teachers received professional development in the past 12 months

Topic areas	% ERF teachers who received training in topic
Language Development and Early Literacy	
Phonemic & phonological awareness	100.0
Literacy-rich environments	97.8
Concepts of print writing & prewriting	96.7
Oral language	96.7
Facilitating emergent literacy	95.7
Alphabetic knowledge	92.4
Oral comprehension & cognition	88.0
Child Assessment	
Child Development and Behavior	90.2
Early childhood growth & development	76.1
Classroom management	76.1
Other Topics	56.5
Number of topics	**% ERF teachers who received training in number of topics**
0	0.0
1 to 4	1.1
5 to 8	21.7
9 or 10	77.2
Mean # of topics (SD)	9.6 (1.7)
Sample Size	92

SOURCE: Spring teacher surveys.

ERF teachers reported that most of the professional development topics on which they received training were covered through in-service training (see Table 5.3). Teachers potentially could have received professional development training in 11 areas, including topics that fell under the "other" category. In-service training covered an average of 7.6 out of 11 topics. Several topics were also covered by mentoring or tutoring (4.7 out of 11 topics) and by workshops (4.5 out of 11 topics). While these patterns reflect the flexibility of each training method in covering a variety of topics, it may not reflect the relative number of hours teachers participated in each type of training. We did not ask teachers how their professional development hours were distributed across the various types of training.

Table 5.3. Mean number of professional development topics for ERF Teachers, by method of training

Training method	Mean number of topics (SD)
In-service	7.60 (3.48)
Mentor or tutor	4.73 (4.54)
Workshops	4.52 (4.42)
Continuing education courses	2.48 (4.00)
National meetings	1.20 (2.81)
Other	0.55 (1.76)
Sample Size	92

SOURCE: Spring teacher surveys.

Formal education was a substantial source of professional development for ERF teachers. ERF teachers reported that they received training on an average of 2.5 topics through continuing-education courses. More than 40 percent of ERF teachers reported taking courses toward certification or degree programs in the past year (see Table 5.4). Many (17 percent) ERF teachers were working toward a graduate degree.

Table 5.4. Current ERF teacher enrollment in formal education

	% of ERF teachers currently enrolled
Any teacher-related training or education	42.4
Type of formal education	
Child development associate (CDA)	4.3
Teaching certificate program	2.2
Special education teaching degree	0.0
Associate's degree	0.0
Bachelor's degree	5.4
Graduate degree	17.4
Other	13.0
Sample size	92

SOURCE: Spring teacher surveys.

ERF teachers' professional development activities were funded by a variety of sources (see Table 5.5). Teachers in nearly all of the ERF programs received training funded by ERF on multiple topics. Except for ERF funds, school district and Head Start funds were the most widely used sources for teachers in ERF programs, paying for training of 56.5 percent and 31.5 percent of ERF teachers, respectively. This is consistent with the finding in Chapter 4 that many preschools in the sample received state or local education funding or Head Start funding (or both). Notably, approximately 1 in 10 teachers paid for his or her own professional development on at least one of the topics. Because we do not know how the hours of professional development activities were covered by various funding sources, this descriptive analysis cannot assess the extent to which ERF might have contributed to the professional development hours reported by teachers. We address the question of how ERF influenced teachers' professional development in the impact analysis in Chapter 6.

Table 5.5. Sources of funding for professional development for ERF teachers, by number of topics

Funding source	% ERF teachers receiving training on topics thru funding source
ERF	
No topics	17.4
One topic	0.0
Multiple topics	82.6
School district	
No topics	43.5
One topic	6.5
Multiple topics	50.0
Head Start	
No topics	68.5
One topic	4.3
Multiple topics	27.2
State	
No topics	80.4
One topic	2.2
Multiple topics	17.4
Teacher (self)	
No topics	87.0
One topic	4.3
Multiple topics	8.7
Other	
No topics	82.6
One topic	10.9
Multiple topics	6.5
Sample Size	92

SOURCE: Spring teacher surveys.

Curricula and Assessment Practices

The statute requires ERF grantees to identify and provide activities and instructional materials that are designed according to scientifically based reading research for developing children's oral language, phonological awareness, print awareness and alphabet knowledge. ERF programs are also required to use assessments to monitor children's attainment of skills and to guide instruction.[44] ERF programs are expected to integrate assessments of child progress with teaching so that instruction can build on what children already know and bring them to the next level (U.S. Department of Education 2003.) Accordingly, the choice of assessments is important in providing critical information about children's progress and about useful next steps in supporting their learning. The following section describes curricula and assessment instruments used in the ERF classrooms.

[44] U.S. Department of Education. *Guidance for the Early Reading First Program*. Washington, DC, March 2003, p. 5.

Curricula Used by Teachers

Recommendations for the practice of early-childhood education call for a classroom curriculum that articulates learning objectives and that teachers can use to plan daily activities for preschool-age children throughout the year.[45] A widely used set of professional guidelines recommends choosing a curriculum that is consistent with the program's goals for children's development across the cognitive, language, social, emotional, and physical domains.[46]

Guidance from ED recommended that ERF teachers "organize and present instructional materials in a systematic and coherent manner." ED's guidance specified that curricula should be "intellectually engaging, have meaningful content, and provide multiple opportunities for developing phonological awareness, print awareness, oral-language skills, and alphabet knowledge, including the use of explicit, contextualized, and scaffolded instruction."[47] In their grant applications, some grantees explicitly said that they sought ERF funding to support the purchase and implementation of a new curriculum designed according to scientifically based reading research, either as a replacement or a supplement for a curriculum that they were already using. The legislation that authorized ERF and the written guidance from ED to ERF grantees do not recommend particular curricula.[48]

All ERF teachers reported using a curriculum (see Table 5.6). In ERF preschool classrooms, 39 percent of the teachers reported following one curriculum, and 61 percent reported using a combination of curricula.

Table 5.6. Number of curricula used by ERF teachers

% ERF teachers using	
A single curriculum	39.1
A combination of curricula	60.9
No curriculum	0.0
Average number of curricula used (SD)	1.88 (1.00)
Sample Size	92

SOURCE: Spring teacher surveys.

Most ERF teachers used the Creative Curriculum or the High/Scope (Educating Young Children) curriculum (see Table 5.7). Roughly 46 percent of the teachers used the Creative Curriculum; 24 percent used the High/Scope curriculum. The widespread use of these two curricula is consistent with reported curriculum choices among a nationally representative sample of Head

[45] For example, Head Start Program Performance Standards require that programs have a curriculum, but do not prescribe one. (*Head Start FACES 2000: A Whole-Child Perspective on Program Performance*, Fourth Progress Report. U.S. Department of Health and Human Services, Washington, DC, May 2003). In addition, non-regulatory guidance for Title I preschools recommends that the preschools use a curriculum. (*Serving Children Under Title I: Non-Regulatory Guidance*. U.S. Department of Education Washington, DC, March 2004.)

[46] *NAEYC Early Childhood Program Standards and Accreditation Criteria: The Mark of Quality in Early Childhood Education*. Washington, DC: National Association for the Education of Young Children (NAEYC), 2005.

[47] U.S. Department of Education March 2003, p. 9.

[48] No Child Left Behind Act of 2001, Sections 1221 and 1222 and U.S. Department of Education, March 2003.

Start programs. In the Head Start Family and Child Experiences Study (FACES) 2000 cohort, 59 percent of Head Start teachers reported using either the Creative Curriculum or High/Scope.[49]

For language and early literacy, each of four curricula was used by more than 10 percent of the teachers in ERF programs: Building Language for Literacy (an online early literacy activity site designed for children to use); Doors to Discovery (curriculum and materials to foster language and early literacy); Let's Begin with the Letter People (a language and literacy curriculum with materials that include "letter people"), and Opening the World of Learning (a curriculum with books, songs, and poetry to foster language and literacy).

Table 5.7. Curricula used by ERF teachers

Curriculum	% of ERF teachers using
Creative Curriculum	45.7
High/Scope (Educating Young Children)	23.9
Building Language for Literacy	16.3
Doors to Discovery	15.2
Let's Begin with the Letter People	15.2
Opening the World of Learning	12.0
We Can!	8.7
DLM Early Childhood Express	7.6
Breakthrough to Literacy	6.5
Creating Child-Centered Classrooms	4.3
Scholastic Curriculum	3.3
CIRCLE	3.2
SRA Open Court Reading	2.2
Montessori	2.2
High Reach Learning	0.0
Other	21.7
Sample Size	92

NOTE: Percentages exceed 100 because teachers may be using multiple curricula. "Other" includes all curricula reported by four or fewer teachers.
SOURCE: Spring teacher surveys.

Assessment Usage

The statute requires ERF programs to acquire, provide training on, and use screening assessments or other appropriate measures designed according to scientifically based reading research to determine whether preschool age children are developing the cognitive skills they need for later reading success. ED's guidance reiterates that requirement and states that teachers are expected to be trained on using the assessments and to use the assessments to tailor a plan of instruction to the needs of individual children.[50] ED did not require the FY 2003 grantees to use any specific child assessment tools.[51]

[49] U.S. Department of Health and Human Services (2003), *Head Start FACES 2000: A Whole-Child Perspective on Program Performance*, Fourth Progress Report.
[50] U.S. Department of Education (2003), *Guidance for the Early Reading First Program*, p. 9.
[51] Early Reading First 2005 and 2006 Performance Plans (U.S. Department of Education 2004 and 2005), accessed at http://www.ed.gov/about/reports/annual/2006plan/edlite-g2eseaearlyread.html *...Footnote continued on page 40.*

Nearly all ERF teachers (97.8 percent) reported using at least one assessment tool for children in their classes, reflecting the current interest in at least screening children's developmental progress during the preschool year (see Table 5.8). Since the Head Start program's reauthorization in 1998, teachers have been required to assess all children in their classes (using tools of their choice) on a broad range of outcomes and to use the information to plan instruction. Many curricula, including the two most widely used curricula, include assessment tools that reflect the curriculum's learning goals. Results of these assessments are intended to help teachers tailor the curriculum and instruction to children's developmental levels.

Table 5.8. Number of assessments used by ERF Teachers

	% of ERF teachers using
Assessments per classroom:	
No assessment	2.2
Single assessment	33.7
Combination assessments	64.1
Mean (SD)	2.11 (1.21)
Sample Size	92

SOURCE: Spring teacher surveys.

A majority of ERF teachers (64 percent) reported using more than one assessment instrument with children in their classes. Among the most commonly used were the assessment tools associated with the two most widely used curricula; 26 percent of teachers used the Child Observation Record (the assessment tool accompanying the High/Scope curriculum), and 22 percent used the Creative Curriculum Continuum (the assessment tool accompanying the Creative Curriculum) (see Table 5.9).

Substantial percentages of ERF teachers reported using several other assessment tools, including those that focus specifically on language and early literacy skills. The Peabody Picture Vocabulary Test (used by 34 percent of teachers) is a vocabulary assessment with national norms to help interpret children's progress over the course of the year. The Preschool Individual Growth & Development Inventory (used by 22 percent of teachers) measures language through picture naming and measures phonemic awareness through rhyming and alliteration. The Phonological Awareness Literacy Screening—Pre-K (used by 17 percent of teachers) focuses on alphabet knowledge, beginning sounds, print and word awareness, and rhyme awareness. The Teacher Rating of Oral Language and Literacy (TROLL) (used by 12 percent of teachers) rates the child's language use, early reading, and early writing skills. The Work Sampling System (used by 12 percent of teachers) uses observational checklists, portfolios, and teacher and parent summaries to assess the child's development across the full range of outcome domains. The Desired Results assessment (used by nearly 10 percent of teachers) has been under development for the California Department of Education to assess progress toward preschool-learning guidelines across all developmental domains.

and http://www.ed.gov/about/reports/annual/2005plan/edlite-esea-earlyread.html. The two most recent cohorts of grantees, FY 2005 and FY 2006, must use two child assessments for the purpose of GPRA reporting: the PPVT and the Phonological Awareness Literacy Screenings (PALS) Pre-K.

Table 5.9. Instruments used to assess children's progress and needs within the previous 30 days

Assessment Instrument	% of ERF teachers using
Peabody Picture Vocabulary Test	33.7
Child Observation Record	26.1
Creative Curriculum Continuum	21.7
Preschool Individual Growth & Development Inventory	21.7
Phonological Awareness Literacy Screening	17.4
Teacher Rating of Oral Language & Literacy	12.0
Work Sampling	12.0
Desired Results	9.8
Brigance Inventory of Early Development	6.5
Learning Accomplishment Profile—Diagnostic (LAP-D)	4.3
State- or School District-designed	4.3
Galileo	2.2
Expressive One Word Picture Vocabulary Test	0.9
Get Ready to Read	0.0
Other[1]	28.3
Sample Size	92

[1] "Other" includes all assessments reported by four or fewer teachers.
SOURCE: Spring teacher surveys.

Classroom Environments and Teacher Practices

In this section, we describe the classroom-learning environments, including the materials and physical organization of the classroom, the teacher's interactions with children, and the range and quality of instruction about early literacy topics.

Two perspectives on the classroom environment can inform our picture of the quality of ERF classrooms as environments for fostering children's language development and early literacy skills. First, research shows that some characteristics of preschools classrooms are positively correlated with child outcomes (Vandell and Wolfe 2000; NICHD Early Childhood Research Network 2002, 2003, and 2006). Given its correlational nature, this research is not conclusive. Second, ERF requires grantees to provide the types of materials, learning opportunities, and instruction that are intended to support the development of children's language and early literacy skills. ERF also requires regular progress assessments to gauge children's learning. Accordingly, our measures of teacher instructional practice focused on both the general quality of the preschool environment and on the language, early literacy, and assessment practices that are intended to support children's development of language and early literacy skills.

We obtained measures of the classroom environment and instructional practices through direct observation of the classroom and teacher. We completed observations of up to three classrooms per site in the fall and spring. The observation protocols included the Teacher Behavior Rating Scale (TBRS), developed by the Center for Improving the Readiness of Children for Learning and Education (CIRCLE) at the University of Texas-Houston (Landry et al. 2004), and a subset of items from the Early Childhood Environment Rating Scale-Revised (ECERS-R) (Harms,

Clifford, & Cryer 1998).[52] The TBRS was developed to evaluate the early literacy and language qualities in preschool classrooms, but it also includes subscales that measure the general quality of the classroom and the sensitivity of teacher behavior. We selected 11 ECERS-R items that compose the subscale, Teaching and Interactions, on the basis of a previous factor analysis of the instrument (Clifford, Barbarin, et al. 2005), which produced a single score focused on the quality of teaching and interactions in the classroom environment. The full ECERS-R score has been found to be correlated with children's cognitive and emotional outcomes in early childhood settings, although no causal inference can be drawn from these correlational studies (Vandell and Wolfe 2000).

General Quality of the Preschool Classroom

The ECERS-R and the TBRS provided measures of several aspects of the general quality of the preschool environment. The quality of teacher-child interactions refers to the teacher's responsiveness to children, sensitivity to children's needs, consistent, positive guidance, and encouragement. To measure teacher-child interactions, we used the Teaching and Interactions subscale of the ECERS-R (Clifford et al. 2005) and the Teacher Sensitivity subscale from the TBRS (Landry et al. 2004). We also measured the quality of the assistant teacher-child interactions through the TBRS Team Teaching subscale.

The ECERS-R scores each item on a scale ranging from 1 ("inadequate") to 7 ("excellent"). ECERS-R Teaching and Interactions subscale scores averaged 5.7 for the funded classrooms; a score of 5 on the ECERS-R is considered to be "good." Scores on the Teaching and Interactions subscale tend to be higher than scores on the full ECERS-R scale. For example, in spring 2001, Head Start classrooms in the FACES 2000 cohort sample scored an average of 5.5 on the Teaching and Interactions subscale but 4.9 on the full ECERS-R scale.[53]

Table 5.10. General quality of ERF classrooms, based on ECERS-R and TBRS subscales

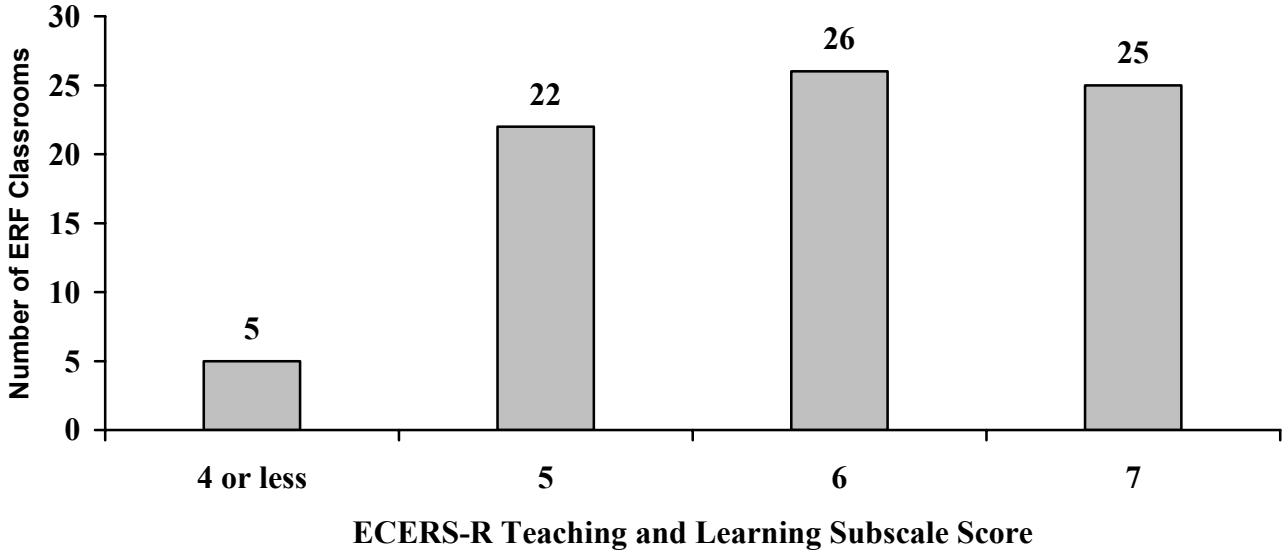
	Mean (SD)		
	Fall	Spring	Diff.
ECERS-R Teaching and Interactions Subscale score	5.67 (1.07)	5.78 (1.03)	+0.12
General teaching behavior	3.14 (0.56)	3.14 (0.52)	–0.00
Classroom community	3.18 (0.59)	3.19 (0.56)	+0.01
Teacher sensitivity	3.11 (0.68)	3.07 (0.62)	–0.04
Lesson planning	3.06 (0.81)	3.05 (0.90)	–0.01
Quality and organization of activity centers	3.12 (0.67)	2.93 (0.73)	–0.19
Quality of team teaching	2.98 (0.83)	2.99 (0.88)	+0.01
Math concepts	2.33 (1.04)	2.35 (1.01)	+0.02
Total TBRS Score	2.71 (0.61)	2.65 (0.65)	–0.06
Sample size	78	78	

SOURCE: Fall and spring classroom observations.

[52] Appendix C provides details on the contents and psychometric properties of the TBRS and ECERS-R.
[53] Authors' calculations using subscale-level ECERS data from the FACES 2000 Cohort microdata (U.S. Department of Health and Human Services, 2005).

The average score on the ECERS-R Teaching and Interactions subscale in the spring was 5.8 for ERF classrooms (slightly higher than in the fall) with all but 5 classrooms scoring at least a "good" or 5 on the subscale (see Figure 5.1).

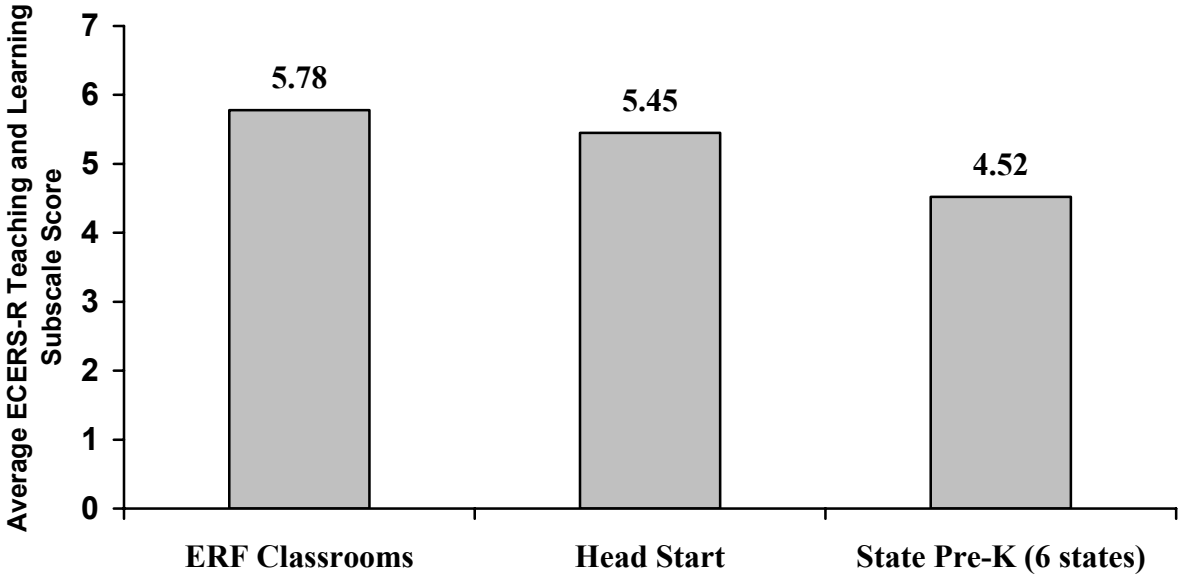
Figure 5.1. Number of ERF classrooms by ECERS-R Teaching and Interactions Subscale, spring 2005



ERF classrooms have similar general quality to Head Start classrooms and better general quality than state pre-kindergarten classrooms (see Figure 5.2). The average score on the ECERS-R Teaching and Interactions subscale for ERF classrooms is similar to those of Head Start classrooms, according to data for the 2000 FACES cohort. Although the means for the ERF funded classrooms look higher, the differences between those means and that for Head Start are not statistically significant.[54] Data for a national sample of state pre-kindergarten programs have not been gathered as they have for Head Start, but a recent study of pre-kindergarten programs in six states found significantly lower ECERS-R Teaching and Interactions scores among classrooms in the study than was found among ERF classrooms (Clifford et al. 2005).[55]

[54] Head Start data are from authors' calculations using subscale-level ECERS data from the FACES 2000 Cohort microdata (U.S. Department of Health and Human Services, 2005).
[55] States included in the study are Georgia, Illinois, Kentucky, Ohio, California, and New York.

Figure 5.2. Average ECERS-R Teaching and Interactions Subscale Score, ERF, Head Start, and state pre-kindergarten classrooms



The TBRS measures several aspects of the general quality of preschool classrooms. The TBRS items are scaled so that higher values represent greater frequency or quality or both, using Likert ratings that range from 1 (low or none) to 4 (high frequency/high quality) for virtually all of the items. Because of a high correlation between quantity and quality item scores, we have averaged them to create a single-item score and created subscales from these composite items.[56]

The average score for General Teaching Behavior, which includes the subscales Classroom Community and Teacher Sensitivity, was 3.1 out of 4 among ERF classrooms in the fall (see Table 5.10). Classroom Community measures the degree to which teachers have established classroom routines for children that help to maintain a calm, orderly, and busy atmosphere throughout the preschool day. Teacher Sensitivity refers to the teacher's responsiveness and emotional supportiveness toward children. The average score for General Teaching Behavior was nearly the same in the fall and spring for ERF classrooms.

Teachers can help to maintain classroom order and prevent conflict by organizing the physical environment. To measure the extent to which teachers have organized the physical environment of the classroom into interesting, diverse, and well-placed activity centers, we used the Quality and Organization of Activity Centers subscale of the TBRS measure. The average score for the Activity Centers subscale among ERF classrooms was 3.1 out of a possible 4 in the fall and 2.9 in the spring. To measure the extent to which teachers plan a variety of learning activities and follow through with their plans, we used Lesson Planning, another subscale of the TBRS. ERF classrooms scored an average of 3.1 out of 4 in the fall and spring.

[56] Appendix C contains additional information about the TBRS subscales used in the ERF evaluation.

Nearly all preschool classrooms are taught by a lead and assistant teacher. The assistant teacher ideally does more than provide an extra pair of hands to help keep order in the classroom. By acting as a knowledgeable teaching-team member, the assistant teacher can extend the guidance, teaching, and emotional support provided by the lead teacher. The assistant teacher can help enrich the classroom language environment and keep learning activities going in a small group after the lead teacher has moved on to another group. The TBRS Team Teaching subscale measures the assistant teacher's contributions to the language and learning environment of the classroom. ERF classrooms scored an average of 3.0 on the Team Teaching scale in both the fall and spring.

Math Concepts is a short, 2-item subscale of the TBRS that measures the extent to which the teacher incorporates mathematics concepts and activities into the preschool day. Early mathematics skills were not a focus of ERF, and they have not received much attention from early-childhood professionals. Nevertheless, because the subscale is a component of the TBRS, we include it here for completeness. ERF classrooms scored an average of 2.3 on this scale in the fall. In the spring, the average score for ERF classrooms was similar to the fall score.

Classroom Language and Early Literacy Environment

Several measures of the language and early literacy aspects of teacher instructional practices and the available classroom materials are available from the TBRS. Table 5.11 shows the fall and spring scores for ERF classrooms for key subscales of the TBRS that measure the language environment, early literacy materials and instruction, and child assessment.

Table 5.11. Classroom language and early literacy environment in ERF classrooms

	Mean (SD)		
Subscales	Fall	Spring	Difference
Oral Language Use by Lead Teacher	2.99 (0.75)	2.88 (0.71)	–0.11
Book-Reading Practices	2.34 (0.90)	2.40 (0.83)	+0.07
Phonological Awareness Activities	2.25 (0.88)	2.05 (1.00)	–0.20
Print and Letter Knowledge	2.32 (0.78)	2.14 (0.83)	–0.18
Written Expression	2.47 (0.78)	2.28 (0.91)	–0.19
Child Portfolios	2.79 (1.63)	2.82 (1.47)	+0.03
Dynamic Assessment	2.84 (1.07)	2.786 (1.13)	–0.05
Sample size	78	78	

SOURCE: Fall and spring classroom observations.

A high-quality language environment that includes exposure to new vocabulary, adults modeling more complex sentences for children, and encouragement of children's expression can help children to expand their vocabulary. A wider vocabulary can help children understand the information they hear in the classroom and recognize words that they sound out as they begin to read (Whitehurst and Lonigan 2001). Oral Language Use measures the language environment provided by the lead teacher in the classroom. ERF classrooms scored 3.0 out of a possible 4 on the Oral Language Use subscale in the fall and 2.9 in the spring.

Book reading in preschool classrooms provides an appealing and flexible foundation for teaching a wide range of language and literacy skills to children. Teachers can use a book-reading session to explain new vocabulary words, teach concepts of print, expose children to the sounds and rhythms of language, and encourage children to express their thoughts and comprehend oral expression. These features of a good-quality book-reading session are all measured by items in the Book-Reading Practices subscale of the TBRS. The average Book-Reading score for ERF classrooms was 2.3 in the fall and 2.4 in the spring out of a possible 4.

To better understand how classrooms performed with respect to the activities associated with book reading, see Table 5.12, which shows average scores for several items that compose the Book-Reading scale.[57]

Table 5.12. Book reading and associated activities in ERF classrooms, fall and spring

	Mean (SD)	
Book-Reading Activity	Fall	Spring
Number of books read during the observation	1.65 (1.09)	1.45 (1.00)
Number of book features discussed (title, author, illustrator, cover)	2.06 (1.01)	2.38 (1.11)
Frequency of introducing and discussing vocabulary words before and during book reading	2.12 (1.15)	2.32 (1.17)
Quality of teacher's use of facial expressions and voice to capture children's attention	2.77 (1.37)	2.79 (1.09)
Quantity and quality of open-ended questions asked to encourage discussion of book	2.59 (1.26)	2.55 (1.23)
Quantity and quality of activities or discussions that extend book reading	2.12 (1.22)	1.78 (1.27)
Sample Size	78	78

SOURCE: Fall and spring classroom observations.

ERF teachers typically read one or more books during the 3-hour observation period. Teachers typically drew children's attention to and discussed two features of the book during book reading—for example, the title, author, or illustrator. Teachers did not consistently use the book-reading session as a springboard for vocabulary or to ask open-ended questions. ERF teachers scored an average of 2.32 on frequency of vocabulary words in the spring, corresponding to "rarely" or "sometimes" introducing new words. Results were similar for the item measuring the frequency of open-ended questions and the extent to which children were permitted time to express their ideas in response. Teachers in ERF classrooms consistently used facial expressions and voice to capture children's attention during book reading. The average score of 2.79 in the spring corresponds to "medium high" quality of this aspect of the book-reading session. Finally, the score for frequency and quality of activities and discussions to extend the book reading (1.78) is in the low- to medium-range, meaning that teachers typically offered at least one activity or discussion to extend the book reading, but the average quality of the extension was low to medium.[58]

[57] Appendix C contains additional information on the Book-Reading scale and the other subscales that comprise the TBRS.

[58] The correlation between quality and quantity of the book-reading extensions items is .94; therefore, the combined quantity and quality score closely reflects the individual scores.

Phonological-awareness activities provide opportunities for children to learn word and letter sounds, which are fundamental skills needed for reading. The TBRS provides indicators of whether the teacher introduced or discussed any of seven phonological awareness activities: listening (to sounds generally or to sounds in spoken words), rhyming, alliteration, sentence segmenting (clap for each word in a sentence or rearrange word cards), onset-rime blending and segmenting (teaching initial consonant sounds by using simple rhyming words as in "bat" and "cat"), syllable blending or segmenting (calling attention to each syllable in a word), and phoneme blending, segmenting, and manipulation (calling attention to each separate sound in a word). Table 5.13 shows the proportion of classrooms in the fall and spring in which each phonological awareness activity was observed.

Table 5.13. Phonological awareness activities in ERF classrooms fall and spring

	Observation time	
Phonological Awareness Activity	Fall	Spring
Activity observed:	% of classrooms where activity observed	
Rhyming (identifying words with the same ending sound)	47.4	64.1
Listening (teacher draws attention to environmental sounds)	52.6	39.7
Alliteration (note initial sounds in words (e.g. lazy lizard lounging))	43.6	32.1
Onset-rime blending and segmenting (working with words that share sounds and varying the first letter or sound—c-at, b-at)	25.6	26.9
Phoneme blending, segmenting and manipulation (isolate sounds in words and replace with other sounds)	25.6	26.9
Sentence segmenting (clapping for each word in a sentence, deleting words in a sentence, using word cards)	25.6	12.8
Syllable blending and segmenting (clapping for each syllable, deleting syllables)	16.7	21.8
Average number of different activities observed	2.4	2.2
Sample Size	78	78

SOURCE: Fall and spring classroom observations.

Rhyming was the most common activity in the spring, and was observed in 64 percent of the classrooms. Listening and alliteration activities were observed in 40 percent and 32 percent of classrooms in the spring, respectively. Other more challenging phonological-awareness activities, such as blending and segmenting words, syllables, initial sounds, and phonemes, were observed in 27 percent or fewer ERF classrooms. We observed an average of 2.2 different phonological-awareness activities during the spring visit to ERF classrooms.

The quality of the phonological awareness activities is measured by the degree to which children seem engaged in the activity. The average score for quantity and quality of Phonological Awareness Activities combines the number of different activities observed, the number of different classroom contexts where those activities were observed, and the level of children's engagement in the activity. ERF classrooms had similar scores on this subscale in the fall (2.2) and spring (2.0).

Knowledge of print and letters is another skill needed for reading. The TBRS Print and Letter Knowledge subscale taps the frequency and level of children's engagement in print and letter-learning opportunities, which include instances when the teacher discusses concepts about print; associates letters with their picture, name, shape, and sound; and talks about contrasting sounds and meanings of words, rhyming words, and uppercase and lowercase letters. This subscale also measures the classroom print environment, which includes theme- and topic-related books available in the classroom, charts, posters, and labels on materials in activity centers and around the classroom, and a complete letter wall, showing pictures with printed words for each letter of the alphabet (to support teaching the names and sounds of letters). The average score for Print and Letter Knowledge in the spring was 2.1 for ERF classrooms (reflecting some observed learning opportunities at medium quality, on average).

Providing children with opportunities for writing and showing them how to write letters can help children's letter-recognition skills and help them to understand that writing and reading are complementary literacy activities. The Written Expression subscale measures the extent to which teachers provide learning opportunities that model writing and provide materials for writing in the classroom. ERF classrooms scored an average of 2.3 on this subscale in the spring, reflecting that some learning opportunities and materials of average quality and variety were observed during the visit.

ERF requires programs to assess children's progress in language development and literacy skills so that instruction can build more effectively on what children have learned and help them progress to the next level. TBRS subscales, Child Portfolios and Dynamic Assessment, measure the extensiveness, completeness, and recency of progress assessments and samples of children's work. ERF classrooms scored an average of 2.8 in the spring on the Portfolios subscale, meaning that over half of children's portfolios contained at least one work sample and an anecdotal teacher note. On Dynamic Assessment, ERF classrooms scored an average of 2.8 in the spring. Fewer than half of the classrooms had recent (within 30 days) documentation of children's developmental progress across a range of emergent literacy areas, while more than half of the teachers said that they plan for instruction on the basis of children's assessments and could identify an average of two ways in which they use results from child assessments.

The total TBRS score summarizes all of the TBRS general quality and language, literacy, and assessment subscales described in this chapter and reported in Tables 5.10 and 5.11. The average TBRS total score was 2.7 in the fall and 2.6 in the spring (see Table 5.10).

Chapter 6. Impacts on Teachers and Classroom Practices

The Early Reading First (ERF) program provides funding to preschools to improve classroom environments and teacher practices particularly to help economically disadvantaged preschool children develop language and early literacy skills. To support development of these skills, ERF grantees are required to use the funds to provide:

- Professional development (according to scientifically based reading research) for teachers to enhance children's specific language, cognitive, and early reading skills.
- A high-quality oral-language and literature-rich classroom environment.
- Learning activities and instructional materials designed according to scientifically based reading research that cover oral language, phonological awareness, print awareness, and alphabetic knowledge.
- Assessments and other appropriate measures developed according to scientifically based reading research to determine reading skills that children are learning.
- Integration of the materials, activities, tools, and measures into the preschool's existing programs.

In this chapter, we analyze the program's impacts on teachers' professional development and classroom-learning environments. ERF funding for the 2003 cohort of grantees was awarded in October 2003, and programs were expected to train teachers and purchase materials in the fall of 2003 so that ERF would be fully implemented in classrooms by January 2004. Accordingly, we examined the impacts of ERF in both fall 2004 and spring 2005 because both time points were expected to reflect full implementation of ERF. However, to avoid repetition, we present only the spring impacts in this chapter. Fall impacts are presented in Appendix D. We obtained impact estimates by using the methods discussed in Chapter 2 and Appendix A.[59] Impacts for selected subgroups are presented in Appendix F. The analysis methods accounted for the fact that some outcome domains contained multiple measures. The tables presented include checkmarks for domains in which impacts are jointly statistically significant once the adjustment for multiple comparisons is made. The tables also include p-values for tests of statistical significance of individual outcomes that do not reflect adjustments for multiple comparisons. The conclusions are unaffected when adjustments for multiple comparisons are applied. (see Appendix A for further details on adjustments for multiple comparisons.)

We find that ERF had positive impacts on teachers' professional development in the spring. We also find statistically significant impacts on several domains of classroom quality and the language, early literacy, and assessment practices.

[59] Appendix A demonstrates that the results are robust to a variety of functional forms. In Appendix A, plots of the data provide graphical evidence of the impacts and the proper functional form of the models.

Outcome Measures

ERF funds were intended to give teachers the knowledge, skills, and materials necessary to support a literature-rich classroom environment and age-appropriate activities through which preschool children can learn language and early literacy skills. Teacher knowledge and skills are likely to be imparted primarily through professional development but can also be acquired through formal education and teaching experience.

We focus on the following aspects of the classroom environment that can potentially contribute to children's learning:

- general quality of the preschool environment
- language, early literacy, and assessment practices

The general quality measures, including teacher behaviors and aspects of the classroom environment, have been found by previous research to be positively correlated with young children's cognitive skills and emotional development (Vandell and Wolfe 2000; NICHD Early Childhood Research Network 2002, 2003, and 2006). However, given its correlational nature, this research is not conclusive.

The language, early literacy, and assessment practices in the classroom include aspects of teacher-instructional practices and the classroom environment that relate closely to the requirements of ERF. ERF specifies that grantees must provide the types of materials and learning opportunities that can support the development of children's language and early literacy skills. Grantees also should conduct regular progress assessments to gauge children's learning.

Accordingly, we examined the impacts of ERF on

- teacher knowledge and skills
- the general quality of the preschool environment
- the quality of language, early literacy, and child-assessment practices and environments

Within each of these areas, we examined measures within several domains. Table 6.1 summarizes the outcomes, domains, and measures developed for this study; we describe the domains, measures, and our hypotheses in the following text.

Table 6.1. Domains and measures for the analysis of ERF impacts on teachers and classrooms

Outcome	Domain	Measure
Teacher knowledge and skills	Teaching experience	Years experience as a preschool teacher
		Years experience teaching at this center or preschool
	Hours of professional development	Hours in the past year focusing on teaching language and literacy
		Hours in the past year focusing on curriculum
	Mode of professional development	Mode of training: mentoring
		Mode of training: workshops
		Mode of training: mentoring
		Mode of training: workshops
	Earnings	Hourly earnings
General quality of the preschool classroom	Quality of teacher-child interactions	Teaching and interactions (ECERS-R) Teacher sensitivity (TBRS) Quality of team teaching (TBRS)
	Organization of the environment	Classroom community (TBRS) Quality and organization of activity centers (TBRS)
	Planning	Lesson planning (TBRS)
	Adequacy of supervision	Child-staff ratio
Quality of language, early literacy, and assessment practices and environments	Oral language environment	Oral language use by lead teacher (TBRS) Oral language use by assistant teacher (TBRS)
	Book reading	Number of book-reading sessions (TBRS)
		Book-reading practices (TBRS)
	Phonological awareness activities	Number of different phonological awareness activities observed (TBRS)
		Quality of phonological awareness activities (TBRS)
	Print and letter knowledge	Learning opportunities (TBRS) Classroom print environment (TBRS)
	Written expression	Learning opportunities (TBRS) Opportunities and materials for writing (TBRS)
	Child screening and progress assessment	Child portfolios (TBRS) Dynamic assessment (TBRS)

ECERS-R = Early Childhood Environment Rating Scale—Revised (Harms, Clifford, and Cryer 1998).
TBRS = Teacher Behavior Rating Scale (Landry et al. 2004).

Teacher knowledge and skills were measured indirectly through teaching experience and professional development (hours and modes of training), which contribute to knowledge and skills. Exhibit 6.1 describes these measures.

Exhibit 6.1. Domains and measures of teacher experience and professional development

Teaching experience

***Years teaching preschool*—**Teachers' reports of the number of years they have taught in any preschool, at the assistant- or head-teacher level.

***Years teaching at this school*—**Teachers' reports of the number of years they have taught in their current center or school, at the assistant- or head-teacher level.

Professional development

***Professional development hours*—**Teachers' reports of the number of hours of professional development received in the past 12 months. Teachers reported about training received in two different contexts, which are not mutually exclusive:

> ***Professional development on language and literacy topics*—**Teachers' reports of the number of hours and modes of training used to learn about any language or early literacy topic in the previous 12 months.
>
> ***Professional development on curriculum*—**Teachers' reports of the number of hours and modes of training used to learn about a particular curriculum. If teachers were trained to use a curriculum focusing on language and early literacy skills, the hours and modes of training reported for this activity might be reported both as training on curriculum and as training on language and literacy topics.

***Professional development modes of training*—**Teachers' indications of whether the training they received was through mentoring or workshops.

> ***Mentoring or tutoring*—**Intensive, one-on-one training that entails an experienced or master teacher observing the mentored teacher at work in her classroom and then meeting with her later to discuss strengths and weaknesses of her practice and to suggest strategies for improvement.
>
> ***Workshops*—**Group instruction on a particular topic in a conference or adult classroom setting.

Earnings

Hourly earnings—Directors' reports of the hourly earnings of one teacher in their preschool whose classroom was observed.

We expected that ERF preschools would enhance teachers' knowledge and skills through professional development. Professional development may focus either on techniques for helping children develop language and literacy skills or on curricula designed for these purposes. ERF encouraged grantees to use intensive modes of professional development, particularly mentoring or tutoring. In addition to examining mentoring, we also measured the use of workshops for professional development. Because of their relatively low cost, workshops may be equally available to teachers in the funded and unfunded groups. Finally, higher teacher earnings can help to reduce turnover that might occur after teachers have improved their skills by receiving more training. Accordingly, we examined whether ERF increased teachers' earnings.

We examined several aspects of the general quality of the preschool environment; specific measures used in this study are described in Exhibit 6.2.

Exhibit 6.2. Measures of general quality of the preschool classroom

Early Childhood Environment Rating Scale—Revised (ECERS-R; Harms, Clifford, and Cryer 1998)— This scale is used widely to measure the quality of the classroom environment for children ages 2.5 through 5 years. Items measure the quality of space, materials, and teacher interactions with children, the range and quality of activities, and program support for parents and staff. The full scale includes 43 items, each scored from 1 (inadequate) to 7 (excellent). The ERF evaluation used a subscale of the ECERS-R:

> ***Teaching and Interactions (Clifford et al. 2005)—***This 11-item subscale was created on the basis of a factor analysis of the ECERS-R in 240 pre-kindergarten classrooms sampled from 6 states (Clifford et al. 2005). The items include those measuring the emotional and educational quality of teacher-child interactions and the encouragement of language development during the preschool day. Items are scored higher if the teacher models language or encourages the child to use language in the context of the activity.
>
> *For example, the Discipline item is scored 1 if discipline is severe, lax, or reflects inappropriate expectations; 3 if staff maintain basic control, do not use severe methods, and have generally appropriate expectations; 5 if staff use positive discipline methods (attention to positive behavior and redirection), set up the environment to promote positive interactions, and use consistent methods; and 7 if staff work with children to actively solve conflicts through discussion in conflict situations and through storybooks and if they consult professionals about behavior problems.*

Teacher Behavior Rating Scale (TBRS; Landry et al. 2004)—This scale is a research measure of the general quality and early literacy and language qualities of preschool classrooms. Originally developed as an implementation-fidelity tool linked to CIRCLE's preschool-literacy curriculum (Landry et al. 2006), the TBRS has been revised and refined for use in the Preschool Curriculum Evaluation Research (PCER) and ERF evaluations. Most items have a quantity aspect (rated 1 to 4, based on frequency) and a quality aspect (rated 0 if not observed or 1 to 4, based on low to high quality). Subscale scores are computed by first averaging, for each item, the quantity and quality scores and then averaging across these mean items. (See Appendix C for details.) Five subscales relate to the general quality of classrooms and teacher practices:

> ***Teacher Sensitivity***—The teacher offers encouragement and positive feedback; is sensitive and responsive to children's cues; provides positive guidance and encourages children to regulate behavior; and uses varied and playful techniques to engage children in literacy, language, and math activities. (4 items; same as Teacher Sensitivity)
>
> ***Quality of Team Teaching***—The teaching assistant improves the teaching environment by working with small groups of children, helping maintain classroom regulation, responding to children, engaging children, and scaffolding children's language. (5 items; same as Team Teaching)
>
> ***Classroom Community***—The classroom is arranged to encourage safe movement, positive interactions, and child independence; children's work is displayed; and rules and routines are established with children's input. (5 items; same as Classroom Community)
>
> ***Quality and Organization of Activity Centers***—Activity centers cover critical learning objectives and are linked to theme. Materials are refreshed and rotated; centers have clear boundaries, and children understand how to move between centers and use materials appropriately. Centers provide space that encourages interaction; table arrangement supports activities linked with centers. (7 items; same as Quality and Organization of Activity Centers)
>
> ***Lesson Planning***—Written lesson plans have strong thematic connections, and lessons are implemented through observed activities and materials located throughout the room. (3 items; same as Lesson Plans)

TBRS Total Score—The total TBRS score is the average score across all subscale scores.

Child-Staff Ratio—The child-staff ratio is the ratio of the observed number of children in the room to the observed number of paid staff.

The general quality of the preschool classroom environment provides a foundation for teaching and learning. We examined the impacts of ERF on these aspects of the environment because preschools may focus on these areas in order to support the language and literacy activities that are central to ERF.

The quality of language, early literacy, and child-assessment practices and environments is a major focus of ERF, and we have developed several measures for this study, based on the TBRS. The measures examine teacher instructional practices and the materials available in the classroom environment (see Exhibit 6.3); the measures are scaled so that higher values represent greater frequency or quality or both. Most TBRS items measure both the frequency and the quality of a teacher activity or classroom feature, but these ratings are highly correlated (see Appendix C for details about the TBRS and the measures used in this chapter).

Exhibit 6.3. Measures of language, early literacy, and assessment practices in preschool classrooms

Teacher Behavior Rating Scale (TBRS; Landry et al. 2004)—This scale is a research measure of the general quality and early literacy and language qualities of preschool classrooms. Originally developed as an implementation fidelity tool linked to CIRCLE's preschool literacy curriculum (Landry et al. 2006), the TBRS has been revised and refined for use in the Preschool Curriculum Evaluation Research (PCER) and ERF evaluations. Most items have a quantity aspect (rated 1 to 4 based on frequency) and a quality aspect (rated 0 if not observed or 1 to 4 based on low to high quality). Subscale scores are computed by first averaging, for each item, the quantity and quality scores and then averaging across these mean items. (See Appendix C for original TBRS measures and ERF adaptations.) The following 12 outcome measures relate to the language and literacy environment of classrooms and teacher practices in these areas and in tracking children's progress:

Oral language use by lead teacher—The teacher models language, speaks clearly and grammatically, uses rich labels, descriptors, and verbs, uses open-ended "thinking" questions, relates previously learned words and concepts to activity, encourages children's use of language, and engages children in turn-taking conversations. (7 items; same as Oral Language Use with Students in original TBRS)

Oral language use by assistant teacher—The assistant teacher uses rich labels, descriptors, and verbs; asks open-ended questions; and encourages conversations in small-group work as she moves around the classroom. (2 items out of 5 from the Team Teaching Ability subscale in original TBRS)

Number of book-reading sessions observed—Observations note the number of times the teacher reads a book to children, either in large or small groups, during the two-hour observation period. (1 descriptive observation item coded in conjunction with (but not part of) the Book-Reading Behaviors subscale in original TBRS)

Book-reading practices—Teacher and children discuss features of the book (for example, the title and illustrator); teacher discusses vocabulary words and uses pictures or objects as props for the words before reading; teacher captures attention using facial expression, voice, and modulation; paces reading; and allows children to comment; teacher asks open-ended questions and initiates activities and discussions to extend the book reading. (8 items; same as Book Reading Behaviors subscale in original TBRS)

Number of different phonological awareness activities observed—Observations note the number of distinct activities carried out during the two-hour period, including listening, rhyming, alliteration, sentence segmenting, syllable blending and segmenting, onset-rhyme blending and segmenting, phoneme blending, segmenting, and manipulation. (1 item based on count of 7 possible activities from Phonological Awareness Activity in original TBRS)

Quality of phonological awareness activities—The level of child engagement is noted in the observed phonological awareness activities. (1 item average of 7 possible observations from Phonological Awareness Activity in original)

Print and letter knowledge learning opportunities—The teacher engages children in activities that promote children's knowledge of the names and shapes of letters, the sounds of letters, and concepts about print; score reflects number of such opportunities and children's level of engagement. (3 items out of 6 from Print and Letter Knowledge in original)

Classroom print environment—The classroom has a letter wall with letters, pictures, and related activities; activity centers include books and printed words that relate to the center, topic, or theme. (3 items out of 6 from Print and Letter Knowledge in original TBRS)

Written expression learning opportunities—The teacher models writing in large or small groups. (1 item out of 3 from Written Expression in original TBRS)

Opportunities and materials for writing—The classroom includes many types of materials for children's writing, and writing materials are included in a large number of activity centers. (2 items of 3 from original Written Expression)

Child portfolios—A large proportion of children's portfolios contain diverse samples of children's work and recently dated teacher-written observations. (2 items; same as Portfolios in original TBRS)

Dynamic Assessment—Portfolios include documentation of assessment across a range of emergent literacy areas within the past 30 days; teachers use assessments to plan instruction and a variety of activities. (3 items; same as Dynamic Assessment in original TBRS)

Outcome measures for the teacher- and classroom-level analyses were obtained from three sources. Teacher characteristics, experience, formal education, and professional development were measured by a *teacher self-administered survey* completed in fall and spring. Hourly earnings for one randomly selected teacher per preschool were reported by the preschool director in the *fall and spring director survey*. Classroom environments and teacher practices in the classroom were measured by trained observers, who completed semistructured *observation protocols during 3-hour classroom visits* in the fall and spring.

Impacts on Teachers and Classroom Environments

Overall, we find that in the spring, ERF had positive impacts on teachers' professional development. The program increased hours of professional development during the 12 months preceding the survey and the proportion of teachers receiving professional development through mentoring. ERF also had pervasive impacts on the general quality of the preschool classroom; on the classroom language environment, materials, and teaching practices that support early literacy; and on child-assessment practices.

Impacts on Teachers' Qualifications

One way in which ERF preschools could have improved teacher knowledge and skills was to hire new teachers with higher levels of experience. However, we find no evidence of an impact of ERF on years of teaching experience, measured as either teaching preschool generally or teaching at the current school or center.

ERF had a positive impact on teachers' professional development in spring 2005 (see Table 6.2). The program increased the number of hours of professional development that focused on language and early literacy topics by 50 hours (approximately 6 days) over the 12 months preceding the spring survey. ERF also had a positive impact on the mode of training. A higher proportion of ERF teachers than teachers in unfunded programs reported receiving professional development on language or literacy topics and on curriculum topics through mentoring or tutoring. The estimated impact on the proportion of teachers receiving mentoring or tutoring on language and literacy topics was 41 percentage points. Over half of ERF teachers reported receiving mentoring or tutoring in the previous year on language and literacy topics (56 percent, using regression-adjusted percentages), compared with 15 percent of unfunded teachers. A larger proportion of ERF teachers than teachers in unfunded programs also reported receiving workshop training on language and literacy topics. The estimated impact on the proportion of teachers receiving workshop training on language and literacy topics was 41 percentage points. Seventy-three percent of ERF teachers reported receiving mentoring in the previous year on language and literacy topics (using regression-adjusted percentages), compared with 38 percent of unfunded teachers.

Table 6.2. ERF impacts on teachers' experience, training, and earnings, spring 2005

Domain/Outcome (range)	Unadjusted means		Regression-adjusted means				
	Funded	Unfunded	Funded	Unfunded	Estimated impact[a]	Effect size[b]	P-value of impact
Teaching Experience							
Years at current school or center (0–30)	5.44	6.27	6.33	4.45	1.88	0.32	0.248
Years at any preschool (0–36)	9.34	9.37	9.93	8.37	1.56	0.21	0.405
Professional Development ✓							
Professional development focusing on early-language and literacy topics:							
Hours (1–160)	67.77	30.27	72.03	22.09	49.94	1.04	0.002*
Received professional development through:							
Mentoring or tutoring (%)	60.00	15.00	55.60	14.90	40.70	0.86	0.009*
Workshops (%)	64.44	38.00	72.80	32.03	40.77	0.82	0.000*
Professional development focusing on curriculum:							
Hours (0–160)	43.37	19.00	39.91	24.51	15.41	0.39	0.209
Received professional development through:							
Mentoring or tutoring (%)	46.67	17.00	49.32	14.25	35.07	0.78	0.027*
Workshops (%)	56.67	40.00	53.05	46.46	6.59	0.13	0.675
Sample Size							
Number of teachers			90	100			
Number of sites			28	37			
Earnings							
Teachers' hourly earnings (6.09–54.44)	20.20	17.98	20.46	17.28	3.18	0.30	0.517
Sample Size							
Number of preschools			43	45			
Number of sites			23	30			

*p-value (of adjusted difference in means) < 0.05; two-tailed test.
✓Impact on domain is positive and statistically significant after adjustments for multiple comparisons (see Appendix A).
[a]All estimates except those for earnings were obtained from a regression model of the outcome variable on an indicator variable of ERF grant receipt; grant application score; and teacher's education, age, and an indicator variable of nonwhite, using SAS's PROC MIXED procedure for continuous outcome measures and SUDAAN logit for binary outcome measures. Missing values of covariates were mean-imputed by site. For earnings, the regression model included only an indicator variable of ERF grant receipt and grant application score without any teacher demographic controls.
[b]The effect size was calculated by dividing the estimated impact by the standard deviation of the outcome measure (that is, the impact expressed as a percentage of the standard deviation).
NOTE: All figures were estimated by using sample weights to account for the sample and survey designs. Standard errors of the impact estimates account for design effects due to unequal weighting of the data and clustering at site level.
SOURCE: ERF spring teacher surveys and director surveys.

We found no statistically significant differences in the hourly earnings of teachers in ERF programs relative to those in unfunded programs in the spring. We conclude that ERF did not induce preschools to raise the wages of their teachers, who had received additional professional development through the program.[60]

Impacts on General Quality of Preschool Classrooms

In the spring, ERF had positive impacts on each of the domains of the general quality of preschool classrooms except adequacy of supervision (see Table 6.3). ERF increased the lead teachers' sensitivity and the quality of interactions toward children by approximately one standard deviation relative to what we would have expected in the absence of the program. In addition, team teaching, which measures the extent to which the assistant teacher contributes to the language environment and acts as a team player to extend the lead teacher's activities, was improved by 0.79 standard deviations.

Impacts on the two measures of the organization of the classroom environment—classroom community and the quality and organization of activity centers—exceed one standard deviation. ERF also significantly improved lesson planning.

ERF increased the overall quality of the classroom-learning environment, measured by the total TBRS score (the average across subscales measuring general classroom quality and the language and early literacy environment). In ERF classrooms, the regression-adjusted average total TBRS score was 1.44 standard deviations higher than it would have been in the absence of ERF.

[60] The teacher hourly earnings data are reported by center directors, not teachers.

Table 6.3. ERF impacts on classroom outcomes: general quality of the preschool classroom, spring 2005

Domain/Outcome (range)	Unadjusted means		Regression-adjusted means				
	Funded	Unfunded	Funded	Unfunded	Estimated impact[a]	Effect size[b]	P-value of impact
Quality of Teacher-Child Interactions ✓							
Teaching and Interactions (ECERS-R) (1.60–7.00)	5.78	5.09	5.94	4.73	1.20	1.12	0.001*
Teacher Sensitivity (TBRS) (0.50-4.00)	3.07	2.69	3.16	2.49	0.67	0.99	0.008*
Quality of Team Teaching (TBRS) (0.80–4.00)	2.99	2.40	3.04	2.29	0.76	0.79	0.049*
Organization of the Environment ✓							
Classroom Community (TBRS) (0.90–4.00)	3.19	2.75	3.33	2.51	0.82	1.22	0.001*
Quality and Organization of Activity Centers (TBRS) (0.86–4.00)	2.93	2.38	3.03	2.14	0.88	1.13	0.003*
Planning							
Lesson Planning (TBRS) (0.50–4.00)	3.05	2.41	3.13	2.27	0.87	0.84	0.016*
Total Teacher Behavior Rating Scale							
Total TBRS Score (0.94–3.89)	2.65	2.07	2.77	1.84	0.93	1.44	0.000*
Adequacy of Supervision							
Child-staff ratio (2.40–20.00)	7.50	7.65	7.06	8.19	–1.13	–0.38	0.336
Sample Size							
Number of Classrooms			78	91			
Number of Sites			28	37			

*p-value (of adjusted difference in means) < 0.05; two-tailed test.
✓Impact on domain is positive and statistically significant after adjustments for multiple comparisons (see Appendix A).
[a]All estimates were obtained from a regression model of the outcome variable on an indicator variable of ERF grant receipt; grant application score; and teacher's education, age, and an indicator variable of nonwhite, using SAS's PROC MIXED procedure. Missing values of covariates were mean-imputed by site.
[b]The effect size was calculated by dividing the estimated impact by the standard deviation of the outcome measure (that is, the impact expressed as a percentage of the standard deviation).
NOTE: All figures were estimated by using sample weights to account for the sample and survey designs. Standard errors of the impact estimates account for design effects due to unequal weighting of the data and clustering at site level.
SOURCE: ERF spring classroom observations.

ERF had no statistically significant impact on observed child-staff ratios in the spring. Ratios for both funded and unfunded programs were between 7 and 8 children per staff member, well within professionally accepted upper limits for ratios in preschool classrooms (10 children per adult).

Impacts on Classroom Support for Language and Early Literacy

In the spring, ERF had positive impacts on all domains of classroom language, early literacy, and assessment practices (see Table 6.4). The Oral Language Use subscale measures the language environment provided by the lead teacher and the assistant teacher in the classroom. Oral language use by both the lead and assistant teachers in ERF classrooms was rated higher than it would have been in the absence of ERF, by 1.11 standard deviations for lead teachers and by 0.89 standard deviations for assistant teachers.

Book-reading practices, which measures the use of a book-reading session to reinforce concepts of print and encourage children's oral expression, were rated higher in ERF classrooms than they would have been in the absence of ERF by 1.03 standard deviations. However, ERF did not increase the number of book-reading sessions (the number of times a teacher sat down with children to read one or more books).

Table 6.4. ERF impacts on classroom outcomes: language, early literacy, and assessment practices, spring 2005

Domain/Outcome (range)	Unadjusted means		Regression-adjusted means				
	Funded	Unfunded	Funded	Unfunded	Estimated impact[a]	Effect size[b]	P-value of impact
Oral Language Environment ✓							
Oral language use by lead teacher (0.50–4.00)	2.88	2.39	3.00	2.17	0.83	1.11	0.002*
Oral language use by assistant Teacher (0.50–4.00)	2.67	1.90	2.77	1.73	1.04	0.89	0.027*
Book Reading ✓							
Number of book-reading sessions observed (0–4)	1.45	1.16	1.41	1.20	0.21	0.23	0.516
Book-reading practices (0.56–3.94)	2.40	1.77	2.49	1.60	0.89	1.03	0.003*
Phonological Awareness Activities ✓							
Number of different phonological awareness activities observed (0–7)	2.24	0.96	2.40	0.67	1.73	1.10	0.004*
Quality of phonological awareness activities (0–4.00)	1.91	1.30	2.04	1.07	0.97	0.79	0.024*
Print and Letter Knowledge ✓							
Learning opportunities (0.50–4.00)	2.04	1.29	2.05	1.20	0.85	0.87	0.022*
Classroom print environment (0.50–4.00)	2.24	1.71	2.28	1.59	0.69	0.81	0.028*
Written Expression ✓							
Learning Opportunities (0.50–4.00)	1.88	0.99	1.99	0.78	1.21	1.06	0.003*
Opportunities and materials for writing (0.50–4.00)	2.34	1.72	2.55	1.32	1.23	1.48	0.000*
Child Screening and Progress Assessments ✓							
Child portfolios (1.00–5.00)	2.82	2.09	3.07	1.72	1.35	0.98	0.012*
Dynamic assessment (0.67–4.33)	2.79	2.34	2.89	2.18	0.71	0.64	0.095
Sample Size							
Number of Classrooms			78	90			
Number of Sites			28	37			

*p-value (of adjusted difference in means) < 0.05; two-tailed test.
✓Impact on domain is positive and statistically significant after adjustments for multiple comparisons (see Appendix A).
[a]All estimates were obtained from a regression model of the outcome variable on an indicator variable of ERF grant receipt; grant application score; and teacher's education, age, and an indicator variable of nonwhite, using SAS's PROC MIXED procedure. Missing values of covariates were mean-imputed by site.
[b]The effect size was calculated by dividing the estimated impact by the standard deviation of the outcome measure (that is, the impact expressed as a percentage of the standard deviation).
NOTE: All figures were estimated by using sample weights to account for the sample and survey designs. Standard errors of the impact estimates account for design effects due to unequal weighting of the data and clustering at site level.
SOURCE: ERF spring classroom observations.

ERF had positive impacts on classroom materials and teacher practices to promote children's letter recognition and the association between sounds and letters (the domains of phonological awareness activities, print and letter knowledge, and written expression). Phonological-awareness activities measured by the TBRS include listening, rhyming, alliteration, sentence segmenting, onset-rime blending and segmenting words), syllable blending or segmenting, and phoneme blending, segmenting, and manipulation. ED guidance on ERF recommends additional phonological awareness activities beyond traditional nursery school rhymes. We expect that ERF teachers will look for more opportunities to introduce phonological awareness activities in class. We found that the number of different phonological awareness activities observed during the 3-hour observation period was higher in ERF classrooms than in unfunded classrooms by 1.73 (or nearly 2) activities, on average. (Appendix D provides details about the percentage of classrooms in which each type of phonological awareness activity was observed.) The quality of these activities, measured by the level of children's engagement, was also significantly higher in ERF classrooms than it would have been in the absence of ERF.

ERF had a positive impact on the classroom print environment (labels, books, and letters displayed with pictures) and the opportunities and materials for writing. Regression-adjusted average scores for the classroom print environment subscale were 0.81 standard deviations higher in ERF classrooms than in unfunded classrooms, and scores for opportunities and materials for writing in ERF classrooms were 1.48 standard deviations higher. ERF also had a positive impact on teacher practices in these areas. Print- and letter-knowledge learning opportunities tap both the frequency that teachers provide lessons or explanations about print and letters and the level of children's engagement in them. The impact of ERF on print- and letter-knowledge learning opportunities is 0.87 standard deviations, and the impact on written-expression learning opportunities (modeling writing) is 1.06 standard deviations.

ERF requires teachers to periodically assess children's language development and literacy skills as a basis for building lessons on what children know, but it does not require teachers to use portfolios. ERF had positive impacts on child screening and progress assessment in the spring. ERF improved the extensiveness and completeness of children's portfolios, although it did not have statistically significant impacts on dynamic assessment.

Chapter 7. Impact Findings: ERF Impacts on Children's Language and Literacy Skills and Social-Emotional Outcomes

Ultimately, through its effects on classroom practices, the ERF Program is intended to provide young children with the necessary language, cognitive, and early reading skills to prevent reading difficulties and ensure school success as they enter kindergarten. In this chapter, we examine whether ERF achieved this goal, through our analysis of the program's impacts on three domains of children's language and early literacy skills: print and letter knowledge, phonological awareness, and oral language. In addition, we examine the program's effects in the *nonliteracy* domain of social-emotional development, in response to concerns that ERF might have had detrimental effects in this domain if it led teachers to focus on improving early literacy skills at the exclusion of other areas of child development. The analytic methods underlying this analysis are discussed in Appendix A.[61] The analysis methods accounted for the fact that some outcome domains contained multiple measures. The tables presented include checkmarks for domains in which impacts are jointly statistically significant once the adjustment for multiple comparisons is made. The tables also include p-values for tests of statistical significance of individual outcomes that do not reflect adjustments for multiple comparisons. The conclusions are unaffected when adjustments for multiple comparisons are applied (see Appendix A for further details on adjustments for multiple comparisons).

We find that the program had a statistically significant positive effect on children's print and letter knowledge. However, we find no statistically significant impacts on either phonological awareness or oral language. We also find no evidence that the program had detrimental effects on any of the nonliteracy outcomes examined.

Outcome Measures

The outcome measures for the child-level analyses were obtained from assessments that were given to children in spring of the school year on their literacy and language skills and behavior.

We examined ERF impacts on children's literacy and language skills in three domains. To measure *print and letter knowledge,* we used the Print Awareness subtest of the Preschool Comprehensive Test of Phonological and Print Processing (Pre-CTOPPP, Lonigan et al. 2002). To measure *phonological awareness,* we used the Elision subtest of the Pre-CTOPPP (Lonigan et al. 2002). To measure *oral language,* we used two separate assessments: the Expressive One-Word Picture Vocabulary Test (EOWPVT, Brownell 2000) and the Auditory Comprehension subtest of the Preschool Language Scale, Fourth Edition (PLS-4, Zimmerman et al. 2002). Higher values for each measure are associated with higher literacy and language skills. Exhibit 7.1 describes these measures and provides sample items.

[61] Appendix A demonstrates that the results are robust to a variety of functional forms. In Appendix A, plots of the data provide graphical evidence of the impacts and the proper functional form of the models.

We also estimated ERF's impacts on children's social-emotional development, as measured by three subscales of the 30-item Social Competence and Behavior Evaluation (SCBE); see Exhibit 7.2. This evaluation is based on assessments of the child by the child's teacher. The three 10-item subscales include a social-competence subscale, an anger-aggression subscale, and an anxiety-withdrawal subscale. Higher values on the social-competence subscale represent a positive outcome (the child is more socially competent) while higher values on the anger-aggression and anxiety-withdrawal subscales indicate negative outcomes (the child is more angry-aggressive or anxious-withdrawn).

Exhibit 7.1. Domains of language and early literacy skills and associated measures

Print and Letter Knowledge—measured by the ***Print Awareness*** subtest of the Preschool Comprehensive Test of Phonological and Print Processing (Pre-CTOPPP; Lonigan, et al., 2002).

The Pre-CTOPPP includes subtests that measure print concepts, letter and word discrimination, letter identification, phonological sensitivity (sound and word blending and elision), and vocabulary for children ages 3 to 6 years. Children are directly assessed by using a standard protocol. The ERF evaluation used a research version of the test available in 2004; however, the slightly revised test with normed scores has been published by ProEd as the Test of Preschool Early Literacy (TOPEL). The TOPEL norms can be used to derive age-adjusted, standardized scores for the Pre-CTOPPP Print Awareness subtest. The Print Awareness normed scores have a mean of 100 and a standard deviation of 15; see Appendix C for more information on how these standard scores were constructed.

The Print Awareness subtest measures print concepts, letter and word discrimination, letter identification, and letter-sound recognition.

For example, the child is asked to point to the title of the book; distinguish letters from numbers; distinguish words from numbers and pictures; identify printed letters; associate letters with sounds; provide the name of particular letters; and provide the sound of particular letters.

Phonological Awareness—measured by the ***Elision*** subtest of the Pre-CTOPPP (see above). Because of differences in the Pre-CTOPPP and the TOPEL, norms cannot be used to derive scores for the Elision subtest, so only raw scores are presented for this measure.

The Elision subtest measures the child's ability to isolate and drop a syllable or phoneme from a word, which is one component of phonological awareness.

For example, the child is asked to say a compound word and drop one part ("toothbrush" without "brush"); say a two-syllable word and drop one part ("candy" without "dee"); and say a one-syllable word and drop one phoneme ("heat" without "t") both with and without multiple-choice picture prompts.

Oral Language—measured by (1) the ***Expressive Vocabulary*** subtest of the Expressive One-Word Picture Vocabulary Test, Third Edition (EOWPVT-III; Brownell 2000) and (2) the ***Auditory Comprehension*** subtest of the Preschool Language Scale, Fourth Edition (PLS-IV; Zimmerman, et al., 2002).

(1) The EOWPVT-III measures English-speaking vocabulary of children ages 24 months to 18 years, 11 months. Children are directly assessed by using a standard protocol. The EOWPVT-III was normed on a nationally representative sample of children of various ages so that raw scores can be converted to age-adjusted, standardized scores with a mean of 100 and a standard deviation of 15.

The Expressive Vocabulary subtest is designed to assess expressive vocabulary and word retrieval.

The child is presented with pictures and is asked to name the objects, actions, and concepts shown in the pictures. Children are asked to name pictures showing a personal computer, a wagon, and a teacup; they are shown a picture of a painter and asked, "What is he doing?" and they are shown a picture of a cow, a bear, a giraffe, and a turkey and asked, "What word names all of these?"

(2) The PLS-IV measures language development of children from birth through 6 years, 11 months. The PLS includes two subtests, Auditory Comprehension and Expressive Communication. Each subtest was normed on a nationally representative sample of children of various ages so that raw scores can be converted to age-adjusted, standardized scores with a mean of 100 and a standard deviation of 15. Children are directly assessed by using a standard protocol.

The Auditory Comprehension subtest measures comprehension of basic vocabulary, concepts, and grammatical markers such as comparatives and superlatives. Test items ask children to identify a named color, identify categories of objects, understand "more" and "most," understand expanded sentences, qualitative concepts, and time concepts, understand the –er ending as one who . . ., and identify objects that do not belong to a group.

For example, the child is asked to point to the bear that is blue; complete analogies such as "Ice cream is cold; a fire is ____;" point to the animal with the longest nose; and identify which item does not belong in a set that includes a car, a truck, a boat, and a chair.

Exhibit 7.2. Measures of social-emotional development

Social-Emotional Development—measured by three subscales of the Social Competence and Behavior Evaluation—Short Form (LaFreniere and Dumas 1996), which measures the child's affect and behavior in relationships with teachers and peers. Teachers rate the child's "typical behavior or emotional state" on 30 items, each scored from 0 (never occurs) to 5 (always occurs). Three subscales were formed from these items:

***Social Competence*—**measures the extent to which the child exhibits cooperative behavior and interacts well in relation to other children. For example, the measure asks about "takes other children and their point of view into account," "comforts or assists another child in difficulty," and "takes pleasure in own accomplishments." The subscale includes 10 items, and the score is the sum of the items.

***Anxiety-Withdrawal*—**measures the extent to which the child tends to withdraw from groups of children or to exhibit sad or anxious behavior. For example, the measure asks about "worries," "doesn't talk or interact in a group," and "sad, unhappy." The subscale includes 10 items, and the score is the sum of the items.

***Anger-Aggression*—**measures the extent to which the child exhibits angry, oppositional, or destructive behavior or tends to be in conflict with others. For example, the measure asks about "screams or yells easily," "hits you or destroys things when angry with you," and "opposes your suggestions." The subscale includes 10 items, and the score is the sum of the items.

Impacts on Child Outcomes

ERF had a statistically significant positive effect on print and letter knowledge (see Table 7.1). The program increased children's Pre-CTOPPP print awareness standard scores by 5.78 points (p-value = 0.042) relative to what we would have expected in the absence of the program. This increase indicates that ERF improved children's ability to recognize letters of the alphabet and associate letters with their sounds. The impact estimate translates into an effect size of 0.34 standard deviations. Results are similar for print awareness raw scores. Comparison of the regression-adjusted standard scores for children in the unfunded sites to the national norms for this subtest indicates that in the absence of ERF, children in the ERF sites would have scored about 3 percentage points below the national average of 100; with exposure to ERF, their average score of 102.69 was slightly above the national average for this subtest.

Table 7.1. ERF impacts on child outcomes in spring, preferred model, without controls for fall value of outcome measure

Outcome (range)	Unadjusted means		Regression-adjusted means		Estimated impact[a]	Effect size[b]	P-value of impact
	Funded	Unfunded	Funded	Unfunded			
Language and Literacy Skills							
Print and Letter Knowledge✓							
Print awareness, raw score (0–36)	22.73	20.10	23.51	19.11	4.40	0.44	0.027*
Print awareness, standard score (58–144)	101.39	98.92	102.69	96.91	5.78	0.34	0.042*
Phonological Awareness							
Elision, raw score (0–18)	9.18	9.20	9.40	8.99	0.41	0.10	0.441
Oral Language							
Expressive vocabulary, raw score (0–99)	38.74	39.56	39.42	39.33	0.09	0.01	0.965
Expressive vocabulary, standard score (53–147)	82.98	83.91	83.90	83.43	0.47	0.03	0.841
Auditory comprehension, raw score (1–62)	51.64	51.33	52.38	50.36	2.01	0.27	0.095
Auditory comprehension, standard score (50–135)	92.59	91.70	94.11	89.82	4.29	0.28	0.088
Number of students	802	846					
Number of sites	28	37					
Social Competence and Behavior Evaluation (Scales range from 0 to 50)							
Social competence	31.46	32.23	32.16	31.24	0.93	0.10	0.617
Anxiety-withdrawal	10.73	10.76	10.80	10.81	-0.01	-0.00	0.992
Anger-aggression	9.03	9.83	8.49	10.73	-2.25	-0.26	0.128
Number of students	801	844					
Number of sites	28	37					

*p-value (of adjusted difference in means) < 0.05, two-tailed test.

[a]All estimates were obtained from a regression model of the outcome variable on an indicator variable of ERF grant receipt; grant application score; and indicator variables of female and nonwhite, using SAS's PROC MIXED procedure. Language and literacy skill models also control for indicator variables of fall assessment taken in Spanish, missing fall assessment data, and age at spring assessment. SCBE models also control for an indicator variable of missing fall SCBE data and age at spring SCBE observation. Missing values of covariates are mean-imputed by site and gender.

[b]The effect size was calculated by dividing the estimated impact by the standard deviation of the outcome measure (that is, the impact expressed as a percentage of the standard deviation).

✓Impact on domain is positive and statistically significant after adjustments for multiple comparisons (see Appendix A).

NOTE: All figures were estimated by using sample weights to account for the sample and survey designs. Standard errors of the impact estimates account for design effects due to unequal weighting of the data and clustering at site and classroom level.

SOURCE: ERF spring child assessments and SCBE evaluations.

We find no evidence that ERF improved children's phonological awareness (see Table 7.1). The estimated impact on Elision scores is small and not statistically significant at conventional levels. The estimate is similar in a model that included the pretest as a covariate (see Table 7.2).

Similarly, we find no evidence that ERF improved children's oral language skills. ERF's impact on the first measure in this domain—the expressive vocabulary subtest—is small and not statistically significant at conventional levels (see Table 7.1). Results are similar in a model that included the pretest as a covariate (see Table 7.2). ERF's estimated impact was an increase of 4.29 points in the standard score on the second measure in the oral language domain—the auditory comprehension subtest—not statistically significant at the 5 percent level (see Table 7.1). Also, tests that adjust for the multiple outcomes in the oral-language domain indicate that there is no statistically significant impact on children's skills in this domain (see Appendix A).

ERF did not affect children's social-emotional skills, as measured by the SCBE-30 anger-aggression, social-competence, and anxiety-withdrawal scales (see Tables 7.1 and 7.2). The estimated impact on children's social competence is positive but not statistically significant. The estimated impact on anxiety-withdrawal is close to zero and not statistically significant. The estimated impact on anger-aggression is negative and points to a reduction in anger-aggression due to ERF. However, this estimate is also not statistically significant. The lack of program effects in this domain is noteworthy in light of concerns that ERF might *adversely* impact these skills by compelling teachers to focus on improving language and literacy at the expense of developing other skills; our null estimates for these outcomes suggest that ERF did not adversely affect children's nonliteracy skills.

ERF thus appears to have had a positive effect on children's print and letter knowledge but not on phonological awareness or oral language. In addition, ERF neither enhanced nor diminished children's social-emotional development during the preschool year.

Table 7.2. ERF impacts on child outcomes in spring, preferred model, with controls for fall value of outcome measure

Outcome (range)	Funded	Unfunded	Estimated impact[a]	Effect size[b]	P-value of impact
Language and Literacy Skills					
Print and Letter Knowledge					
Print awareness, raw score (0–36)	—	—	—	—	—
Print awareness, standard score (58–144)	—	—	—	—	—
Phonological Awareness					
Elision, raw score (0–18)	9.50	8.89	0.61	0.14	0.236
Oral Language					
Expressive vocabulary, raw score (0–99)	39.78	39.17	0.62	0.04	0.659
Expressive vocabulary, standard score (53–147)	83.98	83.44	0.54	0.03	0.727
Auditory comprehension, raw score (1–62)	—	—	—	—	—
Auditory comprehension, standard score (50–135)	—	—	—	—	—
Number of students	802	846			
Number of sites	28	37			
Social Competence and Behavior Evaluation (Scales range from 0 to 50)					
Social competence	32.28	31.56	0.72	0.08	0.591
Anxiety-withdrawal	11.00	10.42	0.58	0.09	0.569
Anger-aggression	9.03	10.15	–1.12	–0.13	0.249
Number of students	801	844			
Number of sites	28	37			

*p-value (of adjusted difference in means) < 0.05, two-tailed test.
— Not available. Impact estimates controlling for fall values of outcome measures are not presented for these outcomes, because of evidence of early impacts on fall measures that would bias impact estimates on spring measures. See Appendix A for additional discussion.
[a]All estimates were obtained from a regression model of the outcome variable on an indicator variable of ERF grant receipt; grant application score; and an indicator variable of nonwhite, using SAS's PROC MIXED procedure. Language and literacy skill models also control for indicator variables of fall assessment taken in Spanish and missing fall assessment data and age at spring assessment. SCBE models also control for an indicator variable of missing fall SCBE data and age at spring SCBE observation. Missing values of covariates are mean-imputed by site and gender.
[b]The effect size was calculated by dividing the estimated impact by the standard deviation of the outcome measure (that is, the impact expressed as a percentage of the standard deviation).
NOTE: All figures were estimated by using sample weights to account for the sample and survey designs. Standard errors of the impact estimates account for design effects due to unequal weighting of the data and clustering at site and classroom level.
SOURCE: ERF spring child assessments and SCBE evaluations.

Chapter 8. Analysis of Mediators of ERF's Impacts on Classroom Instructional Practice and Children's Language and Literacy Skills

Through its focus on teacher training and professional development, ERF seeks to improve language and literacy instruction in the classroom and, in turn, to improve children's language and early literacy skills. Chapter 6 of this report documents ERF's positive impacts on several measures of the classroom learning environment, and Chapter 7 documents ERF's positive impact on children's print and letter knowledge. In this chapter, we explore potential channels, or mediators, through which ERF generated its positive impacts on classroom and child outcomes. Unlike the impact analyses presented in previous chapters, this analysis is correlational, rather than quasi-experimental, because we cannot use the regression-discontinuity design to identify the causal effects of particular mediators. Consequently, any observed effect of mediators on child or classroom outcomes might be due to the effects of unobserved factors that happen to be correlated with these mediators, rather than to the mediators themselves.

Models of Professional Development, Classroom Practice, and Children's Language and Literacy Skills

This report has shown that ERF had positive, statistically significant impacts on several classroom and teacher outcomes and on one child outcome. As shown in Chapter 7, ERF had positive impacts on the number of hours of professional development that teachers received and on the use of mentoring as a mode of training. ERF also had positive impacts on aspects of classroom environments and teacher practices that were major program focuses, including the language environment of the classroom, book-reading practices, the variety of phonological-awareness activities and children's engagement in them; materials and teaching practices to support print and letter knowledge and writing; and the extensiveness and recency of child-assessment practices. ERF also had positive impacts on other, more general aspects of classroom quality, including the quality of teacher-child interactions, the organization of the classroom, and the planning of activities for children. Finally, as shown in Chapter 7, ERF had a positive impact on children's print awareness.

For our analysis of the channels through which ERF generated positive impacts on classroom and child outcomes, we hypothesized that the additional hours of professional development attributable to ERF and the increased proportion of teachers receiving professional development through intensive, individualized mentoring account for at least some of ERF's impact on the classroom language and early literacy environment. The impacts on classroom environments, in turn, might account for at least some of the program's impacts on children's language and literacy skills.

To investigate this hypothesis, we first examine the extent to which hours of professional development and the use of mentoring as a mode of training are associated with the classroom outcomes affected by ERF. Table 8.1 shows the outcome variables that we examined and their associated potential mediators.

Table 8.1. Potential mediators of child and classroom outcomes

Outcome	Potential mediators
Classroom outcomes	
Book-reading practices	Hours of professional development
	Whether received any training though mentoring
Number of phonological awareness activities	Hours of professional development
	Whether received any training though mentoring
Print and letter knowledge learning opportunities	Hours of professional development
	Whether received any training though mentoring
Written expression learning opportunities	Hours of professional development
	Whether received any training though mentoring
Classroom print environment	Hours of professional development
	Whether received any training though mentoring
Opportunities and materials for writing	Hours of professional development
	Whether received any training though mentoring
Oral language use by lead teacher	Hours of professional development
	Whether received any training though mentoring
Oral language use by assistant teacher	Hours of professional development
	Whether received any training though mentoring
Child portfolios	Hours of professional development
	Whether received any training though mentoring
Teacher sensitivity	Hours of professional development
	Whether received any training though mentoring
Child outcomes	
Print awareness, standard score	Book-reading practices
	Number of phonological awareness activities
	Print and letter knowledge learning opportunities
	Written-expression learning opportunities
	Classroom print environment
	Opportunities and materials for writing
	Child portfolios
	Teacher sensitivity

We then examine the associations between classroom outcomes and the child outcome on which ERF had a positive impact—print and letter knowledge. The print awareness test used to measure skills in this domain requires children to recognize features of a book, to distinguish print from pictures, to recognize letters, and to associate sounds with letters. The development of these skills could be influenced by the extent to which teachers create or take advantage of opportunities for children to learn the sounds of letters, to learn to distinguish print from pictures, to learn about the sounds of words and parts of words, and to think about the shapes of letters and associate letter names with letter shapes. These skills are also supported by examples of print in the classroom environment and by the availability of materials for writing. Book-reading practices that include introducing features of the book and discussing those features may also help children acquire the skills needed for the print-awareness assessment. Teacher sensitivity and encouragement and regular, comprehensive assessment of children could also contribute to children's performance in this area (Landry 2005). Thus, as shown in Table 8.1, our model of print awareness includes as mediators the number of phonological awareness activities, print- and letter-knowledge learning opportunities, written-expression learning opportunities, the classroom print environment, opportunities and materials for writing, book-reading practices, child portfolios, and teacher sensitivity.

Approach to Estimation

The estimation approach for the mediated analysis has four stages. In the *first* stage, we regress each potential mediator on an indicator of treatment status, grant applicant score, and additional covariates in order to obtain estimates of the impact of ERF on the potential mediator:

(1) $M_i = b_0 + b_1 T + b_2 Score + X_M b_3 + e$

where M_i is mediator i, T is an indicator of treatment status, *Score* is the grant application score (normalized to have a mean of zero), X_M is a vector of covariates, and e is a random error term. Estimates are weighted to account for the sample and survey designs. The estimated coefficient $\hat{b}_1$ provides an estimate of ERF's impact on mediator i, which we denote as I_{Mi}.

In the *second* stage, we regress the outcome variable (child or classroom level) on an indicator for treatment status, *Score*, the potential mediating variables, and a set of exogenous explanatory variables:

(2) $Y = \alpha_0 + \alpha_1 T + \alpha_2 Score + \sum_i M_i \gamma_i + X\beta + \varepsilon$

where X is a vector of additional explanatory variables, ε is a random error term, and the other variables are defined as above. Additional explanatory variables for the classroom-level analysis include teacher age, education, experience, and an indicator of whether the teacher was nonwhite, non-Hispanic. Additional explanatory variables for the child-level analysis include age at spring assessment and indicators of female; nonwhite, non-Hispanic; whether pretest was taken in Spanish; and whether pretest data are missing. Estimates are weighted to account for the sample and survey designs, and standard errors account for design effects that are due to unequal weighting of the data and clustering at the site level.

We then use the estimated coefficient on each mediator, $\hat{\gamma}_i$, as an estimate of the marginal effect of that mediator on the outcome variable, holding constant the other mediators and explanatory variables. It is important to keep in mind that since this model relies on cross-sectional rather than quasi-experimental variation, the estimated coefficients on the mediators represent correlations rather than causal effects. For instance, if any of the mediating variables included in the model are correlated with another mediator that also affects the outcome but is omitted from the model (for instance, teacher motivation), the true causal effect of that omitted variable on the outcome will be attributed to the estimated coefficients on the included mediators, leading them to be biased estimates of the causal effects of each individual mediator. Nonetheless, these estimates can provide useful descriptive information on the association between each mediator and the outcome variable of interest.

In the *third* stage of this analysis, we use the coefficient estimates from model (2) to compute what we term the "implied impacts" of each mediator on the outcome by multiplying the estimate of ERF's impact on mediator i from equation (1), I_{Mi}, by the coefficient on that mediator from model (2), $\hat{\gamma}_i$. The implied impact of a particular mediator provides an estimate of change in the outcome variable that is attributable to the change that ERF caused in that particular mediating variable. This estimate may be biased, however, because it is unlikely that the relationships estimated between the mediators and the outcome variable in model (2) are true causal relationships.

In the *fourth* stage of this analysis, we compute ERF's total implied impact on the outcome variable, I_Y, as the sum of the implied impacts of ERF on each mediator, plus any residual treatment effects (represented by the estimated coefficient on treatment status, $\hat{\alpha}_1$, from model (2):

(3) $I_Y = \hat{\alpha}_1 + \sum_i I_{Mi}\hat{\gamma}_i$

We can then partition the estimate of ERF's total implied impact on the outcome variable into the percentage due to ERF's impact on each individual mediator and the percentage due to residual factors. Although the total implied impact on the outcome computed in (3) are not mathematically identical to the impacts estimates presented in Chapters 7 and 8, they are very close in practice.

Results of the Analysis of Mediators of ERF's Impacts on Classroom Instructional Practice

We conducted the mediated analysis for 10 measures of classroom practice that were positively affected by ERF—book-reading practices, number of different phonological-awareness activities, print- and letter-knowledge learning opportunities, classroom-print environment, written-expression learning opportunities, opportunities and materials for writing, oral-language use by the lead teacher, oral-language use by the assistant teacher, child portfolios, and teacher sensitivity. Because the primary channels through which ERF aimed to improve language and literacy instruction were professional development and mentoring, the mediating variables that we explore for these classroom-level outcomes are hours of professional development and whether mentoring was provided as a mode of training.

Table 8.2 presents the results of the analysis of mediators of ERF's impacts on each of the 10 measures of classroom instructional practice that we examined. Overall, as shown in the "Total" column, the professional development and mentoring mediators explain less than 20 percent of the total implied impact estimates on each of the 10 measures of classroom practice that we examined; the two mediators are jointly statistically significant only for the child-portfolio and teacher-sensitivity models. For child portfolios, however, the two mediators do not account for any of the total implied impact on the outcome.

The estimated marginal effect of hours of professional development on each of the 10 measures is generally small and not statistically significant. The two exceptions are classroom print environment and teacher sensitivity, on which we estimate positive and statistically significant effects of professional development. Similarly, the estimated marginal effect of mentoring on each of the 10 outcomes is generally small and not statistically significant; the exceptions are negative and statistically significant estimates of the marginal effect of mentoring on child portfolios and teacher sensitivity. The mediators are jointly statistically significant only for child portfolios and teacher sensitivity.

Table 8.2. Hours of professional development as potential mediator of ERF's impacts on classroom instructional practice related to language and literacy

Measures of instructional practice	Estimated marginal effect on instructional practice of:			Percentage of ERF's impact on classroom instructional measure associated with:		Total percentage of ERF's impact on classroom instructional measure associated with professional development
	Professional development hours (p-value)	Received mentoring (p-value)	P-value of joint significance of mediators	Professional development hours	Received mentoring	
Book-reading practices	0.00 (0.077)	0.11 (0.510)	0.113	6.50	5.45	11.95
Number of different phonological awareness activities	0.41 (0.285)	0.00 (0.517)	0.527	13.44	–0.67	5.76
Print and letter knowledge learning opportunities	0.00 (0.626)	0.24 (0.230)	0.343	3.89	14.81	18.70
Classroom print environment	0.00 (0.029*)	–0.17 (0.340)	0.065	33.93	–16.67	17.25
Written expression learning opportunities	0.00 (0.127)	0.22 (0.372)	0.183	7.83	7.13	14.96
Opportunities and material for writing	0.00 (0.976)	0.00 (0.350)	0.649	–0.19	3.21	3.03
Oral language use by lead teacher	0.17 (0.232)	0.00 (0.427)	0.283	11.84	4.56	16.40
Oral language use by assistant teacher	0.00 (0.796)	0.24 (0.365)	0.660	–3.02	14.48	11.46
Child portfolios	0.29 (0.277)	–0.01 (0.000*)	0.000*	19.57	–110.21	–90.65
Teacher sensitivity	0.34 (0.005*)	0.000 (0.012*)	0.006*	21.95	–11.65	10.30
Sample size (number of classrooms)	133					

*p-value < 0.05, two-tailed test.
SOURCE: ERF spring Teacher Behavior Rating Scale and fall teacher survey.

Results of the Analysis of Mediators of ERF's Impacts on Children's Print and Letter Knowledge

As shown in Chapter 7, ERF had a positive impact on children's print and letter knowledge. Table 8.3 presents the analysis of the potential mediators of ERF's impact on print and letter knowledge. As shown in this table, the estimated marginal effects on print and letter knowledge are not statistically significant for any of the potential mediators except print- and letter-knowledge learning opportunities, which account for 27 percent of the total implied impact on print awareness scores. Together, all eight mediators account for 60 percent of the total implied impact on print and letter knowledge and are jointly statistically significant at the 5 percent level.

Table 8.3. Potential mediators of ERF's impacts on print and letter knowledge

Mediator	Estimated marginal effect of mediator on print and letter knowledge	P-value of estimated marginal effect*	Percentage of ERF's impact on print and letter knowledge associated with mediator
Book-reading practices	–0.22	0.731	–4.15
Number of phonological awareness activities	0.38	0.424	12.12
Print and letter knowledge learning opportunities	1.56	0.048*	26.97
Written expression learning opportunities	0.53	0.438	13.88
Classroom print environment	0.70	0.549	8.92
Opportunities and material for writing	0.29	0.821	7.73
Child portfolios	0.42	0.381	10.46
Teacher sensitivity	–1.15	0.303	–15.92
Total		0.015*	60.02
Sample size (number of children)	1,223		

*p-value < 0.05, two-tailed test.
SOURCE: ERF spring Teacher Behavior Rating Scale and spring child assessments.

Appendix A. Impact Analysis Methods and Sensitivity of Results

In this technical appendix, we provide additional methodological details about the ERF impact analysis. In the first section, we describe the analytic methods for the child and classroom impact analyses and the specification of our preferred models (those used to produce the results presented in the main text of this report). In the second section, we present sensitivity analyses of the child impact models, and in the third section, we present analogous information on sensitivity tests of the classroom impact models. In the fourth section, we describe our procedures to adjust for multiple comparisons within outcome domains.

Impact Analysis Methods

The National Evaluation of ERF used a regression discontinuity (RD) design to estimate ERF's impact on children's language and literacy skills and on the quality of language and literacy instruction in the classroom. In this section, we describe several aspects of the analytic methods used to estimate these impacts.

- The regression-discontinuity design
- The statistical model
- Selection of the functional form for the application score
- Selection of covariates
- Sample weights
- Statistical power
- Subgroup analysis

The Regression-Discontinuity Design

The RD design makes use of the scoring process that was used to award the ERF grants. In the FY 2003 ERF grant competition, applications were scored according to predetermined criteria. ED then awarded ERF grants to the grant applicant with the highest application score first and progressed down the score distribution until all funding available for the fiscal year had been allocated. In this way, 30 grants were awarded to the grant applicants with scores of 74 or higher; applicants with scores below 74 were not awarded grants.[62]

This "discontinuity" in grant awards based on the application scores was used to identify ERF impacts. We estimated impacts by using regression models to compare child and classroom outcomes in the funded sites (the treatment group) to those in the unfunded sites (the comparison group), controlling for a smooth function of grant application score. If we assume that the outcome variables exhibit a stable continuous relationship with the application score and that we have correctly modeled this relationship, the sharp discontinuity in ERF grant receipt at the score cutoff, conditional on this smooth function of application score, will identify ERF's impacts.

[62] This design is referred to in the literature as a "sharp" regression-discontinuity design (Trochim, 1984) because treatment status is completely determined by an observed measure.

A related requirement for obtaining unbiased impact estimates under the RD design is that the grant application scores were determined independently of the score cutoff value. Stated differently, the raters must not have manipulated application scores based on their knowledge of the score cutoff value. For instance, if peer reviewers knew the threshold for grant receipt, they might have increased scores for sites with "true" scores below the cutoff value but who the reviewers thought might particularly benefit from the ERF grant. Such strategic behavior by scorers, however, was unlikely because the threshold for determining grant receipt was not determined until after applications had been submitted and scored on the basis of funding availability. This perception is supported by the finding that there is no clustering of sites just above the cutoff value, which would likely occur if raters manipulated the application scores to make their preferred sites barely qualify for grants (McCrary 2005).

Ideally, the RD model would compare sites just above the score threshold for ERF grant awards to sites just below this threshold to ensure that the two sets of sites were as comparable as possible.[63] In the case of the ERF evaluation, however, in order to obtain adequate sample sizes to achieve desired precision levels, we needed to select sites from a fairly broad range of the score distribution. Figure A.1 shows the distribution of grant application scores for the funded and unfunded sites in the study sample. The scores are relatively uniformly distributed, ranging from 42.3 to 73.8 in unfunded sites and 74.2 to 94.7 in funded sites.

Figure A.1. Distribution of grant application scores

[63] See Lee and Card (2006) for a more general discussion of this issue.

A handful of studies have evaluated the performance of the RD design in replicating findings from randomized experiments (Aiken et al., 1998; Buddelmeyer and Skoufias, 2003; Black, Galdo, and Smith, 2005). Aiken et al. and Buddelmeyer and Skoufias found similar impact results using RD and experimental methods. Black, Galdo, and Smith, however, find that their RD estimates are sensitive to the estimation sample and econometric models and in some cases fail to replicate the experimental results. They also found that the RD models that generally performed best were those that restricted the sample to individuals within a very narrow window around the discontinuity point, while models that included a wider range of individuals were more sensitive to the model specification. Given that the RD design for the National Evaluation of ERF needed to include sites from a broad range of the score distribution, we conducted a variety of sensitivity tests to examine the robustness of our results to various model specification decisions.

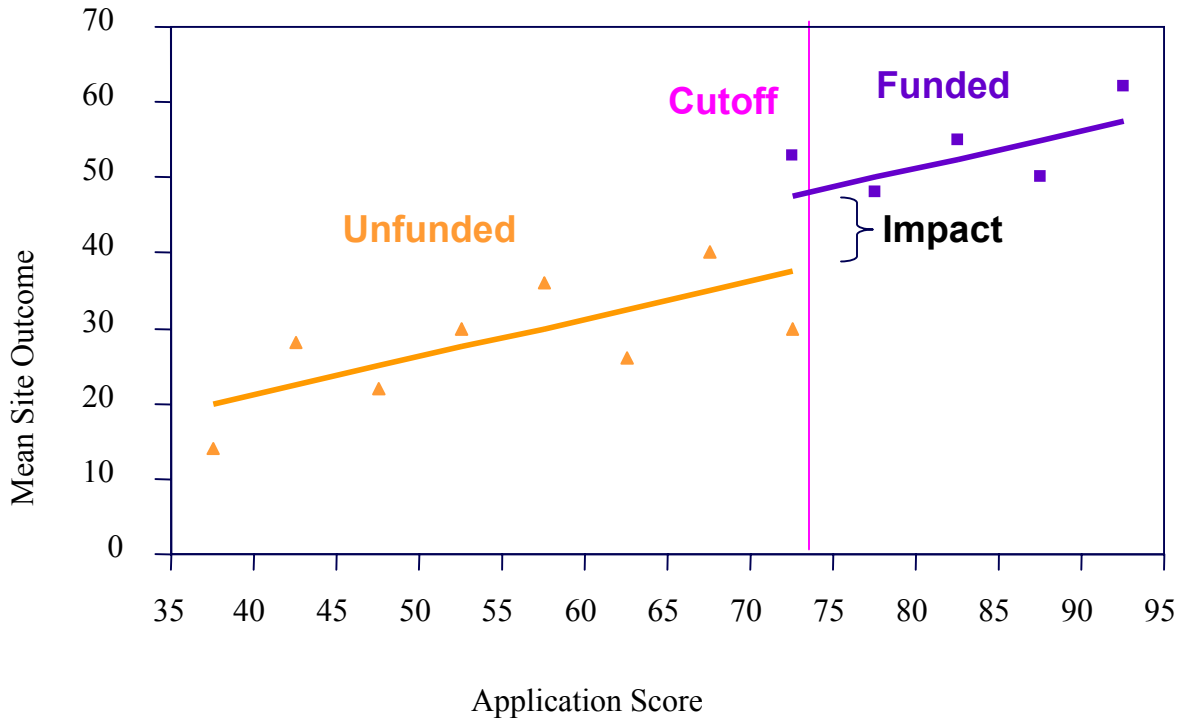
The RD design has implications for the generalizability of the impact estimates. One view is that the impact estimates generalize only to sites that are "similar" to those with application scores just above or below the 74 cutoff and not necessarily to sites with scores farther from 74 or to the average site in the sample. Under this view, the impact estimates are marginal average treatment effects (MATEs) (Bjorklund and Moffitt 1987, Heckman 1997) that represent mean impacts for sites at the margin of ERF funding receipt.

Another view is that if a parametric specification is used for the functional form for *Score*, the fitted regression lines for the treatment and comparison groups can each be "extrapolated" to obtain impact estimates for sites with alternative *Score* values. Estimates of average treatment effects (ATEs) can then be obtained and can be written as weighted averages of MATEs over the full support of *Score* (Heckman and Vytlacil 1999).[64] This approach, however, hinges critically on the extent to which modeling assumptions apply to the full *Score* distribution and could lead to anomalous results. For instance, if the slopes of the regression lines differ for the funded and unfunded sites, then "extrapolated" impacts would be positive for some *Score* values and negative for others.

Before presenting the mathematical framework for estimating impacts, we illustrate the estimation approach graphically for a hypothetical child or classroom outcome. Figure A.2 plots the mean outcome at the site level against the site application score. The figure also displays the fitted regression lines for the unfunded and funded sites, where, for simplicity, the slopes of the two regression lines are assumed to be the same (although this assumption can be relaxed). The estimated impact is the vertical difference between the two regression lines at the cutoff score value of 74 (that is, at the point of discontinuity). In contrast, a simple comparison of mean outcomes across the funded and unfunded sites that does not account for the relationship between score and the outcome will yield biased impact estimates, and thus, standard estimation procedures that are typically used for random assignment designs are *not* applicable for RD designs. Unlike a random assignment design, treatment and comparison sites under an RD design are—*by construction*—likely to have different baseline characteristics and, thus, are *not* directly comparable without conditioning on the appropriate function of application score.

[64] If treatment effects are homogeneous for all *Score* values, then MATE and ATE parameters are the same.

Figure A.2. The RD method with hypothetical data points and estimated regression lines



Parametric Statistical Model

We used a hierarchical linear modeling framework (Raudenbush and Bryk 2002) to estimate impacts under the RD design in our preferred models. This framework accounts for the clustering of children within classrooms and sites in the variance calculations.[65] We used regression models to estimate impacts, controlling for functions of the application score.

The hierarchical linear model for a *child* outcome consists of three levels that are indexed by children (*i*), classrooms (*c*), and sites (*s*):

$$(1)\quad \begin{aligned} &Level\ 1: Students: && Y_{ics} = \alpha_{0cs} + e_{ics} \\ &Level\ 2: Classrooms: && \alpha_{0cs} = \gamma_{00s} + u_{0cs} \\ &Level\ 3: Sites: && \gamma_{00s} = \lambda_0 + \lambda_1 T_{00s} + f([Score_{00s} - 74], T_{00s})\theta + \eta_{00s}, \end{aligned}$$

where Y_{ics} is a child outcome measure; α_{0cs} is a classroom-level random intercept; γ_{00s} is a site-level random intercept; T_{00s} is an indicator variable equal to 1 for funded sites and 0 for unfunded sites; $f([Score_{00s}\text{-}74], T_{00s})$ is a vector containing polynomial functions of the application score (centered at the 74 cutoff value) and terms formed by interacting T with the *Score* variables; e_{ics} are assumed to be *iid* $(0, \sigma^2_e)$ child-level random error terms; u_{0cs} are *iid* $(0, \sigma^2_u)$ classroom-

[65] We discuss nonparametric estimation approaches later in this appendix.

specific error terms that capture the correlation between the outcomes of children in the same classrooms; η_{00s} are *iid* $(0, \sigma^2_{\eta})$ site-specific error terms that capture the correlation between the outcomes of children in the same sites; and λ_0, λ_1, and θ are fixed parameters to be estimated. The random error terms across equations are assumed to be distributed independently of each other.[66]

For ease of presentation, we hereafter refer to the following single-equation version of the hierarchical linear model (see, for example, Murray 1998) by recursively inserting the Level 2 and 3 equations into the Level 1 equation and also adding to the model a vector of child-, classroom-, and site-level baseline covariates, *X*, that can increase precision by explaining some of the variation in outcomes between units:

$$(2) \quad Y_{ics} = \lambda_0 + \lambda_1 T_{00s} + f([Score_{00s} - 74], T_{00s})\theta + X_{ics}\beta + [e_{ics} + u_{0cs} + \eta_{00s}].$$ [67]

In this formulation, the estimate of the parameter, λ_1, is the regression-adjusted impact estimate and represents the difference between the intercepts of the fitted regression lines (curves) for the treatment and comparison groups. T-tests are used to gauge the statistical significance of the impact estimates, which are less precise under the RD design than would be the case under a simple random-assignment design because of the substantial correlation between *T* and the *Score* terms. This design effect is about 3.75. The SAS procedure, PROC MIXED, was used to estimate equation (2).[68]

To estimate impacts for *classroom* (teacher) outcomes in our preferred models, we employed a 2-level hierarchical linear model where Level 1 pertains to classrooms and Level 2 to sites. For these outcomes, we estimated a variant of the model in equation (2) by dropping the child-level subscript (*i*) from all terms and omitting the child-level error terms (e_{ics}).

Selection of the Functional Form for the Application Score

The statistical model in equation (2) produces unbiased and internally valid impact estimates if the functional form of the continuous relationship between *y* and *Score* is correctly specified. The functional form for *Score* in equation (2) can include linear, quadratic, or higher order *Score* terms, as well as terms formed by interacting *T* with the *Score* variables. The appropriate functional form depends on the true relationship between the application scores and the outcomes of interest and could vary by outcome. Determining the appropriate functional form is a particularly important issue for the ERF study, given the broad range of scores for the sites in our sample.

We used several methods to assess the appropriate functional form for each outcome measure: (1) graphically inspecting the relationship between the application score and the average value of

[66] The model does not account for preschool-level clustering, because there was no sampling of preschools; rather, classrooms were sampled with probabilities proportional to size without regard to the preschool where they were located.

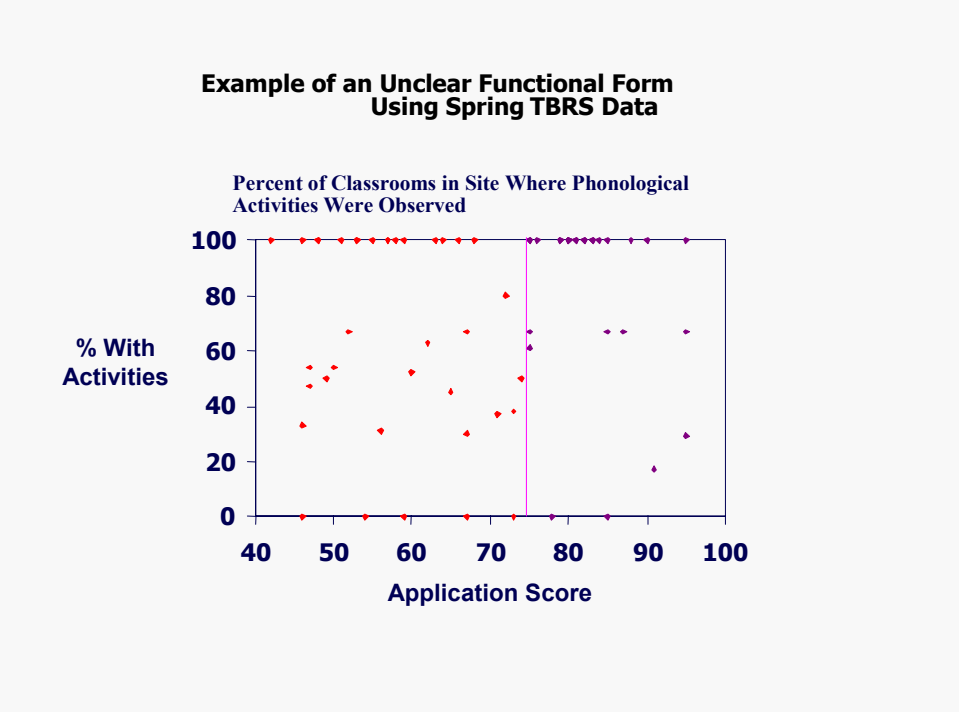
[67] For simplicity, we use Level 1 subscripts for the vector, *X*, although the vector can also include Level 2 and 3 covariates.

[68] The impact estimates obtained by using alternative statistical packages are similar to those obtained using PROC MIXED.

the outcome measure in each site, (2) gauging the statistical significance of the *Score*-related polynomial and interaction terms, and (3) conducting several specification tests found in the literature that are presented with the sensitivity analyses later in this appendix. Based on these examinations, we used a *linear* function of *Score* and no interaction terms for the child and classroom outcomes in our preferred models. The impact findings are robust to alternative functional-form specifications, as shown in the sensitivity analysis.

For some classroom outcomes, it was difficult to identify the correct functional-form specification. These variables tend to be binary outcomes that are typically either always 1 or always 0 within a site and include whether specific phonological awareness activities were observed in the classroom and whether the teacher used specific curricula or child assessments.[69] Figure A.3 provides an example of such a binary outcome—whether or not any of seven phonological awareness activities were observed in the classroom—whose mean value at the site level is plotted against site-application scores. Because many site-level values are either 0 or 100 percent for both the treatment and comparison groups, it is difficult to identify the correct functional form specification for *Score*. Furthermore, it is problematic that the impact estimates for these types of outcomes vary substantially by specification and thus are not robust. Thus, we do not present impact estimates for most of these outcomes. (Impacts on whether specific phonological awareness activities were observed are presented in Appendix D; however, we note that these estimates may not be robust.)

Figure A.3. Example of an unclear functional form relationship: whether any of seven phonological awareness activities were observed in the classroom in the spring



[69] These outcomes would not pose a problem under a random assignment design; they pose a problem under the RD design because of the modeling process that is required to obtain unbiased impact estimates.

Selection of Covariates

Under the RD design, the inclusion of baseline covariates in equation (2) is not required to obtain unbiased impact estimates if the *Score* variable fully reflects the selection rule used to award ERF funds and if we have correctly modeled the relationship between the outcome and *Score*. However, baseline covariates can increase the precision of the impact estimates to the extent that they are correlated with the outcome variables. Improving power is an important issue for the ERF evaluation because of large design effects from clustering and the RD design. In addition, covariates can adjust for residual differences between the baseline characteristics of those in the funded and unfunded sites (conditional on the appropriate function of the application score).

The use of baseline covariates in the ERF evaluation poses several analytic challenges. First, the fall child assessments and classroom observations do not yield "true" baseline measures. This is problematic because in most school-based experimental evaluations, pre-intervention measures of the outcome variables (pretests) are typically the most important predictors of corresponding postintervention measures (posttests) and, thus, are important for improving precision. Second, for some model covariates, the impact results become sensitive to the functional form specification for the application score. These covariates are typically binary variables that vary substantially across sites and are difficult to model as a function of *Score*. Thus, it is difficult to assess the true correlation between these covariates and treatment status, conditional on *Score*.

To address these issues, we adopted a conservative approach for including covariates in our preferred models. Specifically, we selected covariates according to two criteria: (1) their inclusion should not materially change the impact findings relative to models that exclude the covariates; and (2) they should have predictive power in the regression models. We include a limited set of covariates in our preferred models and more extensive sets of covariates in our sensitivity analysis to examine the robustness of study findings. We also estimated models without covariates.

Our preferred models for the *child outcomes* included a limited set of demographic covariates:

- indicators of whether the child is female
- whether the child is white and non-Hispanic
- whether fall assessment data were missing
- age at spring assessment
- whether the fall assessment was taken in Spanish (for language and literacy outcomes)

Some models also included fall assessment scores as covariates (see the following subsections). Our preferred models for the *classroom outcomes* included the following covariates: teacher education level (in years), teacher age, and whether the teacher is white and non-Hispanic.

The following subsections discuss

- our approach for using the fall assessment scores in the analysis because of their importance in improving precision

- our approach for imputing missing covariates

Baseline Assessments

The fall *child* assessments are not true baseline measures. Due to various constraints, the first round of assessments was not conducted until one to four months after the school year began, at a point when all children had started their preschool year and the treatment group had already received some exposure to the intervention. Furthermore, because of challenges in recruiting unfunded sites, the assessments were typically conducted earlier in the funded sites than in the unfunded sites. Thus, including fall assessment scores as covariates in the model could bias the impact estimates because the fall assessment scores may be correlated with treatment status. For instance, if ERF had a positive impact on child outcomes within the first four months of the school year, this effect would be incorrectly attributed to differences in baseline abilities, and the impact estimate for the spring outcomes would be biased downward. Alternatively, if average fall assessment scores were higher in the comparison group than the treatment group simply because the comparison group was tested later in the school year, impact estimates for spring outcomes may be biased upward.

We adopted a conservative approach for including the fall assessment scores as covariates in our preferred child-level models, recognizing the tradeoff between bias and precision. If there is no statistical evidence of a difference in fall assessment scores between the funded and unfunded sites, then we present results that both include and exclude that fall assessment score as a covariate. This is the case for the Elision and expressive vocabulary skill scores and the three behavioral outcomes. However, if there is evidence of a difference in fall assessment scores, then we present only results that exclude that score as a covariate. This is the case for the print and letter knowledge and auditory comprehension outcomes. Although impacts on these fall assessment scores are not statistically significant, the point estimates appear larger than what one might expect by chance. Furthermore, positive impacts on spring posttests were found for these outcomes, suggesting that the fall assessment scores could be capturing early treatment effects. Thus, we are concerned that the inclusion of these fall assessment scores in the regression models could lead to impact estimates that are biased downward.

The fall *teacher* and *classroom* assessments are also not true baseline measures. Because ERF classrooms were expected to reach full implementation by September 2004, key training activities occurred during the spring and summer *before* the start of the school year. Thus, teacher and classroom outcomes should have already been affected by ERF at the time of the fall data collection (which would be the case even if the assessments were conducted at the start of the school year). Consequently, we treat the fall teacher assessments as outcome measures rather than baseline measures, and thus, we do *not* include them as covariates in the regression models.

Imputation of Missing Values of Covariates

For our preferred models, we imputed missing values of covariates by assigning the mean value of the covariate by site and gender for the child-level analysis and by site for the classroom-level analysis. If covariates were missing for an entire site, we assigned the mean value of the covariate by treatment status and gender for the child-level analysis or by treatment status for the classroom-level analysis. Thus, we estimated the regression models by using all available outcome data; we did not exclude children or teachers from the analysis with available outcome data who were missing covariates.

In our sensitivity analysis, we adopted other methods for handling missing data. For instance, we estimated models by using only cases that had no missing data, and for child impact models, we also used a hot-deck imputation procedure.

Sample Weights

To obtain our preferred estimates, we used sample weights for the following reasons:

- ***To account for the random selection of classrooms to the analysis sample.*** Within each site, we selected classrooms with probabilities proportional to classroom size.

- ***To give equal weight to each site.*** Because sites are the unit of analysis, we gave each site equal weight in the analysis, regardless of the number of sample members per site.

- ***To account for study nonconsent and interview nonresponse (for the child-level weights).*** We could not use data on baseline child characteristics to construct weights that adjust for study nonconsent and nonresponse, because these data are not available for nonconsenters. Instead, we constructed weights to be proportional to the combined consent and response rate within each classroom. This approach assumes that children in a specific classroom who have follow-up data are representative of all children in that classroom.[70]

We begin this section with a discussion of the construction of base weights to account for the sample design. We then discuss our adjustment of these weights (to account for study nonconsent and interview nonresponse in the child-level analysis) and our normalization of the weights (to give equal weight to each site in the analysis).

Weights to Account for the Sample Design

Under the ERF sample design, classrooms and children had differing probabilities of being selected into the study sample. Classrooms were randomly selected into the study sample from the full list of participating classrooms in the funded and unfunded sites. The classrooms were selected with probabilities proportional to the number of 4-year-olds who were estimated in late spring and summer 2004 to have been enrolled in each classroom in fall 2004. An ordered list of classrooms was created to replace initial selections when either the school director or teacher of the selected class refused to participate. Site recruiters negotiated participation with the individual schools and teachers, replacing selected classrooms as necessary at this stage by moving sequentially down the ordered lists. When agreement on the details of participation had been reached with each classroom and school, information on the specific classes to be included was sent to the data collection staff.

[70] In the sensitivity analysis, we also estimated impacts using weights that do not account for nonconsent and nonresponse and found very similar results to the preferred models.

The eligible child population for the study consists of 4-year-old children in their pre-kindergarten year. However, many classes selected into the sample included both 3- and 4-year-old children, and data on the ages of individual children were not available before parental consent was requested. Therefore, consent forms were distributed to *all* children in the selected classes, and parents provided the child's birth date when they returned the signed consent form. From the list of consenting children, the study team determined which children were eligible for the study based on age and the local cutoff date for entering kindergarten. From the list of eligible children, the team randomly selected up to 15 children into the sample for assessment and parent-survey data collection. In some classes, data collectors selected replacement children because one or more consenting children were unable to complete the assessment (due to language difficulties or disability) or unavailable (due to absence). In classrooms with less than 15 eligible consenters, all eligible consenting children were selected.

To account for the different probabilities of selection into the study sample for each child and classroom in the study, we constructed base weights reflecting the inverse of the probability that each was selected. The base classroom weight for classroom *c*, $baseclassweight_c$, was calculated as follows:

(1) $\text{baseclassweight}_c = 1/[\text{P(class selected)}_c] = 1/[\text{selprob}_c]$,

where:

$selprob_c$ is the probability a class was selected to the sample, equal to $max(n\_classes\_needed_g * n\_4yo_c / n\_4yo_s, 1)$

$n\_classes\_needed_s$ = number of classes needed for sample in site *s*

n_4yo_s = number of 4-year-olds in site *s* at time classes were sampled

n_4yo_c = number of 4-year-olds in class *c* at time classes were sampled

Similarly, the base weight for child *i*, $basechildweight_i$, was calculated as follows:

(2) basechildweight_i

$= 1/[\text{P(class selected)}_c * \text{P(child selected from consenters|class selected)}_c]$

$= 1/[\text{selprob}_c * (\text{n\_selected}_c/\text{n\_elig}_c)]$,

where:

n_elig_c = number of eligible consenting 4-year-olds in classroom *c*

$n\_selected_c$ = number of eligible consenting 4-year-olds selected into sample in classroom *c*

Weights to Account for Study Nonconsent and Interview Nonresponse

Some teachers and children selected into the sample refused to participate in the study, and some consenters did not complete the various surveys, assessments, and observations. Ideally, we would adjust the sample weights to account for differential probabilities of consent and response using detailed baseline data. For classrooms, however, there is little information to construct these adjustments, so we did not adjust the base classroom weights. For children, there is also very little information on those who did not consent. However, if we are willing to assume that child nonconsent and nonresponse was *random* within a classroom and the same for both 3- and 4-year olds, we can construct an adjusted weight, *adjwgt*, for each child outcome (assessment or SCBE observation) and time period (pre or post) as follows:

(3) $adjwgt_c$

$$= 1/[P(\text{class selected})_c * P(\text{child a consenter}|\text{class selected})_c * P(\text{child selected}|\text{eligible consenter in selected class})_c * P(\text{child responded}|\text{selected})]$$

$$= 1/[selprob_c * (n\_consent_c/n\_children_c) * (n\_selected_c/n\_elig_c) * (n\_responded_c/n\_selected_c)]$$

where:

$n\_children_c$ = number of 3- and 4-year-olds in classroom *c*, as reported by teacher[71]

$n\_consent_c$ = number of consenting 3- and 4-year-olds in classroom *c*[72]

n_elig_c = number of eligible consenting 4-year-olds in classroom *c*

$n\_responded_c$ = number of responders in classroom to outcome (parent survey, assessment, or SCBE) in particular time period (pre or post)

$n\_selected_c$ = number of eligible consenting 4-year-olds selected into sample in classroom *c*

The nonresponse weights require the assumption that nonresponse was random within a classroom and the same for both 3- and 4-year-olds. Given that there is no demographic data for the full sample frame to use to predict response probabilities, this was the only feasible approach.

[71] In a few cases, the number of consenters exceeded the number of children as reported by the teacher. In these cases, we replaced *n_children = n_consent*.

[72] In a handful of cases (3.4 percent of total), the reported number of eligible children exceeded the number of consenters. In these cases, we redefined *n_consent = max(n_eligible, n_consent)* because in all cases, *n_consent* was a binding upper limit on *n_selected*. In no case did the number selected exceed the number of consenters or number eligible.

Normalization of Weights

Since the relevant unit of analysis for the evaluation is the site, we rescaled all child and classroom weights to give equal weight to each site in the impact estimates, regardless of the size of the site. Thus, the adjusted child weights were normalized and scaled to sum to the average number of 4-year-olds per site. The normalized child weights, *normadjwgt*, were calculated as follows:

$$(4) \quad normadjwgt_i = \left(adjwgt_i / \sum_{i \in s} adjwgt_i \right) * \left(\sum_{s \in S} n\_4yo_s / n\_sites \right)$$

The base classroom and child weights, $baseclassweight_c$ and $basechildweight_i$, respectively, were similarly normalized to give equal weight to each site.

The normalized weights, $normadjwgt_i$, serve as the benchmark weights for the child-level analysis, while the normalized child base weights are used for sensitivity testing. The normalized classroom base weights serve as the benchmark weights for the classroom analysis.

Statistical Power

To assess statistical power of the preferred impact estimates for the ERF evaluation, we calculated minimum detectable impacts in effect-size units (MDEs) for child and classroom outcomes. MDEs represent the smallest impacts in effect-size units that can be detected with a high probability (80 percent in our case). The MDEs are primarily a function of study sample sizes, the degrees of freedom available for statistical tests, and design effects from the RD design (which is about 3.75) and clustering.[73] Clustering effects are measured by intraclass correlations (ICCs) that reflect the percentage of the total variance in the outcomes that is between sites and between classrooms within sites. Table A.1 displays, for key child and classroom outcomes, ICCs from equation (2) that do not include fall assessment scores as covariates but do include several other covariates, and ICCs adjusted for fall assessment scores (for the child outcomes only).[74] Table A.2 displays MDEs for a typical child and classroom outcome (assuming a 2-tailed test and a 5-percent significance level) and the MDE formula used in the calculations.

The ICCs for the child outcomes are about 1.5 percent at the site level and 2.5 percent at the classroom level when the model excludes fall assessment scores as covariates; the ICCs are slightly smaller when the fall assessment scores are included as covariates (see Table A.1). This

[73] The design effect under the RD design depends largely on the distribution of the application scores. If the scores were normally distributed, then the design effect would be 2.75. However, the scores are much closer to a uniform distribution, which leads to an actual design effect of 3.75. The design effect was calculated as follows:

$$(1) \quad Design\ Effect = \frac{(1 - R2_1)}{(1 - R2_0)} * \frac{1}{(1 - R2_{T|Score})},$$

where $R2_1$ is the regression R^2 value when the outcome is regressed on *T* and *Score*, $R2_0$ is the regression R^2 value under an experimental design, and $R2_{T|Score}$ is the R^2 value when *T* is regressed on *Score*.

[74] As discussed in Chapter 6 and 7, the preferred models for the child outcomes include as covariates a linear function of *Score*; indicator variables of female and nonwhite; and, for the language and literacy outcomes, an indicator variable of whether the fall assessment was given in Spanish. All models for the teacher outcomes include as covariates a linear function of the application score; teacher education level; age; and indicators of white non-Hispanic.

suggests that mean child outcomes do not vary substantially across sites or classrooms. The ICCs, however, are much larger for classroom outcomes (about 33 percent).

For the full sample of 65 sites, the MDE (unadjusted for the fall assessment scores) is about 0.30 standard deviations for a typical child outcome and is 0.89 standard deviations for a typical classroom outcome (see Table A.2).[75] For a 50-percent subgroup of children, preschools (classrooms), or sites, the MDEs for the child outcomes range from about 0.38 to 0.42.[76]

It is important to note that these MDEs were calculated at 80-percent power. Thus, it is possible to find a statistically *significant* impact on an outcome if the true impact on that outcome is smaller than the relevant MDE, although the chance that this will occur is less than 80 percent. Similarly, it is possible to find a statistically *insignificant* impact on an outcome if the true impact on that outcome is larger than the relevant MDE, although the chance that this will occur is less than 20 percent.

Table A.1. Intraclass correlations for key child and classroom outcomes

	ICCs Not Adjusted for Fall Assessment Scores[a]		ICCs Adjusted for Fall Assessment Scores[a]	
Outcome	Site Level	Classroom Level	Site Level	Classroom Level
Child Outcomes				
Print and Letter Knowledge	.027	.016	.014	.012
Elision	.005	.008	.008	.010
Expressive Vocabulary, Raw Score	.011	.020	.007	.019
Expressive Vocabulary, Standard Score	.010	.018	.006	.017
Auditory Comprehension, Raw Score	.017	.011	.016	.009
Auditory Comprehension, Standard Score	.017	.008	.013	.011
Social Competence	.012	.061	.007	.053
Anxiety-Withdrawal	.005	.047	.010	.039
Anger-Aggression	.010	.020	.004	.028
Classroom Outcomes: Teacher Behavior Rating Scales				
Book Reading	.247	—	—	—
Sensitivity Behaviors		—	—	—
Classroom Organization	.389	—	—	—
Phonological Activities	.483	—	—	—
Oral Language	.333	—	—	—
Team Teaching	.370	—	—	—
Math Concepts	.328	—	—	—
Center Activities	.468	—	—	—
Print and Letter	.381	—	—	—
Written Expression	.412	—	—	—
Lesson Plans	.341	—	—	—

[75] For comparison, to achieve the same MDE under a comparable random-assignment design would require a sample of only 17 sites (65/3.75).

[76] The subgroup MDEs for children, preschools, and sites are similar due to the relatively small ICCs.

Notes from Table A.1

[a] All models for the child outcomes include as covariates a linear function of the application score; indicator variables of female and nonwhite; and, for the language and literacy outcomes, an indicator variable of whether the fall assessment was given in Spanish. All models for the teacher outcomes include as covariates a linear function of the application score and teacher education level, age, and an indicator for white, non-Hispanic.
— = Not applicable.
NOTE: All estimates were calculated with sample weights.
SOURCE: ERF spring assessments and observations.

Table A.2. Minimum detectable impacts in effect size units (MDEs) for a typical child and classroom outcome

	MDEs unadjusted for fall assessment scores	
Sample	Child outcome	Classroom outcome
Full sample	0.30	0.79
50 percent subgroup		
Children	0.38	—
Preschools or classrooms	0.39	1.04
Sites	0.42	1.30

— = Not applicable.
NOTE: The MDE formula used in the calculations for a child outcome is as follows:

$$MDE = 2.802 * \sqrt{3.85} * \sqrt{\rho_1(\frac{1}{s_T} + \frac{1}{s_C}) + \rho_2(\frac{1}{s_T k_T} + \frac{1}{s_C k_C}) + (1 - \rho_1 - \rho_2)(\frac{1}{s_T k_T n_T} + \frac{1}{s_C k_C n_C})},$$

where s_T (28) and s_C (37) are the number of treatment and comparison sites in the sample, respectively; k_T (3.2) and k_C (3.2) are the average number of classrooms per site; n_T (8) and n_C (8) are the average number of children per classroom; ρ_1 (.015) is the intraclass correlation (ICC) at the site level; and ρ_2 (.025) is the ICC at the classroom level.

The MDE formula used in the calculations for a teacher outcome is as follows:

$$MDE = 2.802 * \sqrt{3.85} * \sqrt{\rho_{1a}(\frac{1}{s_T} + \frac{1}{s_C}) + (1 - \rho_{1a})(\frac{1}{s_T k_T} + \frac{1}{s_C k_C})},$$

where ρ_{1a} (.33) is the site-level ICC.

Subgroup Analysis

We estimated ERF impacts for several subgroups defined by key child, preschool, and teacher characteristics. The results of the classroom-level subgroup analyses are presented in Appendix E and the results of the child-level subgroup analysis are presented in Appendix F. We selected subgroups by using two criteria. First, we selected subgroups across which we hypothesized that ERF impacts could differ based on theories of change and impact results from previous evaluations of early childhood interventions. Second, due to statistical power considerations, we selected only subgroups with relatively large population shares.

Subgroup Definitions

The examined subgroups differed somewhat for the child and classroom outcomes. For the child outcomes, we estimated impacts for the following demographic subgroups:

- ***Gender.*** Research on early childhood development typically considers the possibility of variations by gender, and gender differences in verbal ability are widely believed to exist, although a careful review of the extensive empirical evidence suggests little or no verbal advantage for girls (Hyde and Linn 1988). We examined ERF impacts by gender to evaluate whether the program is more effective for boys or for girls.

- ***Race and ethnicity.*** Examining impacts by race and ethnicity helps to address whether the program has a greater effect for children of color and therefore whether it helps make progress toward closing the achievement gap.

- ***Primary language spoken at home.*** Children who are English-language learners (ELLs) may make slower progress toward English vocabulary and early literacy skills because they are also learning basic English. Examining impacts separately for children whose home language is English compared to those whose home language is not English can show whether the program's impacts differ for these groups.

- ***Parental education.*** Parents who have more education tend to expose children to a greater variety of language and books in the home, so estimating impacts by parental education helps to address whether the program is providing more compensatory support for children whose parents have less education compared to those whose parents have more education.

For both the child and classroom outcomes, we estimated impacts for the following program-related subgroups:

- ***Whether the preschool received Head Start funding.*** Head Start programs require lower levels of teacher education than some state-funded preschool programs and provide more comprehensive child and family services. Furthermore, the Head Start program implemented an early-childhood literacy initiative in 2002. Thus, looking separately at child and classroom outcomes in Head Start programs versus other programs addresses the effectiveness of implementing ERF in Head Start settings compared to other settings that might differ in teacher education, their service focus, and teacher training on early literacy activities (Frank Porter Graham Center, 2004, U.S. Department of Health and Human Services, May 2004, Irish, Schumacher, and Lombardi, 2004, Ackerman and Barnett, 2006).

- ***Whether the preschool offered full-time or part-time classes.*** Examining child impacts by full-time (30 hours per week) or part-time status provides a rough measure of whether the potential intensity of children's exposure to the ERF program makes a difference in the program's effectiveness, keeping in mind that children in a full-time program may attend only part time.

Finally, for the classroom outcomes, we estimated impacts by teacher education and experience. Early childhood policymakers and researchers are debating the importance of a bachelor's degree for preschool teachers. Thus, examining impacts on the quality of the early language and literacy environment in the classroom by whether or not the teacher has a bachelor's degree helps address whether more-educated teachers change their practice to a greater degree than teachers with less education when they are provided the resources and requirements of ERF. Examining impacts by teacher experience (5 years or more of preschool experience) addresses whether ERF is implemented more easily by newer or by veteran teachers.

Estimation

We obtained subgroup impact estimates by including in equation (2) the terms formed by fully interacting the subgroup indicator variables with the treatment status indicator variable (T), the

specified function of grant application score, and all other covariates. We used these fully interacted models to take into account clustering of children within sites and classrooms (and the clustering of classrooms within sites) across subgroups. We conducted t-tests to determine the statistical significance of impact estimates for each subgroup and conducted F-tests to jointly determine whether impacts differed across levels of a subgroup—for example, across blacks, whites, and Hispanics.

Sensitivity Tests of Child Impact Models

Our preferred specification of the child-impact models controls for a linear function of *Score* along with a limited set of covariates and accounts for design effects due to clustering at the site and classroom levels. Missing values of covariates are imputed, and estimates are weighted to account for the sample design. In this section, we present the results of sensitivity tests to examine the robustness of the child-impact findings to variations in key parameter assumptions. We find that the pattern of child impacts is generally robust to a variety of model specifications. We discuss these alternative specifications in greater detail in this section.

Functional Form Specification for *Score*

We used the following methods to assess the appropriate functional form of the relationship between *Score* and each child outcome measure:

- We graphically inspected the relationship between Score and the average value of the outcome measure in each site.
- We gauged, in the regression models, the statistical significance of polynomial Score variables and terms formed by interacting the Score variables with the treatment status-indicator variable.
- We conducted the following specification tests that use the relation that under the correct specification:
    - There should be few "impacts" on baseline variables.
    - The inclusion of indicator variables pertaining to "artificial" (false) cutoff values as covariates in the model should all be statistically insignificant.
    - The model should fit better (have a higher R2) when the treatment status indicator variable is defined at the actual Score cutoff value of 74 than if it is defined at any other artificial (false) cutoff value.

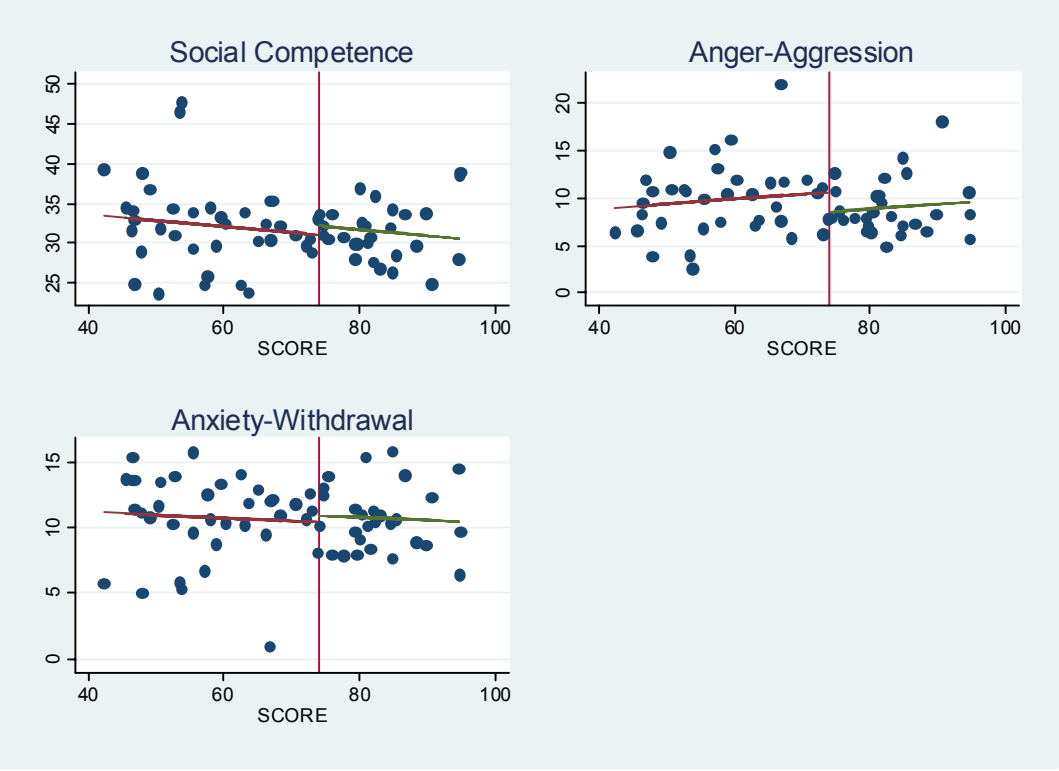
These analyses suggest that the appropriate functional form for the application score for the child impact models is a *linear* function. However, the impact results are robust to alternative functional form specifications.

Graphical Inspection

Figures A.4 and A.5 display plots of site-level mean outcomes versus a linear function of *Score* for seven key child outcome measures.

Figure A.4. Literacy and language skills as a function of *Score*

Figure A.5. SCBE behavioral scales as a function of *Score*



Figures A.6 and A.7 display plots of site-level mean outcomes versus a quadratic function of *Score*.

Figure A.6. Literacy and language skills as a function of *Score* and *Score*-squared

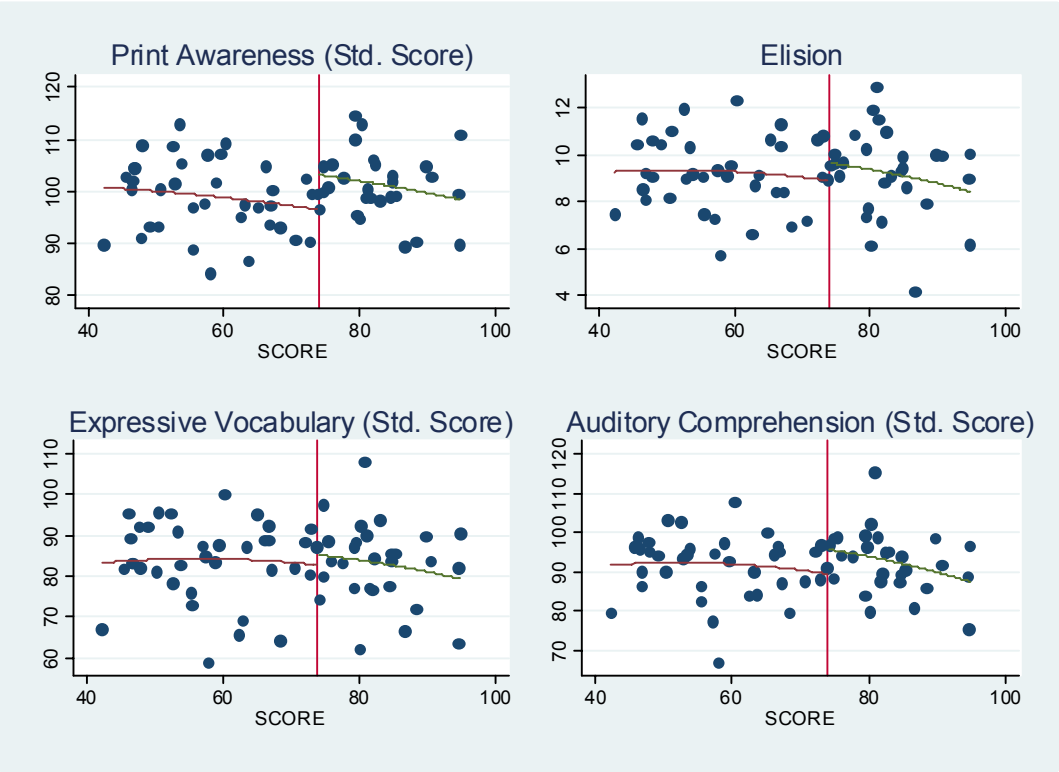
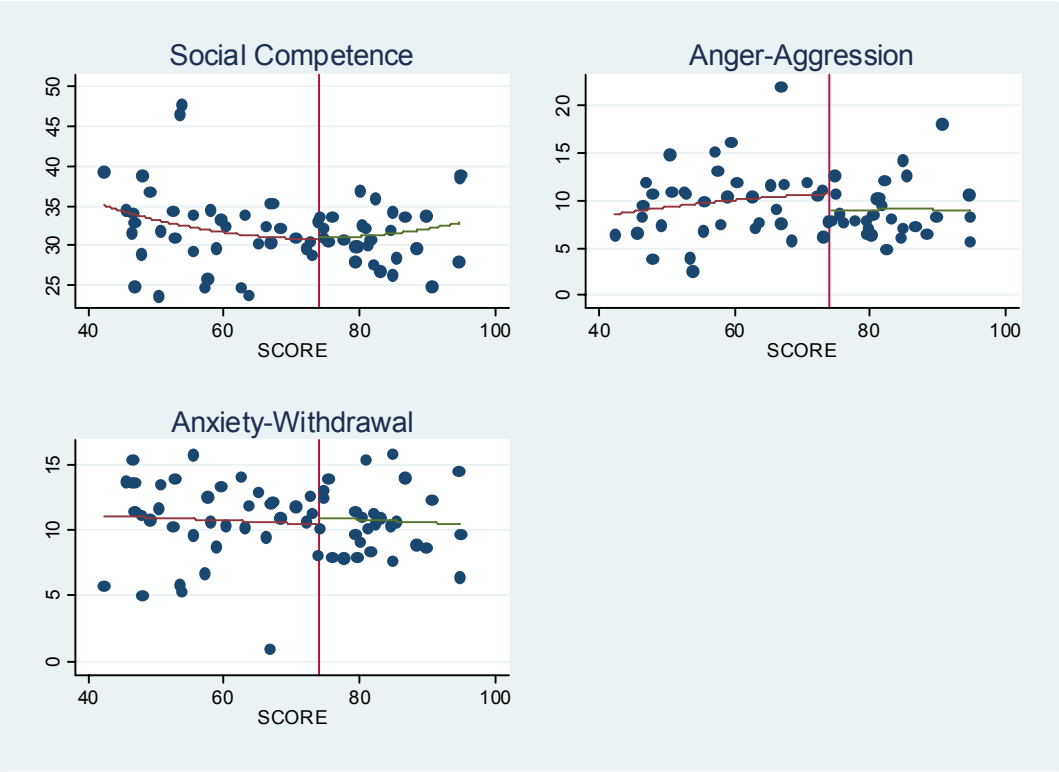


Figure A.7. SCBE behavioral scales as a function of *Score* and *Score*-squared



For six of the seven outcomes, the graphs suggest that a simple linear relationship is appropriate. Furthermore, in the regression models, the estimated polynomial *Score* and interaction terms are not statistically significant at the 5-percent level for any of the outcomes (not shown). For the remaining outcome variable—the SCBE social competence scale—the relationship appears to be quadratic in *Score* (and the quadratic term is statistically significant at the 6-percent level). For simplicity of exposition, however, in our preferred models, we controlled for a linear function of *Score* for all child outcome variables; although the true functional form of the relationship between the social competence scale and *Score* appears to be quadratic, the impact estimates are virtually identical across the two models.

Examining Differences in Baseline Variables

Conditional on the appropriate function of *Score*, there should be few differences between the baseline characteristics of those in the treatment and comparison groups. The strongest specification test would be to examine "impacts" on baseline values of the outcome measures. However, as discussed in Chapters 2, fall assessments were conducted one to four months into the school year and are not true baseline values. Therefore, we cannot use the fall assessment scores to assess the model specification.

We can, however, assess the correct model specification by using data on baseline demographic characteristics of students and sites. Tables A.3, A.4, and A.5 present mean values of key demographic variables in the funded and unfunded sites (columns 1 and 2); differences in these mean values (column 3); and differences in mean values conditional on a linear function of *Score* (column 4), a quadratic function of *Score* (column 5), and a cubic function of *Score* (column 6). The demographic characteristics include child characteristics (such as gender, race and ethnicity,

and age); caregiver characteristics (such as the receipt of public assistance, marital status, number of years in the U.S., education level, and household income); and site characteristics (such as urban or rural status, median income, poverty rate, and unemployment rate).

Under the linear specification for *Score*, there are very few statistically significant baseline differences between the funded and unfunded sites. Of the 45 tests conducted, only 1 is statistically significant at the 5-percent level, which is less than the 2 that we would expect to occur by chance. Under the quadratic specification, however, the baseline differences are statistically significant for 6 variables. Thus, these results further suggest that the linear function of *Score* is appropriate for the analysis.

Table A.3. Characteristics of children in funded and unfunded sites, adjusted for differences in grant applicant score: main covariates (percentages, unless otherwise noted)

	Means		Raw difference		Difference conditional on Score		Difference conditional on quadratic in Score		Difference conditional on cubic in Score	
	Funded	Unfunded	Difference	P-value	Difference	P-value	Difference	P-value	Difference	P-value
Female	49.6	50.2	–0.7	0.783	–1.8	0.653	–1.0	0.821	–1.3	0.770
Child's race/ethnicity (may select multiple categories)										
Black, non-Hispanic	29.1	32.5	–3.4	0.640	4.8	0.736	9.1	0.523	6.2	0.731
White, non-Hispanic	26.8	31.0	–4.2	0.525	6.6	0.614	14.3	0.209	12.8	0.344
Hispanic	41.8	34.5	7.3	0.377	–14.8	0.277	–20.5	0.067	–16.1	0.314
Asian, non-Hispanic	3.2	2.6	0.6	0.695	16.0	0.160	14.6	0.135	10.2	0.353
Other race, non-Hispanic	2.5	1.0	1.5	0.145	2.5	0.433	2.5	0.354	0.5	0.790
Nonwhite	73.2	69.0	4.2	0.525	–6.6	0.614	–14.3	0.209	–12.8	0.344
Age at spring assessment	5.1	5.1	0.0	0.559	0.0	0.720	0.0	0.569	0.0	0.750
Age at spring SCBE	5.1	5.1	0.0	0.489	0.0	0.735	0.0	0.638	0.0	0.951
Fall assessment in Spanish	15.1	8.1	7.0	0.179	0.4	0.965	–1.2	0.886	–1.1	0.904
Missing fall assessment	12.5	10.3	2.2	0.383	–4.8	0.337	–5.6	0.275	–9.4	0.164
Missing fall SCBE	17.2	21.0	–3.8	0.518	–0.6	0.948	–4.9	0.638	4.4	0.738
Missing parent data	25.9	25.5	0.4	0.866	–5.0	0.154	–5.0	0.183	–6.2	0.112
Number of students	**895**	**960**	—		—		—		—	
Number of sites	**28**	**37**	—		—		—		—	

*p-value (of adjusted difference in means) < 0.05, two-tailed test.
NOTE: All figures were estimated using sample weights to account for the sample and survey designs. Difference estimates obtained from a logit model (for binary dependent variables) or ordinary least squares model (for continuous dependent variables) of outcome variable on an indicator variable of ERF grant receipt and the specified function of grant applicant score. Standard errors account for design effects due to unequal weighting of the data and clustering at the site level.
SOURCE: Parent consent forms, fall and spring parent surveys, and fall and spring assessments.

Table A.4. Characteristics of children in funded and unfunded sites, adjusted for differences in grant applicant score: covariates from parent survey (percentages, unless otherwise noted)

	Means		Raw difference		Difference conditional on Score		Difference conditional on quadratic in Score		Difference conditional on cubic in Score	
	Funded	Unfunded	Difference	P-value	Difference	P-value	Difference	P-value	Difference	P-value
In past 6 months family received										
Welfare or TANF	12.4	17.4	–5.0	0.080	–3.2	0.522	–3.1	0.543	0.0	0.997
Unemployment insurance	4.2	3.9	0.2	0.851	–0.9	0.717	–1.3	0.644	–2.4	0.463
Food stamps	29.9	38.6	–8.8	0.087	4.6	0.590	2.9	0.742	–1.3	0.904
WIC	35.4	45.3	–9.9	0.034*	–12.6	0.082	–14.3	0.048*	–13.8	0.117
Child support	15.1	15.5	–0.4	0.891	3.2	0.538	3.9	0.468	3.0	0.631
SSI	8.5	10.6	–2.1	0.328	4.0	0.314	3.1	0.463	2.5	0.592
Foster care assistance	1.2	2.4	–1.2	0.176	–3.7	0.183	–2.1	0.198	–2.5	0.162
Energy assistance	6.6	8.1	–1.4	0.556	–3.3	0.458	–3.6	0.398	–3.1	0.518
Mother's marital status (omitted category is mother not respondent)										
Married	45.4	38.3	7.1	0.078	1.0	0.873	1.5	0.811	3.9	0.647
Unmarried	36.4	42.6	–6.2	0.194	3.1	0.678	–0.1	0.992	–1.6	0.858
Child's age at preschool entry	3.2	3.0	0.2	0.127	0.2	0.363	0.2	0.236	0.3	0.221
Country of birth (omitted category is other or refused to answer)										
Child born in U.S.	75.7	93.7	–18.0	0.000*	–15.4	0.051	–17.3	0.034*	–12.7	0.108
Parent born in U.S.	47.4	60.9	–13.5	0.053	–6.2	0.658	–0.2	0.989	–2.9	0.863
Parent born in Mexico	18.3	17.9	0.4	0.949	–3.8	0.716	–3.7	0.697	–1.7	0.891
Parents years in U.S. (omitted category is parent not respondent or refused to answer)										

Table A.4. Characteristics of children in funded and unfunded sites, adjusted for differences in grant applicant score: covariates from parent survey (percentages, unless otherwise noted) —*Continued*

	Means		Raw difference		Difference conditional on Score		Difference conditional on quadratic in Score		Difference conditional on cubic in Score	
	Funded	Unfunded	Difference	P-value	Difference	P-value	Difference	P-value	Difference	P-value
Less than 5	4.6	3.5	1.0	0.416	0.4	0.853	0.7	0.740	5.8	0.266
Greater than 5	88.0	89.2	–1.2	0.544	2.2	0.531	2.1	0.543	2.2	0.575
Parental education (omitted category is parent not respondent)										
Less than high school	27.8	28.9	–1.1	0.810	–2.3	0.784	–3.9	0.611	2.3	0.819
High school	33.0	29.9	3.2	0.368	–7.2	0.208	–6.8	0.274	–17.8	0.001*
Some college or more	34.5	33.0	1.5	0.750	15.7	0.031*	16.9	0.023*	23.4	0.007*
Household income in past month (omitted category is refused to answer)										
Less than $1000	20.9	24.8	–4.0	0.264	–5.4	0.415	–4.6	0.503	0.4	0.965
$1000–2000	33.6	35.3	–1.7	0.647	5.1	0.472	4.4	0.557	5.5	0.532
More than $2000	35.8	31.1	4.7	0.228	1.5	0.847	3.3	0.676	–3.7	0.716
Homeownership (omitted category is public/subsidized housing or other arrangement)										
Family owns home	38.9	30.1	8.8	0.065	9.5	0.277	12.6	0.148	13.2	0.214
Family rents home	46.1	51.6	–5.5	0.261	–11.3	0.184	–13.7	0.100	–14.7	0.137
Family moved in past year	24.3	28.1	–3.9	0.230	–0.2	0.969	0.7	0.914	–2.1	0.762
Number of students	**690**	**728**	—		—		—		—	
Number of sites	**28**	**37**	—		—		—		—	

*p-value (of adjusted difference in means) < 0.05, two-tailed test.

NOTE: All figures were estimated using sample weights to account for the sample and survey designs. Difference estimates obtained from a logit model of outcome variable on an indicator variable of ERF grant receipt and the specified function of grant applicant score. Standard errors account for design effects due to unequal weighting of the data and clustering at the site level.

SOURCE: Fall and spring parent surveys.

Table A.5. Characteristics of preschool ZIP code areas in funded and unfunded sites, adjusted for differences in grant applicant score (percentages, unless otherwise noted)

	Means		Raw Difference		Difference Conditional on Score		Difference Conditional on Quadratic in Score		Difference Conditional on Cubic in Score	
	Funded	Unfunded	Difference	P-value	Difference	P-value	Difference	P-value	Difference	P-value
Urban	88.2	87.2	1.1	0.895	11.8	0.529	–9.1	0.591	–7.9	0.656
Percent White	63.7	58.6	5.1	0.316	8.5	0.367	12.4	0.160	13.3	0.242
Percent Black	16.9	22.5	–5.6	0.239	–2.4	0.802	0.5	0.957	0.7	0.954
Percent Hispanic	23.7	21.7	1.9	0.745	–10.6	0.312	–18.6	0.052	–17.8	0.143
Median Income ($)	43,371	37,170	6,200	0.024*	8,768.3	0.056	12,033	0.013*	10,760	0.056
Poverty Rate	17.1	21.0	–3.9	0.068	–7.1	0.066	–9.9	0.010*	–8.5	0.076
Unemployment Rate	7.2	9.0	–1.7	0.040*	–2.2	0.192	–3.4	0.036*	–2.6	0.224
Number of Centers	85	80	—		—		—		—	
Number of Sites	28	37	—		—		—		—	

*p-value (of adjusted difference in means) < 0.05, two-tailed test.

NOTE: All figures were estimated using sample weights to account for the sample and survey designs. Difference estimates obtained from a logit model (for binary dependent variables) or ordinary least squares model (for continuous dependent variables) of outcome variable on an indicator variable of ERF grant receipt and the specified function of grant applicant score. Standard errors account for design effects due to unequal weighting of the data and clustering at the site level.

SOURCE: 2000 Census.

Additional Specification Tests

We conducted several additional specification tests to assess whether the linear functional form specification is appropriate.[77] For the first test, we estimated models that allowed for a discontinuity at the true value of the *Score* cutoff value (74) as well as at various *false* values of the cutoff value. To implement this test, we included as an additional model covariate an indicator variable signifying whether the application score was greater than 54, 64, or 84. If the ERF *Score* cutoff value at 74 represents a true discontinuity in the relationship between the outcome variables and *Score* and the relationship is otherwise linear, we would not expect to find evidence of "impacts" at the false values of the cutoff value.

This is indeed the case for the child impact models (see Table A.6). None of the estimated impacts at the false cutoff values are statistically significant. The only exception is a statistically significant estimated impact on social competence with a cutoff value of 54, which may be due to chance. (With a 5-percent critical value, we would expect to find significant estimates for roughly 5 percent of the 30 outcome-cutoff value combinations examined, simply due to chance alone.) Furthermore, the magnitude of the "impacts" at the false cutoffs are smaller than at the true cutoff.

The second (and related) test of the linear specification assumes that the true cutoff value is unknown and attempts to estimate it from the data by (1) sequentially estimating models that allow the discontinuity to occur at different *Score* values, and (2) selecting the model with the largest regression R^2 value.[78] If the linear *Score* specification is correct and ERF had a statistically significant impact on the outcome examined, we would expect the R^2 to be maximized in the model with the true value of the *Score* cutoff value.

Results from this test suggest again that the linear specification is appropriate for the child impact analysis (see Table A.7). For print awareness—the one outcome for which we estimated a statistically significant impact in our main models—the R^2 is larger in the model with the cutoff indicator variable defined at 74 than in models with other cutoff indicator variables.

[77] Ludwig and Miller 2007 provide more details on these tests.

[78] This test differs from the first test because the false cutoff indicator variables are added without controlling for the true cutoff value.

Table A.6. Child impact estimates at true and false values of ERF grant receipt cutoff value

	True value of cutoff		False values of cutoff					
	74		54		64		84	
Outcome	Effect Size[b]	P-value	Effect Size	P-value	Effect Size	P-value	Effect Size	P-value
Language and Literacy Skills								
Print and letter knowledge								
Print awareness, Raw Score	0.44	0.027*	–0.28	0.176	–0.33	0.121	0.09	0.616
Print awareness, Standard Score	0.34	0.042*	–0.22	0.222	–0.22	0.230	–0.01	0.941
Phonological awareness								
Elision, Raw Score	0.10	0.441	–0.18	0.185	–0.15	0.277	0.03	0.799
Oral language								
Expressive Vocabulary, Raw Score	0.01	0.965	–0.26	0.063	0.01	0.972	0.00	0.997
Expressive Vocabulary, Standard Score	0.03	0.841	–0.23	0.104	0.00	0.986	–0.02	0.870
Auditory Comprehension, Raw Score	0.27	0.095	–0.24	0.155	0.05	0.787	0.00	0.977
Auditory Comprehension, Standard Score	0.28	0.088	–0.24	0.159	0.01	0.975	–0.11	0.467
Social Competence and Behavior Evaluation								
Social Competence	0.10	0.617	–0.50	0.020*	0.03	0.892	0.19	0.278
Anxiety-Withdrawal	0.00	0.992	0.18	0.346	–0.07	0.713	0.03	0.858
Anger-Aggression	–0.26	0.128	0.26	0.161	0.02	0.913	0.05	0.732

*p-value (of adjusted difference in means) < 0.05, two-tailed test.

[b]The effect size was calculated by dividing the estimated impact by the standard deviation of the outcome measure (that is, the impact expressed as a percentage of the standard deviation).

NOTE: All figures were estimated using sample weights to account for the sample and survey designs. Standard errors of the impact estimates account for design effects due to unequal weighting of the data and clustering at site and classroom level. All estimates were obtained from a regression model of the outcome variable on an indicator variable of ERF grant receipt; an indicator variable of whether grant application score exceeded the specified false cutoff value; grant application score; and indicator variables of female and nonwhite, using SAS's PROC MIXED procedure. Language and literacy skill models also control for indicator variables of fall assessment taken in Spanish and fall assessment data missing and age at spring assessment. SCBE models also control for an indicator variable of missing fall SCBE data and age at spring SCBE observation. Missing values of covariates mean-imputed by site and gender.

SOURCE: ERF spring child assessments and SCBE evaluations.

Table A.7. R-squared of models with true and false values of ERF cutoff

	True Value	False Values		
Outcome	74	54	64	84
Language and Literacy Skills				
Print and letter knowledge				
Print awareness, Raw Score	0.39	0.37	0.35	0.32
Print awareness, Standard Score	0.37	0.34	0.33	0.30
Phonological awareness				
Elision, Raw Score	0.59	0.60	0.59	0.59
Oral language				
Expressive Vocabulary, Raw Score	0.81	0.82	0.81	0.81
Expressive Vocabulary, Standard Score	0.80	0.81	0.81	0.81
Auditory Comprehension, Raw Score	0.55	0.55	0.52	0.53
Auditory Comprehension, Standard Score	0.64	0.64	0.62	0.61
Social Competence and Behavior Evaluation				
Social Competence	0.30	0.39	0.29	0.32
Anxiety-Withdrawal	0.13	0.13	0.13	0.13
Anger-Aggression	0.26	0.27	0.23	0.23

*p-value (of adjusted difference in means) < 0.05, two-tailed test.
NOTE: All figures were estimated using sample weights to account for the sample and survey designs. Estimates account for design effects due to unequal weighting of the data and clustering at site and classroom level. Estimates were obtained from a regression model of the outcome variable on an indicator variable of whether grant application score exceeded the specified cutoff value; grant application score; and indicator variables of female and nonwhite, using SAS's PROC MIXED procedure. Language and literacy skill models also control for indicator variables of fall assessment taken in Spanish and fall assessment data missing and age at spring assessment. SCBE models also control for an indicator variable of missing fall SCBE data and age at spring SCBE observation. Missing values of covariates mean-imputed by site and gender.
SOURCE: ERF spring child assessments and SCBE evaluations.

Sensitivity Analysis

Despite the evidence in support of the linear functional form of *Score*, we estimated models with alternative parametric functional forms and with nonparametric methods to assess the robustness of the impact findings.

Alternative Parametric Specifications. We find that the results of the child-impact analysis are *not* sensitive to the particular choice of the parametric functional form. Table A.8 presents child impact estimates conditional on a quadratic function of *Score*; although not statistically significant at the 5-percent level, impact estimates for print awareness are comparable in magnitude to those from the main model specification. Impact estimates for auditory comprehension are also comparable in magnitude and significance to those from the main model, and impact estimates for other outcomes remain small and statistically insignificant at conventional levels. Table A.9 presents child-impact estimates conditional on a cubic function of *Score*; again, impact estimates are comparable in magnitude and significance to those from the main model specification.

Table A.8. ERF impacts on child outcomes in spring, quadratic in grant applicant score

Outcome (Range)	Funded	Unfunded	Estimated Impact[a]	Effect Size[b]	P-value of Impact
Language And Literacy Skills					
Print and letter knowledge					
Print awareness, Raw Score (0–36)	22.89	18.99	3.90	0.39	0.062
Print awareness, Standard Score (58–144)	102.33	96.84	5.49	0.32	0.068
Phonological awareness					
Elision, Raw Score (0–18)	9.24	8.96	0.28	0.07	0.616
Oral language					
Expressive Vocabulary, Raw Score (0–99)	38.95	39.24	–0.29	–0.02	0.892
Expressive Vocabulary, Standard Score (53–147)	83.48	83.35	0.13	0.01	0.956
Auditory Comprehension, Raw Score (1–62)	52.37	50.36	2.01	0.27	0.115
Auditory Comprehension, Standard Score (50–135)	94.45	89.88	4.57	0.30	0.086
Number of Students	**802**	**846**			
Number of Sites	**28**	**37**			
Social Competence and Behavior Evaluation (Scales Range from 0 to 50)					
Social Competence	30.85	30.97	–0.11	–0.01	0.951
Anxiety-Withdrawal	10.99	10.85	0.14	0.02	0.911
Anger-Aggression	8.80	10.80	–2.00	–0.23	0.198
Number of Students	**801**	**844**			
Number of Sites	**28**	**37**			

*p-value (of adjusted difference in means) < 0.05, two-tailed test.

[a]All estimates were obtained from a regression model of the outcome variable on an indicator variable of ERF grant receipt; a quadratic in grant application score; and indicator variables of female and nonwhite, using SAS's PROC MIXED procedure. Language and literacy skill models also control for indicator variables of fall assessment taken in Spanish and fall assessment data missing and age at spring assessment. SCBE models also control for an indicator variable of missing fall SCBE data and age at spring SCBE observation. Missing values of covariates mean-imputed by site and gender.

[b]The effect size was calculated by dividing the estimated impact by the standard deviation of the outcome measure (that is, the impact expressed as a percentage of the standard deviation).

SOURCE: ERF spring child assessments and SCBE evaluations.

NOTE: All figures were estimated using sample weights to account for the sample and survey designs. Standard errors of the impact estimates account for design effects due to unequal weighting of the data and clustering at site and classroom level.

Table A.9. ERF impacts on child outcomes in spring, cubic in grant applicant score

Outcome (Range)	Funded	Unfunded	Estimated Impact[a]	Effect Size[b]	P-value of Impact
Language And Literacy Skills					
Print and letter knowledge					
Print awareness, Raw Score (0–36)	23.49	17.45	6.04	0.60	0.017*
Print awareness, Standard Score (58–144)	103.05	94.99	8.07	0.48	0.028*
Phonological awareness					
Elision, Raw Score (0–18)	9.28	8.86	0.42	0.10	0.545
Oral language					
Expressive Vocabulary, Raw Score (0–99)	39.01	39.08	–0.07	–0.00	0.979
Expressive Vocabulary, Standard Score (53–147)	83.61	83.04	0.57	0.03	0.851
Auditory Comprehension, Raw Score (1–62)	52.26	50.65	1.61	0.22	0.300
Auditory Comprehension, Standard Score (50–135)	94.61	89.46	5.14	0.34	0.114
Number of Students	**802**	**846**			
Number of Sites	**28**	**37**			
Social Competence and Behavior Evaluation (Scales Range from 0 to 50)					
Social Competence	31.00	30.59	0.40	0.04	0.860
Anxiety-Withdrawal	11.12	10.50	0.62	0.09	0.676
Anger-Aggression	9.14	9.94	–0.80	–0.09	0.669
Number of Students	**801**	**844**			
Number of Sites	**28**	**37**			

*p-value (of adjusted difference in means) < 0.05, two-tailed test.
[a]All estimates were obtained from a regression model of the outcome variable on an indicator variable of ERF grant receipt; a cubic in grant application score; and indicator variables of female and nonwhite, using SAS's PROC MIXED procedure. Language and literacy skill models also control for indicator variables of fall assessment taken in Spanish and fall assessment data missing and age at spring assessment. SCBE models also control for an indicator variable of missing fall SCBE data and age at spring SCBE observation. Missing values of covariates mean-imputed by site and gender.
[b]The effect size was calculated by dividing the estimated impact by the standard deviation of the outcome measure (that is, the impact expressed as a percentage of the standard deviation).
NOTE: All figures were estimated using sample weights to account for the sample and survey designs. Standard errors of the impact estimates account for design effects due to unequal weighting of the data and clustering at site and classroom level.
SOURCE: ERF spring child assessments and SCBE evaluations.

Nonparametric Methods. We also estimated impacts by using nonparametric methods, which relax assumptions about the appropriate functional form for *Score* (Porter 2003; Ludwig and Miller 2005). This approach estimates local linear regressions (Fan 1992) to the left and right of the discontinuity. We implemented this approach in three steps:

Step 1. Using data from the funded sites, we estimated weighted local linear regressions. The weight for a child (or classroom) in a site was inversely proportional to the absolute difference between the site Score value and 74 (that is, sites with scores closer to 74 were given more weight than sites with scores further from 74). The weight for child (or classroom) *i* in site *s* was defined using a tricube kernel:

$$(1)\quad Weight_{is} = \left[1-\left(\frac{|Score_s - 74|}{h}\right)^3\right]^3 \quad for \ \frac{|Score_s - 74|}{h} < 1$$

$$= \quad 0 \qquad\qquad for \ \frac{|Score_s - 74|}{h} \geq 1,$$

where *h* is the bandwidth (smoothing parameter). We selected *h* to be 20, 30, or 40 based on empirical analyses examining how quickly the site weights decrease as *Score* becomes further from 74. The regression models included a linear specification for (*Score*-74) and several baseline covariates from our preferred specification.

Step 2. We repeated Step 1 using data points from the unfunded sites. We used the tricube kernel and bandwidths discussed *in Step 1* to construct the weights for the regression models.

Step 3. We estimated impacts as the difference between the estimated intercepts from the regression models in Steps 1 and 2. Impact estimates were computed as the difference between the left and right limits of the local linear regressions at the Score cutoff value. These impact estimates are less precise than those under the parametric models because of design effects due to unequal weighting of the data and because of smaller sample sizes due to the fact that some sites were given zero weight in this analysis.

Table A.10 presents results from the nonparametric regression model of child impacts with the bandwidth of 20. We find again that results are similar to those from the main model. Results are also similar using bandwidths of 30 and 40 (not shown).

Table A.10. ERF impacts on child outcomes in spring, nonparametric model

Outcome (Range)	Funded	Unfunded	Estimated Impact[a]	Effect Size[b]	P-value of Impact
Language And Literacy Skills					
Print and letter knowledge					
Print awareness, Raw Score (0–36)	22.96	17.34	5.62	0.57	0.007*
Print awareness, Standard Score (58–144)	102.86	95.22	7.64	0.46	0.012*
Phonological awareness					
Elision, Raw Score (0–18)	9.36	8.84	0.52	0.12	0.449
Oral language					
Expressive Vocabulary, Raw Score (0–99)	39.02	39.78	–0.76	–0.05	0.767
Expressive Vocabulary, Standard Score (53–147)	83.56	83.77	–0.22	–0.01	0.944
Auditory Comprehension, Raw Score (1–62)	52.36	51.11	1.25	0.18	0.327
Auditory Comprehension, Standard Score (50–135)	94.56	90.25	4.31	0.28	0.146
Number of Students	**695**	**556**			
Number of Sites	**25**	**23**			
Social Competence and Behavior Evaluation (Scales Range from 0 to 50)					
Social Competence	31.97	31.60	0.37	0.04	0.833
Anxiety-Withdrawal	10.91	10.67	0.24	0.04	0.853
Anger-Aggression	8.63	9.35	–0.72	–0.08	0.688
Number of Students	**690**	**562**			
Number of Sites	**25**	**23**			

*p-value (of adjusted difference in means) < 0.05, two-tailed test.
[a]All estimates were obtained from a locally weighted kernel regression of the outcome variable on an indicator variable of ERF grant receipt; grant application score; grant application score interacted with grant receipt; and indicator variables of female and nonwhite, using SAS's PROC MIXED procedure. Language and literacy skill models also control for indicator variables of fall assessment taken in Spanish and fall assessment data missing and age at spring assessment. SCBE models also control for an indicator variable of missing fall SCBE data and age at spring SCBE observation. Missing values of covariates mean-imputed by site and gender.
[b]The effect size was calculated by dividing the estimated impact by the standard deviation of the outcome measure (that is, the impact expressed as a percentage of the standard deviation).
NOTE: All figures were estimated using sample weights to account for the sample and survey designs. Standard errors of the impact estimates account for design effects due to unequal weighting of the data and clustering at site and classroom level.
SOURCE: ERF spring child assessments and SCBE evaluations.

Restricting the Sample to Unfunded Sites Close to the 74 Cutoff Value. As another test of the sensitivity of results to the functional form of *Score* (which is similar in spirit to the nonparametric approach), we estimated models, controlling for a linear function of *Score* but restricting the sample to the 56 sites with grant application scores closest to the cutoff value (all 28 funded sites and the highest scoring 28 unfunded sites). Results from this version of the child impact model are also similar in magnitude and significance to those from the main model specification (see Table A.11).

Table A.11. ERF impacts on child outcomes in spring, 56 sites closest to cutoff value

Outcome (Range)	Funded	Unfunded	Estimated Impact[a]	Effect Size[b]	P-value of Impact
Language And Literacy Skills					
Print and letter knowledge					
Print awareness, Raw Score (0–36)	23.39	19.08	4.31	0.43	0.040*
Print awareness, Standard Score (58–144)	103.04	96.57	6.47	0.38	0.036*
Phonological awareness					
Elision, Raw Score (0–18)	9.34	8.99	0.35	0.08	0.558
Oral language					
Expressive Vocabulary, Raw Score (0–99)	39.07	39.24	–0.17	–0.01	0.941
Expressive Vocabulary, Standard Score (53–147)	83.55	83.17	0.38	0.02	0.885
Auditory Comprehension, Raw Score (1–62)	52.33	50.32	2.00	0.26	0.147
Auditory Comprehension, Standard Score (50–135)	94.30	89.31	4.99	0.32	0.080
Number of Students	**802**	**674**			
Number of Sites	**28**	**28**			
Social Competence and Behavior Evaluation (Scales Range from 0 to 50)					
Social Competence	31.65	31.67	–0.03	–0.00	0.989
Anxiety-Withdrawal	10.93	10.64	0.29	0.04	0.811
Anger-Aggression	8.87	10.43	–1.57	–0.18	0.341
Number of Students	**801**	**674**			
Number of Sites	**28**	**28**			

*p-value (of adjusted difference in means) < 0.05, two-tailed test.
[a] All estimates were obtained from a regression model of the outcome variable on an indicator variable of ERF grant receipt; grant application score; and indicator variables of female and nonwhite, using SAS's PROC MIXED procedure. Language and literacy skill models also control for indicator variables of fall assessment taken in Spanish and fall assessment data missing and age at spring assessment. SCBE models also control for an indicator variable of missing fall SCBE data and age at spring SCBE observation. Missing values of covariates are mean-imputed by site and gender. Sample was limited to all 28 funded sites and 28 highest scoring unfunded sites.
[b] The effect size was calculated by dividing the estimated impact by the standard deviation of the outcome measure (that is, the impact expressed as a percentage of the standard deviation).
NOTE: All figures were estimated using sample weights to account for the sample and survey designs. Standard errors of the impact estimates account for design effects due to unequal weighting of the data and clustering at site and classroom level.
SOURCE: ERF spring child assessments and SCBE evaluations.

Assessing Site Nonresponse Bias. As discussed in Chapter 2, 28 out of 30 (93 percent) of the funded sites agreed to participate in the study, but only 37 of the 62 unfunded sites recruited for the study were included in the study sample, for a response rate of 60 percent. Among the unfunded sites, the distribution of application scores is similar for the participants and nonparticipants. Furthermore, the observable characteristics of the two groups of sites are similar. Nonetheless, nonresponse in the unfunded sites could affect the impact estimates (that is, the intercepts and slopes of the fitted regression lines) to the extent that child or classroom outcomes differ in the nonparticipating and participating sites.

To place realistic bounds on the effects of site nonresponse bias on the impact estimates, we "imputed" site-level outcomes for a nonparticipant site, using observed site-level outcomes for the six participating sites with the closest application scores. We sequentially estimated impacts where missing site outcomes were imputed using the second smallest outcome value among the six comparison values; then, we followed the same procedure, using the third, fourth, and fifth smallest outcome values. We believe that the third and fourth smallest values (corresponding to

the fortieth and sixtieth percentiles of the outcome distributions across the six comparison sites) are the most realistic bounds.

Table A.12 presents analysis results for child outcomes. Although the point estimates change somewhat as missing site values are imputed using extreme values, the general pattern of results is similar to the results from the preferred model. In particular, the impact on the print and letter awareness score is statistically significant at the 5-percent level in all specifications but one (which is statistically significant at the 7-percent level), and impacts on all other measures are typically statistically insignificant across the imputation schemes.

Table A.12. ERF impacts on child outcomes in spring where child outcomes for nonparticipating unfunded sites are imputed

	Estimated impact (p-value) [a]				
		Imputations based on the 20th to 80th value of the outcome distribution for the six sites with the closest application scores			
Outcome (Range)	No Imputation	20th	40th	60th	80th
Language and Literacy Skills					
Print and letter knowledge					
Print awareness, Raw Score (0–36)	0.49	0.30	0.55	0.70	0.73
	(0.031)*	(0.072)*	(0.001)*	(0.000)*	(0.000)*
Print awareness, Standard Score (58–144)					
Phonological awareness					
Elision, Raw Score (0–18)	0.13	–0.08	0.12	0.19	0.33
	(0.493)	(0.557)	(0.385)	(0.158)	(0.024)*
Oral language					
Expressive Vocabulary, Raw Score (0–99)	0.10	–0.34	–0.12	0.12	0.56
	(0.831)	(0.313)	(0.710)	(0.710)	(0.112)
Expressive Vocabulary, Standard Score (53–147)	0.08	–0.12	–0.01	0.06	0.36
	(0.780)	(0.571)	(0.959)	(0.776)	(0.119)
Auditory Comprehension, Raw Score (1–62)	0.32	0.09	0.14	0.34	0.54
	(0.178)	(0.607)	(0.395)	(0.034)*	(0.002)*
Auditory Comprehension, Standard Score (50–135)	0.31	0.09	0.29	0.37	0.47
	(0.198)	(0.596)	(0.093)	(0.032)*	(0.011)*
Social Competence and Behavior Evaluation (Scales Range from 0 to 50)					
Social Competence	0.12	0.06	0.13	0.18	0.32
	(0.612)	(0.767)	(0.412)	(0.259)	(0.075)
Anxiety-Withdrawal	0.06	–0.04	–0.01	0.05	0.08
	(0.708)	(0.706)	(0.918)	(0.680)	(0.477)
Anger-Aggression	–0.24	–0.29	–0.26	-0.17	–0.12
	(0.200)	(0.030)*	(0.047)*	(.198)	(0.399)

*p-value (of adjusted difference in means) < 0.05, two-tailed test.
[a]All estimates were obtained from a regression model of the outcome variable at the ***site level*** on an indicator variable of ERF grant receipt and grant application score. Because these estimates were estimated using site-level data, the estimates in this table differ slightly from previous tables that were estimated using child-level data.
NOTE: Standard errors of the impact estimates account for design effects due to clustering at site and classroom level. The sample includes 28 funded and 64 unfunded sites; site values were imputed for 28 nonparticipants using values of the six sites with the closest application scores.
SOURCE: ERF spring child assessments and SCBE evaluations.

Model Covariates

Our preferred child impact models included a limited set of covariates: indicators of whether the child is female; whether the child is white and non-Hispanic; whether fall assessment data were missing; age at spring assessment, and, for language and literacy outcomes, whether the fall assessment was taken in Spanish. Some models also included fall assessment scores as covariates.

As a specification test, we also estimated models with no covariates and models that included more extensive sets of covariates. Table A.13 presents results from a child-impact model with no covariates other than *Score* and an indicator of ERF grant receipt. Table A.14 presents results from a child-impact model that controls for all the covariates included in the preferred model; indicator variables of the racial/ethnic categories described in Table A.3 (instead of the nonwhite indicator variable); and the full set of covariates from the parent survey listed in Table A.4, including information on the family's public-assistance receipt, child's country of origin, parent's country of origin, mother's marital status, educational attainment of responding parent, monthly household income, homeownership, and whether the family moved in the past year. Table A.15 presents results from a child impact model that controls for all these covariates plus the preschool ZIP code covariates, including an indicator of whether the preschool ZIP code was in an urban or nonurban location; the percent of the ZIP code population that was African American, white, and Hispanic; and the median income, poverty rate, and unemployment rate in the ZIP code.

Across all these specifications, results are similar in magnitude and significance level to those from the preferred child-impact model. Thus, our impact results are robust to the choice of model covariates.

Table A.13. ERF impacts on child outcomes in spring, no covariates

Outcome (Range)	Funded	Unfunded	Estimated Impact[a]	Effect Size[b]	P-value of Impact
Language and Literacy Skills					
Print and letter knowledge					
Print awareness, Raw Score (0–36)	23.46	18.80	4.66	0.47	0.034*
Print awareness, Standard Score (58–144)	102.76	96.46	6.31	0.37	0.039*
Phonological awareness					
Elision, Raw Score (0-18)	9.42	8.78	0.63	0.15	0.403
Oral language					
Expressive Vocabulary, Raw Score (0–99)	39.39	38.39	1.00	0.07	0.805
Expressive Vocabulary, Standard Score (53–147)	83.79	82.45	1.34	0.08	0.767
Auditory Comprehension, Raw Score (1–62)	52.34	50.08	2.25	0.30	0.173
Auditory Comprehension, Standard Score (50–135)	93.97	89.21	4.76	0.31	0.192
Number of Students	**802**	**846**			
Number of Sites	**28**	**37**			
Social Competence and Behavior Evaluation (Scales Range from 0 to 50)					
Social Competence	32.17	31.21	0.96	0.1	0.619
Anxiety-Withdrawal	10.76	10.85	–0.09	–0.01	0.935
Anger-Aggression	8.51	10.66	–2.15	–0.25	0.163
Number of Students	**801**	**844**			
Number of Sites	**28**	**37**			

*p-value (of adjusted difference in means) < 0.05, two-tailed test.
[a]All estimates were obtained from a regression model of the outcome variable on an indicator variable of ERF grant receipt and grant application score, using SAS's PROC MIXED procedure.
[b]The effect size was calculated by dividing the estimated impact by the standard deviation of the outcome measure (that is, the impact expressed as a percentage of the standard deviation).
NOTE: All figures were estimated using sample weights to account for the sample and survey designs. Standard errors of the impact estimates account for design effects due to unequal weighting of the data and clustering at site and classroom level.
SOURCE: ERF spring child assessments and SCBE evaluations.

Table A.14. ERF impacts on child outcomes in spring, including additional race and parent covariates

Outcome (Range)	Funded	Unfunded	Estimated Impact[a]	Effect Size[b]	P-value of Impact
Language and Literacy Skills					
Print and letter knowledge					
Print awareness, Raw Score (0–36)	23.27	19.36	3.90	0.39	0.050*
Print awareness, Standard Score (58–144)	102.18	97.35	4.84	0.29	0.092
Phonological awareness					
Elision, Raw Score (0–18)	9.26	9.11	0.15	0.04	0.774
Oral language					
Expressive Vocabulary, Raw Score (0–99)	38.93	39.88	–0.94	–0.06	0.582
Expressive Vocabulary, Standard Score (53–147)	83.27	84.13	–0.86	–0.05	0.657
Auditory Comprehension, Raw Score (1–62)	52.17	50.59	1.58	0.21	0.205
Auditory Comprehension, Standard Score (50–135)	93.65	90.31	3.34	0.22	0.189
Number of Students	**802**	**846**			
Number of Sites	**28**	**37**			
Social Competence and Behavior Evaluation (Scales Range from 0 to 50)					
Social Competence	31.89	31.31	0.58	0.06	0.762
Anxiety-Withdrawal	10.92	10.73	0.19	0.03	0.865
Anger-Aggression	8.76	10.62	–1.86	–0.22	0.175
Number of Students	**801**	**844**			
Number of Sites	**28**	**37**			

*p-value (of adjusted difference in means) < 0.05, two-tailed test.
[a]All estimates were obtained from a regression model of the outcome variable on an indicator variable of ERF grant receipt; grant application score; indicator variables of female and the racial/ethnic categories described in Table A.1; and parent covariates described in Table A.2, with the omitted categories for dummy variables as noted in that table, using SAS's PROC MIXED procedure. Language and literacy skill models also control for indicator variables of fall assessment taken in Spanish and fall assessment data missing and age at spring assessment. SCBE models also control for an indicator variable of missing fall SCBE data and age at spring SCBE observation. Missing values of covariates were mean-imputed by site and gender.
[b]The effect size was calculated by dividing the estimated impact by the standard deviation of the outcome measure (that is, the impact expressed as a percentage of the standard deviation).
NOTE: All figures were estimated using sample weights to account for the sample and survey designs. Standard errors of the impact estimates account for design effects due to unequal weighting of the data and clustering at site and classroom level.
SOURCE: ERF spring child assessments and SCBE evaluations.

Table A.15. ERF impacts on child outcomes in spring, including additional race, parent, and ZIP code covariates

Outcome (Range)	Funded	Unfunded	Estimated Impact[a]	Effect Size[b]	P-value of Impact
Language and Literacy Skills					
Print and letter knowledge					
Print awareness, Raw Score (0–36)	23.31	19.24	4.07	0.41	0.044*
Print awareness, Standard Score (58–144)	101.97	97.48	4.49	0.26	0.114
Phonological awareness					
Elision, Raw Score (0–18)	9.23	9.09	0.14	0.03	0.783
Oral language					
Expressive Vocabulary, Raw Score (0–99)	38.84	40.06	–1.23	–0.08	0.496
Expressive Vocabulary, Standard Score (53–147)	83.10	84.36	–1.26	–0.07	0.535
Auditory Comprehension, Raw Score (1–62)	52.03	50.74	1.29	0.17	0.313
Auditory Comprehension, Standard Score (50–135)	93.27	90.61	2.66	0.17	0.284
Number of Students	**802**	**846**			
Number of Sites	**28**	**37**			
Social Competence and Behavior Evaluation (Scales Range from 0 to 50)					
Social Competence	32.07	31.06	1.01	0.11	0.608
Anxiety-Withdrawal	10.88	10.92	–0.05	–0.01	0.966
Anger-Aggression	8.70	10.69	–1.99	–0.23	0.162
Number of Students	**801**	**844**			
Number of Sites	**28**	**37**			

*p-value (of adjusted difference in means) < 0.05, two-tailed test.
[a]All estimates were obtained from a regression model of the outcome variable on an indicator variable of ERF grant receipt; grant application score; indicator variables of female and the racial/ethnic categories described in Table A.1; parent covariates described in Table A.2, with the omitted categories for dummy variables as noted in that table; and zipcode covariates described in Table A.3, using SAS's PROC MIXED procedure. Language and literacy skill models also control for indicator variables of fall assessment taken in Spanish and fall assessment data missing and age at spring assessment. SCBE models also control for an indicator variable of missing fall SCBE data and age at spring SCBE observation. Missing values of covariates were mean-imputed by site and gender.
[b]The effect size was calculated by dividing the estimated impact by the standard deviation of the outcome measure (that is, the impact expressed as a percentage of the standard deviation).
NOTE: All figures were estimated using sample weights to account for the sample and survey designs. Standard errors of the impact estimates account for design effects due to unequal weighting of the data and clustering at site and classroom level.
SOURCE: ERF spring child assessments and SCBE evaluations.

Imputation of Missing Values of Covariates

For our preferred child impact models, we imputed missing values of covariates by assigning the mean value of the covariate by site and gender. For our sensitivity analysis, we estimated impact models using alternative methods for handling missing data. In Table A.16, we present results from a child-level model that includes no imputation of missing values of covariates, and in Table A.17, we present results from a model in which missing values of covariates are imputed via a hotdeck imputation procedure, which replaces the value of the missing covariate with the value of that covariate from a randomly selected child within the same site/gender cell (Rubin 1987).[79]

[79] Rubin, Donald. 1987. Multiple Imputation for Nonresponse in Surveys. New York: John Wiley and Sons, Inc.

Again, results with these alternative imputation approaches are similar in magnitude and significance to those from the main impact models. Thus, the child impact findings are not sensitive to the way in which covariates are imputed.

Table A.16. ERF impacts on child outcomes in spring, no imputation of missing covariates

Outcome (Range)	Funded	Unfunded	Estimated Impact[a]	Effect Size[b]	P-value of Impact
Language and Literacy Skills					
Print and letter knowledge					
Print awareness, Raw Score (0–36)	23.93	19.19	4.75	0.48	0.017*
Print awareness, Standard Score (58–144)	103.24	96.72	6.52	0.39	0.008*
Phonological awareness					
Elision, Raw Score (0–18)	9.57	8.98	0.59	0.14	0.278
Oral language					
Expressive Vocabulary, Raw Score (0–99)	39.66	39.40	0.27	0.02	0.892
Expressive Vocabulary, Standard Score (53–147)	84.16	83.51	0.65	0.04	0.775
Auditory Comprehension, Raw Score (1–62)	52.48	50.27	2.22	0.30	0.064
Auditory Comprehension, Standard Score (50–135)	94.46	89.69	4.76	0.31	0.059
Number of Students	**732**	**760**			
Number of Sites	**28**	**37**			
Social Competence and Behavior Evaluation (Scales Range from 0 to 50)					
Social Competence	32.19	31.28	0.91	0.10	0.623
Anxiety-Withdrawal	10.71	10.85	–0.14	–0.02	0.903
Anger-Aggression	8.51	10.72	–2.21	–0.26	0.135
Number of Students	**796**	**838**			
Number of Sites	**28**	**37**			

*p-value (of adjusted difference in means) < 0.05, two-tailed test.
[a]All estimates were obtained from a regression model of the outcome variable on an indicator variable of ERF grant receipt; grant application score; and indicator variables of female and nonwhite, using SAS's PROC MIXED procedure. Language and literacy skill models also control for an indicator variable of fall assessment taken in Spanish and age at spring assessment. SCBE models also control for age at spring SCBE observation.
[b]The effect size was calculated by dividing the estimated impact by the standard deviation of the outcome measure (that is, the impact expressed as a percentage of the standard deviation).
NOTE: All figures were estimated using sample weights to account for the sample and survey designs. Standard errors of the impact estimates account for design effects due to unequal weighting of the data and clustering at site and classroom level.
SOURCE: ERF spring child assessments and SCBE evaluations.

Table A.17. ERF impacts on child outcomes in spring, hotdeck imputation of missing covariates

Outcome (Range)	Funded	Unfunded	Estimated Impact[a]	Effect Size[b]	P-value of Impact
Language and Literacy Skills					
Print and letter knowledge					
Print awareness, Raw Score (0–36)	23.49	19.11	4.38	0.44	0.029*
Print awareness, Standard Score (58–144)	102.75	96.85	5.90	0.35	0.020*
Phonological awareness					
Elision, Raw Score (0–18)	9.40	8.99	0.41	0.10	0.452
Oral language					
Expressive Vocabulary, Raw Score (0–99)	39.38	39.35	0.03	0.00	0.988
Expressive Vocabulary, Standard Score (53–147)	83.85	83.45	0.41	0.02	0.868
Auditory Comprehension, Raw Score (1–62)	52.37	50.37	2.00	0.27	0.103
Auditory Comprehension, Standard Score (50–135)	94.09	89.82	4.27	0.28	0.096
Number of Students	**802**	**846**			
Number of Sites	**28**	**37**			
Social Competence and Behavior Evaluation (Scales Range from 0 to 50)					
Social Competence	32.16	31.24	0.93	0.10	0.616
Anxiety-Withdrawal	10.80	10.81	–0.01	–0.00	0.994
Anger-Aggression	8.49	10.73	–2.24	–0.26	0.128
Number of Students	**801**	**844**			
Number of Sites	**28**	**37**			

*p-value (of adjusted difference in means) < 0.05, two-tailed test.
[a]All estimates were obtained from a regression model of the outcome variable on an indicator variable of ERF grant receipt; grant application score; and indicator variables of female and nonwhite, using SAS's PROC MIXED procedure. Language and literacy skill models also control for indicator variables of fall assessment taken in Spanish and fall assessment data missing and age at spring assessment. SCBE models also control for an indicator variable of missing fall SCBE data and age at spring SCBE observation. Missing values of covariates imputed via the hotdeck procedure by site and gender.
[b]The effect size was calculated by dividing the estimated impact by the standard deviation of the outcome measure (that is, the impact expressed as a percentage of the standard deviation).
NOTE: Standard errors of the impact estimates account for clustering at site and classroom level.
SOURCE: ERF spring child assessments and SCBE evaluations.

Sample Weights

We estimated our preferred child-impact models with sample weights that account for the sample design, study nonconsent, and interview nonresponse. As a sensitivity test, we estimated a model with base weights that accounted for the sample design but were not adjusted for nonconsent and nonresponse (see Table A.18). Results estimated with this alternative set of weights are similar in magnitude and significance to those from our preferred child-impact model.

Error Structure and Software Packages

We estimated our preferred child-impact models with the SAS software package's PROC MIXED procedure, with random effects at the site and classroom levels for the child impact analysis. As a sensitivity test, we estimated models with PROC MIXED that allowed for random effects at the site level only (see Table A.19). This approach did not the change the magnitude and significance of the impact estimates.

Table A.18. ERF impacts on child outcome in spring, no nonresponse adjustment to weights

Outcome (Range)	Funded	Unfunded	Estimated Impact[a]	Effect Size[b]	P-value of Impact
Language and Literacy Skills					
Print and letter knowledge					
Print awareness, Raw Score (0–36)	23.53	19.07	4.46	0.45	0.021*
Print awareness, Standard Score (58–144)	102.72	96.92	5.80	0.35	0.029*
Phonological awareness					
Elision, Raw Score (0–18)	9.41	8.92	0.49	0.12	0.333
Oral language					
Expressive Vocabulary, Raw Score (0–99)	39.31	39.06	0.25	0.02	0.897
Expressive Vocabulary, Standard Score (53–147)	83.77	83.19	0.58	0.03	0.797
Auditory Comprehension, Raw Score (1–62)	52.28	50.31	1.97	0.27	0.077
Auditory Comprehension, Standard Score (50–135)	93.85	89.72	4.13	0.27	0.085
Number of Students	**802**	**846**			
Number of Sites	**28**	**37**			
Social Competence and Behavior Evaluation (Scales Range from 0 to 50)					
Social Competence	32.24	31.28	0.97	0.10	0.604
Anxiety-Withdrawal	10.74	10.91	–0.17	–0.03	0.883
Anger-Aggression	8.43	10.66	–2.23	–0.26	0.120
Number of Students	**801**	**844**			
Number of Sites	**28**	**37**			

*p-value (of adjusted difference in means) < 0.05, two-tailed test.
[a]All estimates were obtained from a regression model of the outcome variable on an indicator variable of ERF grant receipt; grant application score; and indicator variables of female and nonwhite, using SAS's PROC MIXED procedure. Language and literacy skill models also control for indicator variables of fall assessment taken in Spanish and fall assessment data missing and age at spring assessment. SCBE models also control for an indicator variable of missing fall SCBE data and age at spring SCBE observation. Missing values of covariates were mean-imputed by site and gender.
[b]The effect size was calculated by dividing the estimated impact by the standard deviation of the outcome measure (that is, the impact expressed as a percentage of the standard deviation).
NOTE: All figures were estimated using sample weights to account for the sample and survey designs but that do not adjust for survey nonresponse. Standard errors of the impact estimates account for design effects due to unequal weighting of the data and clustering at site level.
SOURCE: ERF spring child assessments and SCBE evaluations.

Table A.19. ERF impacts on child outcome in spring, clustering at site level only

Outcome (Range)	Funded	Unfunded	Estimated Impact[a]	Effect Size[b]	P-value of Impact
Language and Literacy Skills					
Print and letter knowledge					
Print awareness, Raw Score (0–36)	23.64	18.97	4.68	0.47	0.023*
Print awareness, Standard Score (58–144)	102.75	96.85	5.90	0.35	0.043*
Phonological awareness					
Elision, Raw Score (0–18)	9.41	9.02	0.39	0.09	0.494
Oral language					
Expressive Vocabulary, Raw Score (0–99)	39.62	39.23	0.39	0.03	0.851
Expressive Vocabulary, Standard Score (53–147)	84.17	83.30	0.88	0.05	0.713
Auditory Comprehension, Raw Score (1–62)	52.40	50.27	2.14	0.29	0.092
Auditory Comprehension, Standard Score (50–135)	94.20	89.73	4.47	0.29	0.086
Number of Students	**802**	**846**			
Number of Sites	**28**	**37**			
Social Competence and Behavior Evaluation (Scales Range from 0 to 50)					
Social Competence	32.16	30.97	1.19	0.12	0.569
Anxiety-Withdrawal	10.93	10.45	0.48	0.07	0.722
Anger-Aggression	8.55	10.72	–2.16	–0.25	0.156
Number of Students	**801**	**844**			
Number of Sites	**28**	**37**			

*p-value (of adjusted difference in means) < 0.05, two-tailed test.
[a]All estimates were obtained from a regression model of the outcome variable on an indicator variable of ERF grant receipt; grant application score; and indicator variables of female and nonwhite, using SAS's PROC MIXED procedure. Language and literacy skill models also control for indicator variables of fall assessment taken in Spanish and fall assessment data missing and age at spring assessment. SCBE models also control for an indicator variable of missing fall SCBE data and age at spring SCBE observation. Missing values of covariates were mean-imputed by site and gender.
[b]The effect size was calculated by dividing the estimated impact by the standard deviation of the outcome measure (that is, the impact expressed as a percentage of the standard deviation).
NOTE: All figures were estimated using sample weights to account for the sample and survey designs. Standard errors of the impact estimates account for design effects due to unequal weighting of the data and clustering at site level.
SOURCE: ERF spring child assessments and SCBE evaluations.

As an additional sensitivity test, we estimated impacts using procedures from alternative statistical packages—SUDAAN's PROC REGRESS procedure and Stata's *svy regress* command—that account for clustering effects in slightly different ways than SAS's PROC MIXED. SAS's PROC MIXED uses a maximum likelihood approach to general linear mixed models, whereas the SUDAAN and Stata procedures are based on the Taylor-series linearization method, combined with variance estimation formulas specific to the sample design. Estimates from both the SUDAAN and Stata models are similar in magnitude and significance to those from the main child impact models (see Table A.20 and Table A.21).[80]

[80] Although the estimated impact on auditory comprehension in the SUDAAN and Stata models has a p-value of 0.030, this impact is not statistically significant at the 5-percent level once we take into account the multiple comparisons within the language development domain using the Benjamini-Hochberg procedure, as described later in this appendix.

Table A.20. ERF impacts on child outcomes in spring, estimated in SUDAAN

Outcome (Range)	Funded	Unfunded	Estimated Impact[a]	Effect Size[b]	P-value of Impact
Language and Literacy Skills					
Print and letter knowledge					
Print awareness, Raw Score (0–36)	23.68	18.93	4.75	0.47	0.011*
Print awareness, Standard Score (58–144)	102.82	96.81	6.01	0.35	0.016*
Phonological awareness					
Elision, Raw Score (0–18)	9.41	9.02	0.38	0.09	0.427
Oral language					
Expressive Vocabulary, Raw Score (0–99)	39.63	39.30	0.33	0.02	0.855
Expressive Vocabulary, Standard Score (53–147)	84.19	83.39	0.80	0.05	0.710
Auditory Comprehension, Raw Score (1–62)	52.42	50.28	2.14	0.29	0.019*
Auditory Comprehension, Standard Score (50–135)	94.24	89.76	4.48	0.29	0.030*
Number of Students	**802**	**846**			
Number of Sites	**28**	**37**			
Social Competence and Behavior Evaluation (Scales Range from 0 to 50)					
Social Competence	32.16	30.97	1.19	0.13	0.355
Anxiety-Withdrawal	10.93	10.44	0.49	0.07	0.685
Anger-Aggression	8.55	10.73	–2.18	–0.25	0.139
Number of Students	**801**	**844**			
Number of Sites	**28**	**37**			

*p-value (of adjusted difference in means) < 0.05, two-tailed test.
[a]All estimates were obtained from a regression model of the outcome variable on an indicator variable of ERF grant receipt; grant application score; and indicator variables of female and nonwhite, using SUDAAN. Language and literacy skill models also control for indicator variables of fall assessment taken in Spanish and fall assessment data missing and age at spring assessment. SCBE models also control for an indicator variable of missing fall SCBE data and age at spring SCBE observation. Missing values of covariates were mean-imputed by site and gender.
[b]The effect size was calculated by dividing the estimated impact by the standard deviation of the outcome measure (that is, the impact expressed as a percentage of the standard deviation).
NOTE: All figures were estimated using sample weights to account for the sample and survey designs. Standard errors of the impact estimates account for design effects due to unequal weighting of the data and clustering at site level.
SOURCE: ERF spring child assessments and SCBE evaluations.

Table A.21. ERF impacts on child outcomes in spring, estimated in Stata

Outcome (Range)	Funded	Unfunded	Estimated Impact[a]	Effect Size[b]	P-value of Impact
Language and Literacy Skills					
Print and letter knowledge					
Print awareness, Raw Score (0–36)	23.64	18.89	4.75	0.47	0.011*
Print awareness, Standard Score (58–144)	102.95	96.94	6.01	0.34	0.016*
Phonological awareness					
Elision, Raw Score (0–18)	9.31	8.93	0.38	0.10	0.427
Oral language					
Expressive Vocabulary, Raw Score (0–99)	38.85	38.52	0.33	0.02	0.855
Expressive Vocabulary, Standard Score (53–147)	83.42	82.62	0.80	0.05	0.710
Auditory Comprehension, Raw Score (1–62)	52.34	50.20	2.14	0.30	0.019*
Auditory Comprehension, Standard Score (50–135)	94.06	89.58	4.48	0.30	0.030*
Number of Students	**802**	**846**			
Number of Sites	**28**	**37**			
Social Competence and Behavior Evaluation (Scales Range from 0 to 50)					
Social Competence	32.41	31.22	1.19	0.12	0.355
Anxiety-Withdrawal	10.99	10.50	0.49	0.07	0.685
Anger-Aggression	8.31	10.49	–2.18	–0.25	0.139
Number of Students	**801**	**844**			
Number of Sites	**28**	**37**			

*p-value (of adjusted difference in means) < 0.05, two-tailed test.
[a]All estimates were obtained from a regression model of the outcome variable on an indicator variable of ERF grant receipt; grant application score; and indicator variables of female and nonwhite, using Stata's *svy regress* command. Language and literacy skill models also control for indicator variables of fall assessment taken in Spanish and fall assessment data missing and age at spring assessment. SCBE models also control for an indicator variable of missing fall SCBE data and age at spring SCBE observation. Missing values of covariates were mean-imputed by site and gender.
[b]The effect size was calculated by dividing the estimated impact by the standard deviation of the outcome measure (that is, the impact expressed as a percentage of the standard deviation).
NOTE: All figures were estimated using sample weights to account for the sample and survey designs. Standard errors of the impact estimates account for design effects due to unequal weighting of the data and clustering at site and classroom level.
SOURCE: ERF spring child assessments and SCBE evaluations.

Sensitivity Tests of Classroom Impact Models

Our preferred specification of the classroom-impact models controls for a linear function of *Score* and a limited set of covariates and accounts for design effects due to site-level clustering in the error structure. Missing values of covariates are imputed, and estimates are weighted to account for the sample design. In this section, we discuss (1) the specific parameter assumptions under our preferred model specification for the classroom-impact analysis and (2) the results of sensitivity tests to examine the robustness of the classroom-impact findings to variations in key parameter assumptions. For brevity, we focus our specification tests on a subset of the full set of child- and teacher-outcome variables. These outcome variables, along with the impact estimates from our preferred classroom models, are shown in Table A.22. We find that the pattern of classroom impacts is generally robust to a variety of model specifications. In the following text, we discuss these alternative specifications in greater detail.

Table A.22. ERF impacts on selected spring teacher and classroom outcomes, main model

Outcome (Range)	Funded	Unfunded	Estimated Impact[a]	Effect Size[b]	P-value of Impact
Teachers' Earnings, Experience, and Training					
Professional Development Hours—Early Language and Literacy	72.03	22.09	49.94	1.04	0.002 *
Received professional development through mentoring / tutoring	59.00	15.94	43.07	0.91	0.002 *
Professional Development Hours—Curriculum	39.91	24.51	15.41	0.39	0.209
Received professional development through mentoring/tutoring	47.90	12.46	35.44	0.78	0.022 *
Number of Teachers	90	100			
Number of Sites	28	37			
General Quality of the Preschool Classroom					
ECERS-R Teaching and Interactions	5.94	4.73	1.20	1.12	0.001 *
Teacher sensitivity	3.16	2.49	0.67	0.99	0.008 *
Classroom community	3.33	2.51	0.82	1.22	0.001 *
Total score	2.77	1.84	0.93	1.44	0.000 *
Language, Early Literacy, and Assessment Practices					
Oral Language Use in the Classroom					
Oral Language Use by Lead Teacher (0.86–4.00)	3.00	2.17	0.83	1.11	0.002 *
Oral Language Use by Assistant Teacher (0.50–4.00)	2.77	1.73	1.04	0.89	0.027 *
Book Reading					
Number of Book Reading Sessions Observed (0–4)	1.41	1.20	0.21	0.23	0.516
Book Reading Practices (0.56–3.94)	2.49	1.60	0.89	1.03	0.003 *
Phonological Awareness					
Number of Different Phonological Awareness Activities Observed (0–7)	2.40	0.67	1.73	1.10	0.004 *
Quality of Phonological Awareness Activities (0–4.00)	2.04	1.07	0.97	0.79	0.024 *
Print and Letter Knowledge					
Learning Opportunities (0.50–4.00)	2.05	1.20	0.85	0.87	0.022 *
Classroom Print Environment (0.50–4.00)	2.28	1.59	0.69	0.81	0.028 *
Written Expression					
Learning Opportunities (0.50–4.00)	1.99	0.78	1.21	1.06	0.003 *
Opportunities and Materials for Writing (0.50–4.00)	2.55	1.32	1.23	1.48	0.000 *
Child Assessments					
Child Portfolios (1.00–5.00)	3.07	1.72	1.35	0.98	0.012 *
Dynamic Assessment 0.67–4.33)	2.89	2.18	0.71	0.64	0.095
Number of Classrooms	78	91			
Number of Sites	28	37			

Notes from Table A.22

*p-value < 0.05, two-tailed test.
[a]All estimates were obtained from a regression model of the outcome variable on an indicator variable of ERF grant receipt; grant application score; teacher's age, education, and an indicator variable of nonwhite, using SAS's PROC MIXED procedure. Missing values of covariates were mean-imputed by site.
[b]The effect size was calculated by dividing the estimated impact by the standard deviation of the outcome measure (that is, the impact expressed as a percentage of the standard deviation).
NOTE: All figures were estimated using sample weights to account for the sample and survey designs. Standard errors of the impact estimates account for design effects due to unequal weighting of the data and clustering at site level.
SOURCE: ERF spring director and teacher surveys and classroom observations.

Functional Form Specification for *Score*

Our preferred specification for the classroom impact models, as for our child impact models, includes a linear function of *Score*. We determined that this was the appropriate specification on the bases of graphical inspection of the outcome variables, the examination of baseline values of covariates at the site level (shown in "Specification and Sensitivity Tests on Child Impact Models earlier in this appendix), and additional specification tests. Nonetheless, results are not sensitive to this specification decision.

Graphical Inspection

Figure A.8 displays plots of site-level mean outcomes versus a linear function of *Score* for nine teacher and classroom outcome measures. Figure A.9 displays plots of these same site-level mean outcomes versus a quadratic function of *Score*. In general, the graphs suggest that the linear function of *Score* is appropriate, although for some outcome variables, the relationship with *Score* appears to be quadratic. In our main impact models, we include a linear function of *Score,* but as shown later in this section, impact estimates are generally similar when we instead control for a quadratic or cubic function of *Score*.

Figure A.8. Teacher training and classroom instructional practice scales as a function of *Score*

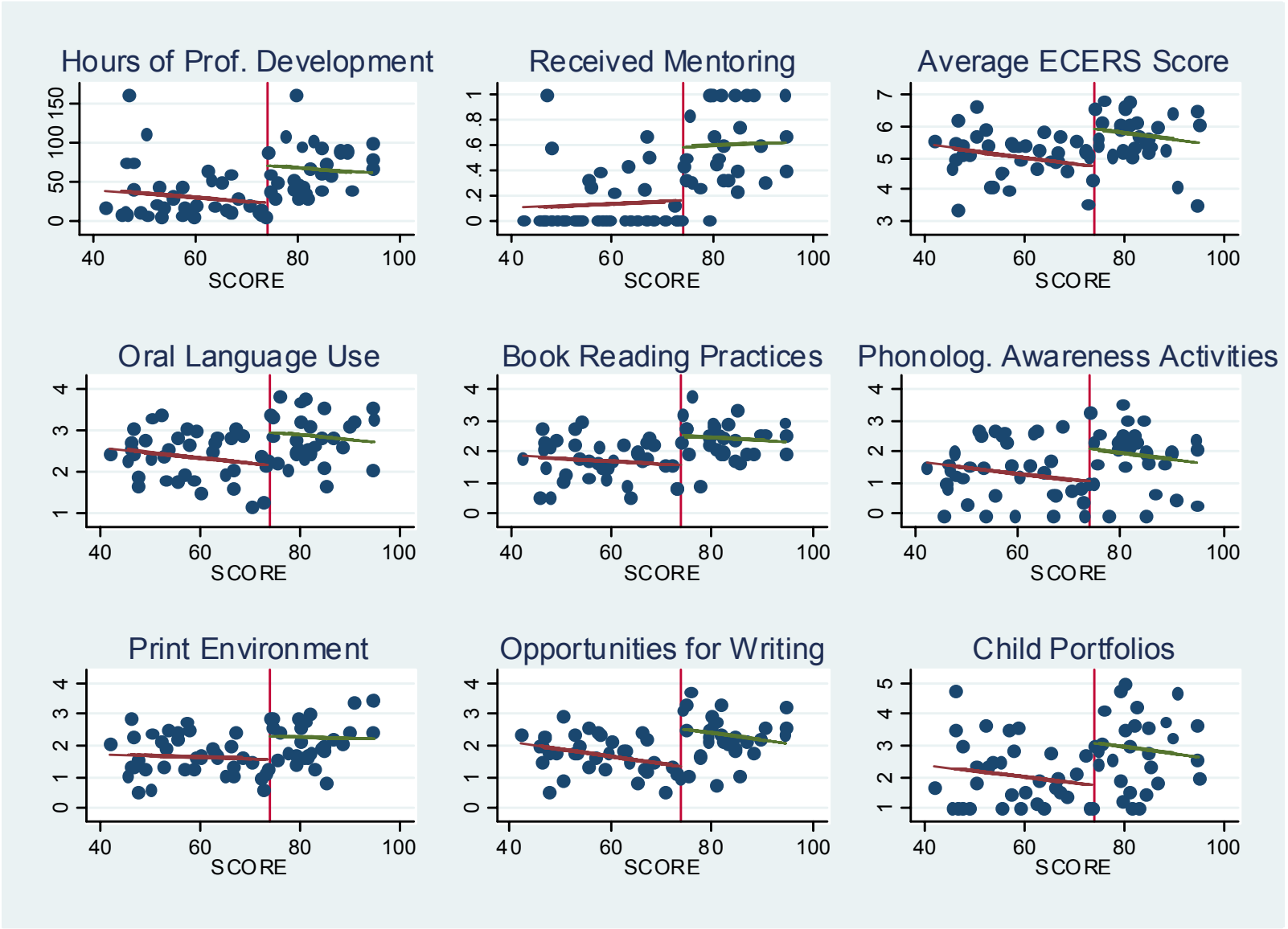
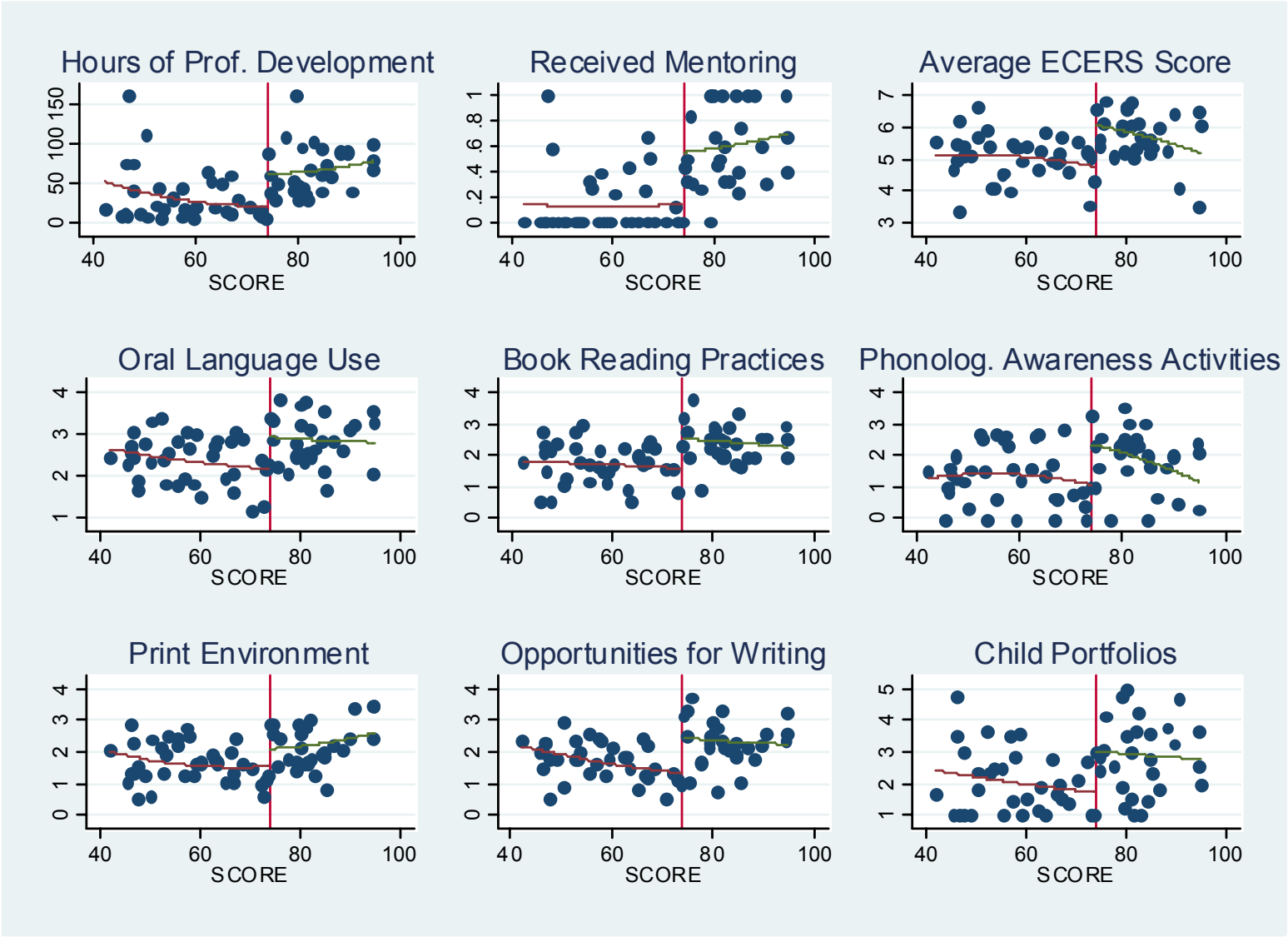


Figure A.9. Teacher training and classroom instructional practice scales as a function of *Score* and *Score*-squared



Additional Specification Tests

As a specification test, we focused on a limited set of outcomes on which we found impacts in our main classroom-impact models, and we estimated alternative models that allowed for a discontinuity at the true value of the *Score* cutoff value and at various false values of the cutoff. If the actual ERF *Score* cutoff value represents a true discontinuity in the relationship between the outcome variables and *Score* and the relationship is otherwise linear, we would not expect to find evidence of impacts at the false values of the cutoff. As shown in Table A.23, this is indeed the case. With only one exception, there are no statistically significant impacts at any of the false values of the cutoff that we examined. The one exception is for the classroom print-environment scale at the cutoff value of 64. This significant effect may be due to chance rather than to any true discontinuities between *Score* and the outcome variable at the false value of the cutoff.

As an additional specification test, we estimated models that allowed for a discontinuity at various false values of the *Score* cutoff rather than at the true value, and we compared the R^2 values across these models. If the linear *Score* specification is correct and ERF had a statistically significant impact on the outcome examined, we would expect the R^2 to be maximized in the model with the true value of the *Score* cutoff. As shown in Table A.24, this is generally the case. The two exceptions, oral language use by assistant teacher and written-expression learning opportunities, may be due to chance.

Sensitivity Analysis

We also examined whether our classroom impact estimates were sensitive to specification of the linear functional form of *Score*. Table A.25 presents results from a model that controls for a quadratic in *Score*; Table A.26 presents results from a model that controls for a cubic in *Score*. Table A.27 presents results from a nonparametric model. Table A.28 presents results of a model that controls for a linear function of *Score* but restricts the sample to the 56 sites with grant applications closest to the cutoff value. Across all these specifications, the pattern of results is generally similar to that from the main model. Thus, we conclude that our results are not sensitive to the linear functional form of *Score* in the regression-discontinuity model.

Model Covariates

The main classroom impact models controlled for the teacher's age, education, and an indicator of whether she was nonwhite. We included teacher's education as a covariate because there appeared to be a difference between funded and unfunded teachers in the proportion of teachers with a bachelor's degree—81 percent compared to 51 percent, based on regression-adjusted averages ($p = 0.016$)—which was not attributable to the ERF program and not accounted for by the score variable. Differential hiring could not be responsible for the difference, because a similar number of teachers in funded and unfunded programs (20 and 19 respectively) reported that they were hired within one year of the fall interview. The education levels of the new hires matched the overall education distribution by funding status, suggesting no substantial change in the educational requirements of new hires following receipt of the ERF grant.

The results were not sensitive to this choice of covariates. There were few additional covariates to add to the models for sensitivity testing; however, as a specification test, we did estimate a model with no covariates other than *Score* and an indicator of ERF grant receipt (see Table A.29). Results from this specification are similar in magnitude and significance level to those from the main classroom-impact model.

Imputation of Missing Values of Covariates

In our preferred classroom impact models, we imputed missing values of covariates by assigning the mean value of the covariate by site. Results were not sensitive to this imputation procedure, however. As shown in Table A.30, results are similar to those from the main model when no imputation is used.

Table A.23. Spring classroom "impact" estimates at true and false values of ERF grant receipt cutoff value

	True value of cutoff		False values of cutoff					
	74		54		64		84	
Outcome	Effect Size[a]	P-value	Effect Size[a]	P-value	Effect Size[a]	P-value	Effect Size[a]	P-value
Oral Language Use in the Classroom								
Oral Language Use by Lead Teacher (0.86–4.00)	1.11	0.002 *	–0.21	0.58	–0.29	0.450	0.02	0.951
Oral Language Use by Assistant Teacher (0.50–4.00)	0.89	0.027 *	–0.54	0.179	–0.31	0.467	–0.14	0.680
Book Reading								
Number of Book Reading Sessions Observed (0–4)	0.23	0.516	–0.26	0.487	0.11	0.772	0.01	0.977
Book Reading Practices (0.56–3.94)	1.03	0.003 *	–0.32	0.366	0.46	0.214	–0.10	0.737
Phonological Awareness								
Number of Different Phonological Awareness Activities Observed (0–7)	1.10	0.004 *	0.27	0.493	–0.13	0.749	–0.46	0.169
Quality of Phonological Awareness Activities (0–4.00)	0.79	0.024 *	0.60	0.097	–0.46	0.221	–0.47	0.125
Print and Letter Knowledge								
Learning Opportunities (0.50–4.00)	0.87	0.022 *	–0.06	0.874	–0.30	0.459	-0.04	0.918
Classroom Print Environment (0.50–4.00)	0.81	0.028 *	0.00	0.997	–0.83	0.033 *	0.34	0.291
Written Expression								
Learning Opportunities (0.50–4.00)	1.06	0.003 *	–0.56	0.131	–0.24	0.538	0.11	0.720
Opportunities and Materials for Writing (0.50–4.00)	1.48	0.000 *	0.07	0.837	–0.52	0.161	–0.05	0.873
Child Assessments								
Child Portfolios (1.00–5.00)	0.98	0.012 *	–0.26	0.512	–0.02	0.966	0.10	0.767
Dynamic Assessment 0.67–4.33)	0.64	0.095	0.31	0.443	–0.67	0.106	0.24	0.494

*p-value < 0.05, two-tailed test.
[a]The effect size was calculated by dividing the estimated impact by the standard deviation of the outcome measure (that is, the impact expressed as a percentage of the standard deviation).
NOTE: All figures were estimated using sample weights to account for the sample and survey designs. Standard errors of the impact estimates account for design effects due to unequal weighting of the data and clustering at site level. All estimates were obtained from a regression model of the outcome variable on an indicator variable of ERF grant receipt; an indicator variable of whether grant application score exceeded the specified false cutoff value; grant application score; teacher's age, education, and an indicator variable of nonwhite, using SAS's PROC MIXED procedure. Missing values of covariates were mean-imputed by site.
SOURCE: ERF spring director and teacher surveys and classroom observations.

Table A.24. R-squared of spring classroom impact models with true and false values of ERF cutoff

Outcome	True Value	False Values		
	74	54	64	84
Oral Language Use in the Classroom				
Oral Language Use by Lead Teacher (0.86–4.00)	0.33	0.31	0.26	0.25
Oral Language Use by Assistant Teacher (0.50–4.00)	0.20	0.21	0.16	0.16
Book Reading				
Book Reading Practices (0.56–3.94)	0.30	0.27	0.07	0.21
Phonological Awareness				
Number of Different Phonological Awareness Activities Observed (0–7)	0.26	0.18	0.18	0.19
Quality of Phonological Awareness Activities (0–4.00)	0.20	0.14	0.16	0.17
Print and Letter Knowledge				
Learning Opportunities (0.50–4.00)	0.32	0.32	0.29	0.29
Classroom Print Environment (0.50–4.00)	0.21	0.17	0.21	0.17
Written Expression				
Learning Opportunities (0.50–4.00)	0.27	0.29	0.20	0.20
Opportunities and Materials for Writing (0.50–4.00)	0.32	0.17	0.16	0.13
Child Assessments				
Child Portfolios (1.00–5.00)	0.16	0.11	0.08	0.08

*p-value < 0.05, two-tailed test.
NOTE: All figures were estimated using sample weights to account for the sample and survey designs. Standard errors of the impact estimates account for design effects due to unequal weighting of the data and clustering at site level. All estimates were obtained from a regression model of the outcome variable on an indicator variable of whether grant application score exceeded the specified false cutoff value; grant application score; teacher's age, education, and an indicator variable of nonwhite, using SAS's PROC MIXED procedure. Missing values of covariates were mean-imputed by site.
SOURCE: ERF spring director and teacher surveys and classroom observations.

Table A.25. ERF impacts on selected spring teacher and classroom outcomes, quadratic in grant applicant score

Outcome (Range)	Funded	Unfunded	Estimated Impact[a]	Effect Size[b]	P-value of Impact	
Teachers' Earnings, Experience, and Training						
Professional Development Hours—Early Language and Literacy	64.08	20.81	43.28	0.90	0.008	*
Received professional development through mentoring / tutoring	55.41	15.35	40.06	0.85	0.005	*
Professional Development Hours—Curriculum	39.30	24.39	14.91	0.38	0.252	
Received professional development through mentoring / tutoring	41.75	11.45	30.31	0.67	0.060	
Number of Teachers	90	100				
Number of Sites	28	37				
General Quality of the Preschool Classroom						
ECERS-R Teaching and Interactions	6.14	4.77	1.38	1.28	0.000	*
Teacher sensitivity	3.17	2.49	0.67	0.99	0.012	*
Classroom community	3.17	2.48	0.69	1.02	0.007	*
Total score	2.71	1.83	0.88	1.36	0.000	*
Language, Early Literacy, and Assessment Practices						
Oral Language Use in the Classroom						
Oral Language Use by Lead Teacher (0.86 - 4.00)	2.94	2.16	0.78	1.05	0.006	*
Oral Language Use by Assistant Teacher (0.50 - 4.00)	2.71	1.71	1.00	0.86	0.042	*
Book Reading						
Number of Book Reading Sessions Observed (0 - 4)	1.38	1.19	0.19	0.20	0.593	
Book Reading Practices (0.56 - 3.94)	2.51	1.61	0.90	1.04	0.005	*
Phonological Awareness						
Number of Different Phonological Awareness Activities Observed (0 - 7)	2.45	0.68	1.78	1.13	0.005	*
Quality of Phonological Awareness Activities (0 - 4.00)	2.25	1.10	1.15	0.94	0.012	*
Print and Letter Knowledge						
Learning Opportunities (0.50 - 4.00)	2.04	1.20	0.84	0.86	0.034	*
Classroom Print Environment (0.50 - 4.00)	2.05	1.55	0.50	0.59	0.118	
Written Expression						
Learning Opportunities (0.50 - 4.00)	1.75	0.74	1.00	0.88	0.018	*
Opportunities and Materials for Writing (0.50 - 4.00)	2.45	1.30	1.15	1.38	0.000	*
Child Assessments						
Child Portfolios (1.00 - 5.00)	2.95	1.70	1.25	0.91	0.025	*
Dynamic Assessment 0.67 - 4.33)	2.92	2.18	0.74	0.67	0.103	
Number of Classrooms	78	91				
Number of Sites	28	37				

Notes from Table A.25

*p-value < 0.05, two-tailed test.
[a]All estimates were obtained from a regression model of the outcome variable on an indicator variable of ERF grant receipt; a quadratic in grant application score; teacher's age, education, and an indicator variable of nonwhite, using SAS's PROC MIXED procedure. Missing values of covariates were mean-imputed by site.
[b]The effect size was calculated by dividing the estimated impact by the standard deviation of the outcome measure (that is, the impact expressed as a percentage of the standard deviation).
NOTE: All figures were estimated using sample weights to account for the sample and survey designs. Standard errors of the impact estimates account for design effects due to unequal weighting of the data and clustering at site level.
SOURCE: ERF spring director and teacher surveys and classroom observations.

Table A.26. ERF impacts on selected spring teacher and classroom outcomes, cubic in grant applicant score

Outcome (Range)	Funded	Unfunded	Estimated Impact[a]	Effect Size[b]	P-value of Impact	
Teachers' Earnings, Experience, and Training						
Professional Development Hours—Early Language and Literacy	65.79	16.66	49.13	1.02	0.014	*
Received professional development through mentoring / tutoring	55.88	14.24	41.64	0.88	0.017	*
Professional Development Hours—Curriculum	42.66	16.07	26.59	0.68	0.096	
Received professional development through mentoring/tutoring	40.95	13.48	27.48	0.61	0.164	
Number of Teachers	90	100				
Number of Sites	28	37				
General Quality of the Preschool Classroom						
ECERS-R Teaching and Interactions	6.15	4.75	1.40	1.30	0.003	*
Teacher sensitivity	3.19	2.43	0.76	1.12	0.020	*
Classroom community	3.25	2.30	0.94	1.40	0.003	*
Total score	2.80	1.60	1.20	1.86	0.000	*
Language, Early Literacy, and Assessment Practices						
Oral Language Use in the Classroom						
Oral Language Use by Lead Teacher (0.86–4.00)	3.01	1.98	1.03	1.38	0.003	*
Oral Language Use by Assistant Teacher (0.50–4.00)	2.83	1.41	1.42	1.22	0.022	*
Book Reading						
Number of Book Reading Sessions Observed (0–4)	1.49	0.94	0.55	0.59	0.202	
Book Reading Practices (0.56–3.94)	2.56	1.50	1.06	1.22	0.007	*
Phonological Awareness						
Number of Different Phonological Awareness Activities Observed (0–7)	2.56	0.42	2.13	1.36	0.006	*
Quality of Phonological Awareness Activities (0–4.00)	2.36	0.82	1.55	1.27	0.005	*
Print and Letter Knowledge						
Learning Opportunities (0.50–4.00)	2.11	1.02	1.08	1.10	0.026	*
Classroom Print Environment (0.50–4.00)	2.25	1.08	1.17	1.38	0.002	*
Written Expression						
Learning Opportunities (0.50–4.00)	1.93	0.28	1.66	1.46	0.001	*
Opportunities and Materials for Writing (0.50–4.00)	2.58	0.99	1.59	1.91	0.000	*
Child Assessments						
Child Portfolios (1.00–5.00)	3.03	1.49	1.55	1.13	0.028	*
Dynamic Assessment 0.67–4.33)	3.14	1.64	1.50	1.36	0.006	*
Number of Classrooms	78	91				
Number of Sites	28	37				

Notes from Table A.26

*p-value < 0.05, two-tailed test.
[a]All estimates were obtained from a regression model of the outcome variable on an indicator variable of ERF grant receipt; a cubic in grant application score; teacher's age, education, and an indicator variable of nonwhite, using SAS's PROC MIXED procedure. Missing values of covariates were mean-imputed by site.
[b]The effect size was calculated by dividing the estimated impact by the standard deviation of the outcome measure (that is, the impact expressed as a percentage of the standard deviation).
NOTE: All figures were estimated using sample weights to account for the sample and survey designs. Standard errors of the impact estimates account for design effects due to unequal weighting of the data and clustering at site level.
SOURCE: ERF spring director and teacher surveys and classroom observations.

Table A.27. ERF impacts on selected spring teacher and classroom outcomes, nonparametric model

Outcome (Range)	Funded	Unfunded	Estimated Impact[a]	Effect Size[b]	P-value of Impact	
Teachers' Earnings, Experience, and Training						
Professional Development Hours—Early Language and Literacy	68.79	18.87	49.92	1.10	0.007	*
Received professional development through mentoring / tutoring	58.56	13.82	44.75	0.91	0.010	*
Professional Development Hours—Curriculum	39.58	21.33	18.25	0.45	0.285	
Received professional development through mentoring / tutoring	44.99	12.17	32.82	0.71	0.103	
Number of Teachers	80	67				
Number of Sites	25	23				
General Quality of the Preschool Classroom						
ECERS-R Teaching and Interactions	6.20	4.61	1.59	1.60	0.000	*
Teacher sensitivity	3.21	2.40	0.81	1.16	0.007	*
Classroom community	3.33	2.37	0.96	1.37	0.001	*
Total score	2.85	1.66	1.18	1.68	0.000	*
Language, Early Literacy, and Assessment Practices						
Oral Language Use in the Classroom						
Oral Language Use by Lead Teacher (0.86 - 4.00)	3.09	1.99	1.10	1.36	0.002	*
Oral Language Use by Assistant Teacher (0.50 - 4.00)	2.89	1.47	1.41	1.17	0.011	*
Book Reading						
Number of Book Reading Sessions Observed (0 - 4)	1.45	1.02	0.43	0.48	0.324	
Book Reading Practices (0.56 - 3.94)	2.60	1.46	1.13	1.28	0.003	*
Phonological Awareness						
Number of Different Phonological Awareness Activities Observed (0 - 7)	2.69	0.41	2.28	1.31	0.005	*
Quality of Phonological Awareness Activities (0 - 4.00)	2.36	0.85	1.51	1.21	0.005	*
Print and Letter Knowledge						
Learning Opportunities (0.50 - 4.00)	2.18	1.03	1.15	1.14	0.013	*
Classroom Print Environment (0.50 - 4.00)	2.43	1.14	1.28	1.62	0.000	*
Written Expression						
Learning Opportunities (0.50 - 4.00)	2.03	0.43	1.60	1.37	0.000	*
Opportunities and Materials for Writing (0.50 - 4.00)	2.71	1.02	1.69	1.83	0.000	*
Child Assessments						
Child Portfolios (1.00 - 5.00)	3.00	1.62	1.38	0.96	0.035	*
Dynamic Assessment 0.67 - 4.33)	3.18	1.77	1.41	1.24	0.008	*
Number of Classrooms	70	58				
Number of Sites	25	23				

Notes from Table A.27

*p-value < 0.05, two-tailed test.
[a] All estimates were obtained from a locally weighted kernel regression of the outcome variable on an indicator variable of ERF grant receipt; grant application score; teacher's age, education, and an indicator variable of nonwhite, using SAS's PROC MIXED procedure. Missing values of covariates were mean-imputed by site.
[b]The effect size was calculated by dividing the estimated impact by the standard deviation of the outcome measure (that is, the impact expressed as a percentage of the standard deviation).
NOTE: All figures were estimated using sample weights to account for the sample and survey designs. Standard errors of the impact estimates account for design effects due to unequal weighting of the data and clustering at site level.
SOURCE: ERF spring director and teacher surveys and classroom observations.

Table A.28. ERF impacts on selected spring teacher and classroom outcomes, 56 sites closest to cutoff value

Outcome (Range)	Funded	Unfunded	Estimated Impact[a]	Effect Size[b]	P-value of Impact	
Teachers' Earnings, Experience, and Training						
Professional Development Hours—Early Language and Literacy	69.40	22.76	46.64	1.00	0.002	*
Received professional development through mentoring/tutoring	55.97	17.93	38.03	0.79	0.006	*
Professional Development Hours—Curriculum	43.36	21.93	21.43	0.52	0.137	
Received professional development through mentoring / tutoring	45.69	14.15	31.55	0.69	0.058	
Number of Teachers	90	80				
Number of Sites	28	28				
General Quality of the Preschool Classroom						
ECERS-R Teaching and Interactions	6.03	4.65	1.37	1.25	0.001	*
Teacher sensitivity	3.20	2.47	0.73	1.06	0.009	*
Classroom community	3.28	2.54	0.74	1.06	0.006	*
Total score	2.82	1.81	1.01	1.54	0.000	*
Language, Early Literacy, and Assessment Practices						
Oral Language Use in the Classroom						
Oral Language Use by Lead Teacher (0.86–4.00)	3.04	2.14	0.90	1.16	0.004	*
Oral Language Use by Assistant Teacher (0.50–4.00)	2.88	1.66	1.22	1.02	0.020	*
Book Reading						
Number of Book Reading Sessions Observed (0–4)	1.50	1.12	0.37	0.40	0.312	
Book Reading Practices (0.56–3.94)	2.53	1.57	0.96	1.12	0.003	*
Phonological Awareness						
Number of Different Phonological Awareness Activities Observed (0–7)	2.45	0.66	1.78	1.09	0.009	*
Quality of Phonological Awareness Activities (0–4.00)	2.21	0.96	1.25	1.01	0.010	*
Print and Letter Knowledge						
Learning Opportunities (0.50–4.00)	2.12	1.16	0.96	0.96	0.024	*
Classroom Print Environment (0.50–4.00)	2.32	1.57	0.75	0.89	0.027	*
Written Expression						
Learning Opportunities (0.50–4.00)	2.05	0.75	1.30	1.12	0.004	*
Opportunities and Materials for Writing (0.50–4.00)	2.60	1.30	1.30	1.52	0.000	*
Child Assessments						
Child Portfolios (1.00–5.00)	3.13	1.70	1.43	1.05	0.010	*
Dynamic Assessment 0.67–4.33)	3.10	2.04	1.05	0.98	0.017	*
Number of Classrooms	78	72				
Number of Sites	28	28				

Notes from Table A.28

*p-value < 0.05, two-tailed test.
[a] All estimates were obtained from a regression model of the outcome variable on an indicator variable of ERF grant receipt; grant application score; teacher's age, education, and an indicator variable of nonwhite, using SAS's PROC MIXED procedure. Missing values of covariates were mean-imputed by site. Sample limited to all 28 funded sites and 28 highest scoring unfunded sites.
[b]The effect size was calculated by dividing the estimated impact by the standard deviation of the outcome measure (that is, the impact expressed as a percentage of the standard deviation).
NOTE: All figures were estimated using sample weights to account for the sample and survey designs. Standard errors of the impact estimates account for design effects due to unequal weighting of the data and clustering at site level.
SOURCE: ERF spring director and teacher surveys and classroom observations.

Table A.29. ERF impacts on selected spring teacher and classroom outcomes, no covariates

Outcome (Range)	Funded	Unfunded	Estimated Impact[a]	Effect Size[b]	P-value of Impact	
Teachers' Earnings, Experience, and Training						
Professional Development Hours—Early Language and Literacy	71.13	22.61	48.52	1.01	0.002	*
Received professional development through mentoring/tutoring	58.93	15.94	42.99	0.91	0.001	*
Professional Development Hours—Curriculum	39.75	24.76	14.99	0.38	0.211	
Received professional development through mentoring/tutoring	48.02	12.34	35.67	0.79	0.019	*
Number of Teachers	90	100				
Number of Sites	28	37				
General Quality of the Preschool Classroom						
ECERS-R Teaching and Interactions	5.92	4.74	1.18	1.09	0.001	*
Teacher sensitivity	3.15	2.51	0.64	0.95	0.008	*
Classroom community	3.32	2.52	0.80	1.18	0.001	*
Total score	2.76	1.86	0.90	1.39	0.000	*
Language, Early Literacy, and Assessment Practices						
Oral Language Use in the Classroom						
Oral Language Use by Lead Teacher (0.86–4.00)	2.98	2.19	0.79	1.06	0.004	*
Oral Language Use by Assistant Teacher (0.50–4.00)	2.74	1.77	0.97	0.83	0.036	*
Book Reading						
Number of Book Reading Sessions Observed (0–4)	1.38	1.23	0.15	0.17	0.631	
Book Reading Practices (0.56–3.94)	2.50	1.60	0.90	1.04	0.003	*
Phonological Awareness						
Number of Different Phonological Awareness Activities Observed (0–7)	2.42	0.66	1.77	1.12	0.003	*
Quality of Phonological Awareness Activities (0–4.00)	2.08	1.04	1.05	0.86	0.016	*
Print and Letter Knowledge						
Learning Opportunities (0.50–4.00)	2.04	1.23	0.81	0.82	0.031	*
Classroom Print Environment (0.50–4.00)	2.29	1.59	0.69	0.82	0.025	*
Written Expression						
Learning Opportunities (0.50 - 4.00)	1.94	0.85	1.09	0.96	0.006	*
Opportunities and Materials for Writing (0.50 - 4.00)	2.53	1.35	1.18	1.42	0.000	*
Child Assessments						
Child Portfolios (1.00–5.00)	3.05	1.75	1.30	0.95	0.012	*
Dynamic Assessment 0.67–4.33)	2.91	2.17	0.74	0.67	0.080	
Number of Classrooms	78	91				
Number of Sites	28	37				

Notes from table A.29

*p-value < 0.05, two-tailed test.
[a] All estimates were obtained from a regression model of the outcome variable on an indicator variable of ERF grant receipt and grant application score, using SAS's PROC MIXED procedure.
[b]The effect size was calculated by dividing the estimated impact by the standard deviation of the outcome measure (that is, the impact expressed as a percentage of the standard deviation).
NOTE: All figures were estimated using sample weights to account for the sample and survey designs. Standard errors of the impact estimates account for design effects due to unequal weighting of the data and clustering at site level.
SOURCE: ERF spring director and teacher surveys and classroom observations.

Table A.30. ERF impacts on selected spring teacher and classroom outcomes, no imputation of missing covariates

Outcome (Range)	Funded	Unfunded	Estimated Impact[a]	Effect Size[b]	P-value of Impact	
Teachers' Earnings, Experience, and Training						
Professional Development Hours—Early Language and Literacy	72.07	22.45	49.61	1.03	0.002	*
Received professional development through mentoring/tutoring	58.27	16.01	42.26	0.90	0.002	*
Professional Development Hours—Curriculum	40.70	24.64	16.06	0.41	0.192	
Received professional development through mentoring/tutoring	47.65	12.48	35.16	0.78	0.023	*
Number of Teachers	88	99				
Number of Sites	28	37				
General Quality of the Preschool Classroom						
ECERS-R Teaching and Interactions	5.98	4.68	1.30	1.19	0.001	*
Teacher sensitivity	3.16	2.49	0.67	0.98	0.015	*
Classroom community	3.31	2.53	0.77	1.13	0.003	*
Total score	2.72	1.86	0.85	1.28	0.001	*
Language, Early Literacy, and Assessment Practices						
Oral Language Use in the Classroom						
Oral Language Use by Lead Teacher (0.86–4.00)	3.00	2.17	0.83	1.09	0.004	*
Oral Language Use by Assistant Teacher (0.50–4.00)	2.69	1.66	1.03	0.87	0.039	*
Book Reading						
Number of Book Reading Sessions Observed (0–4)	1.47	1.25	0.21	0.23	0.571	
Book Reading Practices (0.56–3.94)	2.49	1.64	0.85	0.97	0.007	*
Phonological Awareness						
Number of Different Phonological Awareness Activities Observed (0–7)	2.40	0.61	1.79	1.08	0.004	*
Quality of Phonological Awareness Activities (0–4.00)	1.93	1.06	0.86	0.70	0.059	
Print and Letter Knowledge						
Learning Opportunities (0.50–4.00)	1.99	1.18	0.81	0.80	0.051	
Classroom Print Environment (0.50–4.00)	2.24	1.62	0.62	0.72	0.054	
Written Expression						
Learning Opportunities (0.50–4.00)	2.01	0.82	1.19	1.03	0.004	*
Opportunities and Materials for Writing (0.50–4.00)	2.50	1.40	1.11	1.30	0.000	*
Child Assessments						
Child Portfolios (1.00–5.00)	2.92	1.79	1.13	0.83	0.036	*
Dynamic Assessment 0.67–4.33)	2.83	2.22	0.61	0.55	0.182	
Number of Classrooms	69	76				
Number of Sites	28	36				

Notes from Table A.30

*p-value < 0.05, two-tailed test.
[a]All estimates were obtained from a regression model of the outcome variable on an indicator variable of ERF grant receipt; grant application score; teacher's age, education, and an indicator variable of nonwhite, using SAS's PROC MIXED procedure.
[b]The effect size was calculated by dividing the estimated impact by the standard deviation of the outcome measure (that is, the impact expressed as a percentage of the standard deviation).
NOTE: All figures were estimated using sample weights to account for the sample and survey designs. Standard errors of the impact estimates account for design effects due to unequal weighting of the data and clustering at site level.
SOURCE: ERF spring director and teacher surveys and classroom observations.

Sample Weights

We estimated our preferred classroom models with base weights that accounted for the sample design but not nonconsent and nonresponse, because information to make these adjustments was not available. Since the base weights are necessary to account for the sample design, we do not conduct any additional sensitivity tests of the weights.

Error Structure and Software Packages

We estimated our preferred classroom impact models were estimated with the SAS software package's PROC MIXED procedure, with random effects at the site level. As a sensitivity test, we estimated impacts with procedures from alternative statistical packages—SUDAAN's PROC REGRESS procedure and Stata's *svy regress* command, both of which also allowed for clustering at the site level. Estimates from both of these models are similar in magnitude and significance to those from the main classroom impact models (see Tables A.31 and A.32).

Table A.31. ERF impacts on selected spring teacher and classroom outcomes, estimated in SUDAAN

Outcome (Range)	Funded	Unfunded	Estimated Impact[a]	Effect Size[b]	P-value of Impact	
Teachers' Earnings, Experience, and Training						
Professional Development Hours—Early Language and Literacy	71.44	22.55	48.89	1.01	0.000	*
Received professional development through mentoring/tutoring	55.60	14.90	40.70	0.86	0.009	*
Professional Development Hours—Curriculum	39.59	24.87	14.72	0.37	0.143	
Received professional development through mentoring/tutoring	49.32	14.25	35.07	0.78	0.027	*
Number of Teachers	90	100				
Number of Sites	28	37				
General Quality of the Preschool Classroom						
ECERS-R Teaching and Interactions	5.92	4.76	1.16	1.08	0.000	*
Teacher sensitivity	3.14	2.51	0.63	0.93	0.012	*
Classroom community	3.32	2.51	0.80	1.19	0.003	*
Total score	2.75	1.85	0.90	1.39	0.000	*
Language, Early Literacy, and Assessment Practices						
Oral Language Use in the Classroom						
Oral Language Use by Lead Teacher (0.86–4.00)	2.97	2.19	0.78	1.05	0.003	*
Oral Language Use by Assistant Teacher (0.50–4.00)	2.73	1.78	0.95	0.81	0.031	*
Book Reading						
Number of Book Reading Sessions Observed (0–4)	1.41	1.21	0.20	0.21	0.506	
Book Reading Practices (0.56–3.94)	2.48	1.61	0.87	1.00	0.004	*
Phonological Awareness						
Number of Different Phonological Awareness Activities Observed (0–7)	2.37	0.69	1.67	1.07	0.005	*
Quality of Phonological Awareness Activities (0–4.00)	2.02	1.08	0.95	0.77	0.013	*
Print and Letter Knowledge						
Learning Opportunities (0.50–4.00)	2.01	1.24	0.76	0.78	0.017	*
Classroom Print Environment (0.50–4.00)	2.28	1.59	0.68	0.80	0.009	*
Written Expression						
Learning Opportunities (0.50–4.00)	1.96	0.81	1.15	1.01	0.001	*
Opportunities and Materials for Writing (0.50–4.00)	2.54	1.32	1.22	1.47	0.000	*
Child Assessments						
Child Portfolios (1.00–5.00)	3.08	1.71	1.37	0.99	0.002	*
Dynamic Assessment 0.67–4.33)	2.86	2.20	0.66	0.60	0.099	
Number of Classrooms	78	91				
Number of Sites	28	37				

Notes from Table A.31

*p-value < 0.05, two-tailed test.
[a] All estimates were obtained from a regression model of the outcome variable on an indicator variable of ERF grant receipt; grant application score; teacher's age, education, and an indicator variable of nonwhite, using SUDAAN's PROC REGRESS procedure. Missing values of covariates were mean-imputed by site.
[b] The effect size was calculated by dividing the estimated impact by the standard deviation of the outcome measure (that is, the impact expressed as a percentage of the standard deviation).
NOTE: All figures were estimated using sample weights to account for the sample and survey designs. Standard errors of the impact estimates account for design effects due to unequal weighting of the data and clustering at site level.
SOURCE: ERF spring director and teacher surveys and classroom observations.

Table A.32. ERF impacts on selected spring teacher and classroom outcomes, estimated in STATA

Outcome (Range)	Funded	Unfunded	Estimated Impact[a]	Effect Size[b]	P-value of Impact	
Teachers' Earnings, Experience, and Training						
Professional Development Hours—Early Language and Literacy	73.24	24.35	48.89	1.05	0.000	*
Received professional development through mentoring/tutoring	0.58	0.15	0.43	1.26	0.001	*
Professional Development Hours—Curriculum	38.02	23.31	14.72	0.56	0.143	
Received professional development through mentoring/tutoring	0.48	0.14	0.34	0.94	0.014	*
Number of Teachers	90	100				
Number of Sites	28	37				
General Quality of the Preschool Classroom						
ECERS-R Teaching and Interactions	5.99	4.83	1.16	1.16	0.000	*
Teacher sensitivity	3.19	2.56	0.63	0.92	0.012	*
Classroom community	3.38	2.57	0.80	1.16	0.003	*
Total score	2.81	1.91	0.90	1.72	0.000	*
Language, Early Literacy, and Assessment Practices						
Oral Language Use in the Classroom						
Oral Language Use by Lead Teacher (0.86–4.00)	3.04	2.25	0.78	1.07	0.003	*
Oral Language Use by Assistant Teacher (0.50–4.00)	2.77	1.82	0.95	0.90	0.031	*
Book Reading						
Number of Book Reading Sessions Observed (0–4)	1.37	1.18	0.20	0.23	0.506	
Book Reading Practices (0.56–3.94)	2.52	1.65	0.87	1.09	0.004	*
Phonological Awareness						
Number of Different Phonological Awareness Activities Observed (0–7)	2.47	0.80	1.67	1.80	0.005	*
Quality of Phonological Awareness Activities (0–4.00)	2.12	1.18	0.95	0.79	0.013	*
Print and Letter Knowledge						
Learning Opportunities (0.50–4.00)	2.05	1.28	0.76	0.99	0.017	*
Classroom Print Environment (0.50–4.00)	2.30	1.62	0.68	0.94	0.009	*
Written Expression						
Learning Opportunities (0.50–4.00)	2.00	0.86	1.15	1.38	0.001	*
Opportunities and Materials for Writing (0.50–4.00)	2.66	1.44	1.22	1.81	0.000	*
Child Assessments						
Child Portfolios (1.00–5.00)	3.17	1.81	1.37	1.16	0.002	*
Dynamic Assessment 0.67–4.33)	2.91	2.24	0.66	0.63	0.099	
Number of Classrooms	78	91				
Number of Sites	28	37				

Notes from Table A.32

*p-value < 0.05, two-tailed test.
[a] All estimates were obtained from a regression model of the outcome variable on an indicator variable of ERF grant receipt; grant application score; teacher's age, education, and an indicator variable of nonwhite, using Stata's *svy regress* procedure. Missing values of covariates were mean-imputed by site.
[b]The effect size was calculated by dividing the estimated impact by the standard deviation of the outcome measure (that is, the impact expressed as a percentage of the standard deviation).
NOTE: All figures were estimated using sample weights to account for the sample and survey designs. Standard errors of the impact estimates account for design effects due to unequal weighting of the data and clustering at site level.
SOURCE: ERF spring director and teacher surveys and classroom observations.

Adjustment for Multiple Comparisons

When impacts are estimated for multiple outcomes within a domain, it is possible that some of the estimated impacts will be statistically significant, even if there is no true effect of the intervention. For instance, when assessing statistical significance at the 5-percent level, we would expect that approximately 5 percent of the outcomes examined would be statistically significant, even if there were no true effect of the intervention, simply due to chance alone.

ED's What Works Clearinghouse has established a set of heuristics for accounting for multiple comparisons within a domain. These heuristics indicate that an impact should be considered positive and statistically significant if any one of the following conditions are met:

- Based on univariate statistical tests, at least half of the effect sizes are positive and statistically significant, and no effect sizes are negative and statistically significant.
- The omnibus impact for all the outcomes measured together is positive and statistically significant on the basis of a multivariate statistical test.
- At least one outcome remains positive and statistically significant, and no outcomes are negative and statistically significant after applying the Benjamini-Hochberg (BH; 1995) procedure to adjust significance levels downward to account for the multiple testing of impacts.
- The impact on the mean of the standardized outcome measures is positive and statistically significant.[81]

To maintain a straightforward presentation of results, the impacts presented in the main text of this report show p-values for tests of statistical significance of individual outcomes that do not reflect adjustments for multiple comparisons. The tables presented include checkmarks for domains in which impacts are jointly statistically significant once the adjustment for multiple comparisons is made. Conclusions are unaffected when we apply the procedures outlined by the What Works Clearinghouse. These procedures are relevant only to domains that contain more than one outcome; significance levels of the sole outcome in a domain are unaffected by these procedures.

[81] The standardized outcome measure is the outcome divided by its standard deviation. In cases in which a domain includes both binary and continuous outcome variables, we do not conduct this test.

Table A.33 shows the results of the multiple comparison adjustments for the child-impact analysis. We conduct these adjustments for the oral language and social-emotional domains—the only child-outcome domains that include multiple outcome measures. These adjustments indicate no evidence of statistically significant impacts in either the oral language or social-emotional development domains—none of the preceding conditions outlined by the What Works Clearinghouse heuristics are met.

Table A.33. Adjustment for multiple comparisons in child-impact analysis

	Unadjusted			Adjustments for multiple comparisons				
			Test 1	Test 2	Test 3	Test 4		
						Impact on mean of standardized outcomes in domain[b]		
Outcome (range)	Effect size[a]	P-value	At least half of impacts in domain significant?	P-value of omnibus multivariate statistical test	Statistically significant with Benjamini-Hochberg adjustment?	Impact	P-value	At least one test shows statistical significance?
Oral language			No	0.144		0.14	0.354	No
Expressive vocabulary, standard score	0.03	0.841			No			
Auditory comprehension, standard score	0.28	0.088			No			
Socioemotional development			No	0.269		0.16	0.420	No
Social competence	0.00	0.991			No			
Anxiety-withdrawal (reverse coded)[c]	0.19	0.208			No			
Anger-aggression (reverse coded)[c]	0.26	0.186			No			

*p-value < 0.05, two-tailed test.

[a]The effect size was calculated by dividing the estimated impact by the standard deviation of the outcome measure—that is, the impact expressed as a percentage of the standard deviation.

[b]The standardized outcome is the outcome divided by its standard deviation.

[c]Anxiety-withdrawal and anger-aggressions scales are reverse coded, with higher values representing less anxious-withdrawn/angry-aggressive behavior, for comparability with the social competence scale in estimating the impact on the mean of standardized outcomes in the domain.

SOURCE: ERF spring child assessment and SCBE evaluations

Table A.34 shows the results of the multiple comparison adjustments for the classroom outcome domains relating to teachers' experience and training that include multiple outcome measures. Across all adjustment procedures, there is evidence of a statistically significant impact in the teacher education and professional development domains, but no evidence of statistically significant impacts in the teaching experience domain.

Table A.34. Adjustment for multiple comparisons in classroom-impact analysis: teacher knowledge and skills

	Unadjusted			Adjustments for multiple comparisons				
			Test 1	Test 2	Test 3	Test 4		
						Impact on mean of standardized outcomes in domain[b]		
Outcome (range)	Effect size[a]	P-value	At least half of impacts in domain significant?	P-value of omnibus multivariate statistical test	Statistically significant with Benjamini-Hochberg adjustment?	Impact	P-value	At least one test shows statistical significance?
Education			Yes	0.032*		NA		Yes
Teacher's education (12–20)	0.28	0.448			No			
Bachelor's or higher degree (%)	0.63	0.016*			Yes			
Teaching experience			No	0.515		0.29	0.278	No
Years of experience at current school or center	0.32	0.248			No			
Years of experience at any preschool (0–36)	0.21	0.405			No			
Professional development			Yes	0.000*		NA		Yes
Professional development focusing on early language and literacy topics (1–60)	1.04	0.002*			Yes			
Received professional development through mentoring or tutoring (%)	0.86	0.009*			Yes			
Received professional development through workshops (%)	0.82	0.000*			Yes			
Professional development focusing on curriculum: hours (1-60)	0.39	0.209			No			
Received professional development through mentoring or tutoring (%)	0.78	0.027*			Yes			
Received professional development through workshops (%)	0.13	0.675						

*p-value < 0.05, two-tailed test.
[a]The effect size was calculated by dividing the estimated impact by the standard deviation of the outcome measure—that is, the impact expressed as a percentage of the standard deviation.
[b]The standardized outcome is the outcome divided by its standard deviation.
NA = This test is not conducted for domains that include both binary and continuous outcome measures.
SOURCE: ERF spring director and teacher surveys and classroom observations.

Table A.35 shows the results of the multiple comparison adjustments for the domains relating to the general quality of the preschool classroom. According to all four tests, there is evidence of positive and statistically significant impacts within each of these domains.

Table A.35. Adjustment for multiple comparisons in classroom-impact analysis: general quality of the preschool classroom

	Unadjusted			Adjustments for multiple comparisons				
			Test 1	Test 2	Test 3	Test 4		
						Impact on mean of standardized outcomes in domain[b]		
Outcome (range)	Effect size[a]	P-value	At least half of impacts in domain significant?	P-value of omnibus multivariate statistical test	Statistically significant with Benjamini-Hochberg adjustment?	Impact	P-value	At least one test shows statistical significance?
Quality of teacher-child interactions			Yes	.003*		1.05	0.006*	Yes
Teaching and interactions (ECERS-R)	1.12	0.001			Yes			
Teacher sensitivity (TBRS) (0.50–4.00)	0.99	0.008			Yes			
Quality of team teaching (TBRS)	0.79	0.049			Yes			
Organization of the classroom environment			Yes	.009*		1.24	0.001	Yes
Classroom community (TBRS)	1.22	0.001			Yes			
Quality and organization of activity centers	1.13	0.003			Yes			

*p-value < 0.05, two-tailed test.
[a]The effect size was calculated by dividing the estimated impact by the standard deviation of the outcome measure—that is, the impact expressed as a percentage of the standard deviation.
[b]The standardized outcome is the outcome divided by its standard deviation.
SOURCE: ERF spring director and teacher surveys and classroom observations.

Table A.36 shows the results of the multiple comparison adjustments for the domains relating to the quality of language, early literacy, and assessment practices and environments. According to all four tests, there is evidence of positive and statistically significant impacts within each of these domains.

Table A.36. Adjustment for multiple comparisons in classroom-impact analysis: quality of language, early literacy, and assessment practices and environments

	Unadjusted			Adjustments for multiple comparisons				
			Test 1	Test 2	Test 3	Test 4		
						Impact on mean of standardized outcomes in domain[b]		
Outcome (range)	Effect size[a]	P-value	At least half of impacts in domain significant?	P-value of omnibus multivariate statistical test	Statistically significant with Benjamini-Hochberg adjustment?	Impact	P-value	At least one test shows statistical significance?
Quality of the oral language environment			Yes	0.011*		1.03	0.013*	Yes
Oral language us by lead teacher	1.11	0.002			Yes			
Oral language use by assistant teacher	0.89	0.027			Yes			
Book reading			Yes	0.019*		0.76	0.036*	Yes
Number of book reading sessions observed	0.23	0.516			No			
Book reading practices (0.56–3.94)	1.03	0.003			Yes			
Phonological awareness activities			Yes	0.013*		1.04	0.005*	Yes
Number of different phonological awareness activities observed (0–7)	1.1	0.004			Yes			
Quality of phonological awareness activities	0.79	0.024			Yes			
Print and letter knowledge activities and materials			Yes	0.007*		1.01	0.005*	Yes
Learning opportunities (0.50–4.00)	0.87	0.022			Yes			
Classroom print environment (0.50–4.00)	0.81	0.028			Yes			
Written expression activities and materials			Yes	0.001*		1.24	0.000*	Yes
Learning opportunities (0.50–4.00)	1.06	0.003			Yes			
Opportunities and materials for writing	1.48	0.000			Yes			
Child screening and progress assessment			Yes	0.078		0.82	0.039*	Yes
Child portfolios (1.00–5.00)	0.98	0.012			Yes			
Dynamic assessment (0.67–4.33)	0.64	0.095			No			

*p-value < 0.05, two-tailed test.

[a]The effect size was calculated by dividing the estimated impact by the standard deviation of the outcome measure—that is, the impact expressed as a percentage of the standard deviation.

[b]The standardized outcome is the outcome divided by its standard deviation.

SOURCE: ERF spring director and teacher surveys and classroom observations.

Appendix B. Data-Collection Methods

The data analyzed for this evaluation were obtained through child assessments; classroom observations; and surveys of teachers, center directors, and parents. We collected these data at two times: fall 2004 and spring 2005. We conducted in-depth interviews with the project directors of the funded sites in the spring of 2005. We collected attendance data from preschools for the students included in the assessment sample. This appendix describes the methods used for recruiting sites; training staff to conduct classroom observations, child assessments, and parent interviews; and collecting and processing data.

Institutional Review Board

In 2004, both the federal Office of Management and Budget and the Institutional Review Board (IRB) of Public/Private Ventures (P/PV) approved the design, parental consent procedures, and data-collection methods and instruments for this study. The P/PV IRB approval was updated in 2005 and 2006. The P/PV IRB was contracted to provide this review function because the prime evaluation contractor does not maintain its own internal IRB.

Site Recruitment Procedures

In April 2004, senior staff at DIR and MPR began recruiting ERF grantees and applicants from the FY 2003 cohort. We recruited the comparison group from unfunded ERF applicants. We ranked all unfunded applicants in descending order according to the score ED awarded their application. We recruited unfunded applicants with application scores of 44 or higher to participate in the study. Initially, we sent letters from ED's Institute of Education Sciences (IES) to the project directors of grantees sites and the center directors or principals of unfunded applicants to introduce the evaluation and request the cooperation of grantees and unfunded applicants. We also sent grantees a letter from the ERF program staff within the Office of Elementary and Secondary Education, requesting their participation in the evaluation. DIR and MPR site recruiters followed these advance letters within a week with telephone calls.

The site recruiters followed a prepared script designed to:

- Identify the appropriate person to talk with about study participation
- Introduce the key elements of the study design and data collection
- Explain the responsibilities associated with study participation and describe the incentives, if any, that would be available to participants in the study
- Collect data about all of the preschools and classrooms serving 4-year-old children, including the enrollment process and school schedule
- Discuss next steps regarding contacting the individual preschools that might be involved and obtain a Memorandum of Understanding (MOU) that documented responsibilities and roles for the study participants and the evaluation team

Once project directors verbally agreed to participate in the study, the most challenging aspect of the recruitment process was obtaining signed MOUs from sites. In some cases, school districts required their research review committees to examine and approve the request for study participation. In other cases, school-district superintendents had to approve preschools' participation in the study. For sites in which multiple jurisdictions were involved—for example, collaborations of school districts, nonprofit providers, and other agencies—approval was required from each participating organization.

If unfunded applicants continued to have responsibility or oversight for the preschools that were included in their application, recruitment efforts focused on obtaining the cooperation of individuals with decision-making authority—typically directors of early childhood centers or assistant superintendents in school districts. However, in the 2003 ERF grant application, ED encouraged collaborations of diverse types of preschools within an area (for example, school-district-administered preschools, Head Start centers, and independent child-care centers). In many cases, unfunded applicants did not exercise management control of preschools that collaborated in the grant application. Preschools that had been part of these FY 2003 grant applications were recruited individually by members of the evaluation team. The need to obtain multiple organizational approvals was greater among unfunded applicant sites where the original applying agency was no longer involved with the preschool programs listed in their applications.

In order to obtain a sufficient sample size, site recruitment for the unfunded applicants continued into early fall 2004. Unfunded sites were given a financial incentive for each classroom that was enrolled in the sample to compensate for distributing and returning parent consent forms and facilitating access to classrooms for assessments and observations.

Table B.1 shows the number of sites (funded and unfunded) that the study team attempted to recruit.[82] Table B.2 displays the participation of preschools that correspond to those sites. Five unfunded sites and their associated 25 preschools were removed from the sample because they received a grant in a subsequent round of ERF funding.[83] Of the 62 remaining unfunded sites that were contacted, 37 sites (60 percent) contained at least one preschool that participated. At the preschool level, however, the participation rate was lower. Only 129, or 46 percent, of potentially available preschools agreed to participate in the study.

[82] Several unfunded sites were not recruited. The lowest scoring 23 applicants—those that scored below 42.5—were not contacted during the recruiting process. In addition, 3 unfunded sites were excluded because they did not meet the criteria for participation in the study (one applicant served only deaf children; one applicant proposed to provide only wraparound care consisting mainly of lunch and nap; and one applicant served only migrant children).

[83] Five unfunded sites were removed because they were awarded 2004 ERF grants for classes that overlapped with 2003 unfunded classrooms. Another four unfunded sites that later received grants in 2004 were included in our sample because there was little to no overlap between the classrooms listed in their 2003 and 2004 applications.

Table B.1. Site agreement to participate in the ERF national evaluation

Participation status	Funded sites	Unfunded sites
Site agreed	28	37
Site refused	0	26
Site replaced because it received a grant in 2004	0	5
Total sites contacted	28	68

Table B.2. Preschool agreement to participate in the ERF national evaluation

Participation status	Funded preschools	Unfunded preschools
Site agreed; preschools agreed and selected into sample	86	75
Site agreed; preschools agreed but no classes selected into sample	70	46
Preschool refusals	1	125
Preschools and sites removed by request from ED program office	9	8
Preschools removed because site received grant in 2004	0	25
Total preschools eligible for study	157	246

Using census data aggregated to the ZIP code level, we examined the characteristics of the areas in which the recruited sites and preschools were located, to see how the participating sites compared to those who refused to participate. Compared to those that did not agree to participate or were removed from the sample, the preschools that agreed to participate had higher ERF grant competition scores (72.3 versus 61.3); a larger percentage of the population of their ZIP codes was white non-Hispanic (60 percent versus 55 percent); and a larger percentage was located in an urban area (88 percent versus 79 percent). However, the two groups were very similar in terms of percent black, percent Hispanic, median income, poverty rate, and unemployment rate of the ZIP code area (see Table B.3).[84]

Table B.3. ZIP-code characteristics of participating versus nonparticipating preschools

	Agreed to participate	Refused to participate or dropped by ED	Difference	P-value difference	P-value of difference of conditional score
Average application score	72.3	61.3	11.0	0.000	—
Percent urban	87.7	79.3	8.3	0.016	0.139
Average percent white	59.9	54.9	5.1	0.030	0.011
Average percent black	22.0	22.2	–0.2	0.936	0.407
Average percent Hispanic	21.0	22.1	–1.1	0.620	0.033
Median household income	39.6	40.6	–0.9	0.482	0.435
Poverty rate	19.8	19.0	0.9	0.355	0.714
Unemployment rate	8.5	9.0	–0.4	0.371	0.160
Number of preschools	285	187			

[84] Preschool-level demographic data were unavailable from the applications.

We also examined the distribution of grant application scores among the unfunded applicant group to determine whether sites that agreed to participate in the study had a different distribution of scores than those who refused. This analysis indicated that cooperating and noncooperating sites had similar score distributions, suggesting that those who refused to participate and those who agreed to participate may be similar.

From the 28 ERF grantees and 37 unfunded applicants that agreed to participate in the study, we selected a sample of classrooms with probability proportional to the number of 4-year-old students. Although the sample was designed so that 3 classrooms per grantee would be selected, more classrooms were selected in some sites and fewer in others.

Table B.4 shows the distribution across sites of the number of classrooms that were selected for and agreed to participate in the study.

Table B.4. Distribution of the number of classrooms

Number of classrooms per site	Funded sites	Unfunded sites
1-classroom sites	0	0
2-classroom sites	1	6
3-classroom sites	14	14
4-classroom sites	8	13
5-classroom sites	3	4
6-classroom sites	2	0
Number of sites in study	28	37
Number of classrooms in study	103	126

Obtaining Parental Consent. After the selected funded and unfunded applicant sites and classrooms in the sample agreed to participate, the study team worked to secure signed parental consent by using the forms and procedures approved by the study's Institutional Review Board. We sent English and Spanish consent forms to teachers and asked them to distribute the forms to all children in their classrooms. The forms were printed on 2-ply carbonless paper so that parents could keep a signed copy. The consent forms provided parents a written explanation of the study and requested that they consent to their child's participation in the study by signing the forms. Parents were also asked to provide their children's date of birth. The signed original parental consent forms were returned by overnight mail to DIR. Data from the consent forms were entered in DIR's study database.

We used these data to determine children's age eligibility; select the evaluation sample (that is, who would be assessed) according to the sampling levels specified for the classroom; and create labels for classroom observations and child assessments. The children's eligibility for the study was based on whether, as determined by their birthdates and local age cutoffs for kindergarten, they were likely to enter kindergarten in the next school year. The parents of approximately 2,840 children (79 percent of the children enrolled in participating classrooms) consented. From the age-eligible children with parental consent, approximately 1,900 were selected into the sample. Table E-5 shows the return rate for parental consent forms.

Table B.5. Status of returned parental consent forms

	Funded sites	Unfunded sites
Total received	1,454	1,630
% agreed for child to participate	93.2%	94.7%
% of children age eligible	79.6%	73.1%

Response Rates for Study Respondent Groups

Assessment and Parent Survey Response Rates. Child assessments were administered by trained assessors during prescheduled site visits. A team of assessors typically completed all of the assigned assessments in a classroom over a 1- or 2-day period. Teachers were asked to complete a social competence and behavior evaluation (SCBE) rating form for all students in their classroom who were participating in the study. A small monetary incentive was provided to teachers for each rating form they returned. Telephone interviewing of parents in each site began soon after the child assessments began in that site. All parents received a small monetary incentive for completing the telephone survey. Response rates were above 85 percent for both the child assessments and the teachers' ratings of children's social-emotional behavior and approximately 61 percent for the parent surveys (see Table B.6).

Table B.6. Data-collection recruitment and response rates: children and parents

	Funded sites	Unfunded sites	Total
Eligible sample of students and parents	935	979	1,914
Language and Literacy Skill Assessments			
Assessments completed (spring)	803	855	1,668
% of students assessed	85.9%	87.3%	87.1%
Social Competence and Behavior Evaluation Assessment			
SCBE rating forms completed (spring)	802	843	1,645
% of students with SCBEs	85.8%	86.1%	85.9%
Parent Survey			
Parent surveys completed (spring)	574	603	1,177
% of students with parent data	61.4%	61.4%	61.4%

Teacher and Director Response Rates. Up to three classrooms in each site were selected for classroom observation. If child assessments were conducted in more than three classrooms in a site, then three were randomly selected for observations. The observations were conducted by trained staff, who typically completed the observation battery in a 3-hour scheduled visit to the selected classroom. In addition, all teachers and preschool directors whose students were included in the child sample were asked to complete surveys. The surveys were sent to center directors for distribution to teachers. Return mailing materials were provided in order for center directors and teachers to return the completed instruments directly to the evaluation contractor. Teachers received a small monetary incentive for returning the completed questionnaire. Response rates for both teacher and director surveys were high (close to 90 percent of attempted surveys completed in both funded and unfunded sites, as shown in Table B.7). Attendance data were requested from all of the preschools but were provided at a higher rate by the funded sites.

Table B.7. Data-collection results: teachers and directors

	Funded sites	Unfunded sites	Total
Classroom Observations			
# of classrooms in sample	103	126	229
Observations completed (spring)[1]	78	91	169
Teacher Surveys			
Teacher surveys attempted [2]	98	114	212
Teacher surveys completed (spring)	92	102	194
% of teachers surveyed	93.9%	89.5%	91.5%
Center Director Surveys			
Number of center director surveys attempted	76	74	150
Center director surveys completed (spring)	64	68	132
% of centers surveyed	84.2%	91.9%	88.0%
Classroom Attendance Records			
Classroom attendance records returned	91	91	182
% of classes reporting attendance	92.9%	78.4%	85.0%
% of students for whom attendance data was reported	86.0	73.4	79.6

[1]In sites with 4-6 classrooms, three classrooms were randomly selected for observation
[2]Some teachers taught multiple classes (for example morning and afternoon half-day sessions). In those instances, only one survey was attempted with the teacher to gather information referencing only one of their randomly selected classes.
SOURCE: ERF spring assessments and observations.

Hiring and Training of Assessment and Observation Data-Collection Staff, Including Quality Assurance

Field staff for conducting the child assessments and classroom observations were recruited nationally. Persons with experience in conducting assessments and other data collection with children, observing classrooms, and working in preschools or other educational settings were given highest priority. For fall 2004, field staff were hired to conduct assessments, record observations, or serve as members of the quality-assurance staff. In the spring, some staff who worked in the fall were hired to do both assessments and observations. All field staff were trained before collecting data during both the fall of 2004 and spring of 2005. Separate training sessions were held for assessors and observers. The 5-day fall 2004 child-assessment training conducted by CIRCLE and DIR personnel included the following sessions:

- background about ERF and the evaluation
- general information about conducting pre-K assessments
- proper administration of the Pre-LAS
- proper administration of the Elision and Print Awareness subtests of the Preschool Comprehensive Test of Phonological and Print Processing (Pre-CTOPPP)
- proper administration of the Expressive One-Word Picture Vocabulary Test (EOWPVT)
- proper administration of the Preschool Language Scale-IV (PLS-IV)
- proper administration of bilingual assessments
- quality assurance
- live practice sessions with DIR and CIRCLE staff
- administrative procedures, including travel, responsibilities, and compensation
- final certification (which consisted of conducting assessments with 2 children from 3 to 5 years of age)

The 6-day fall 2004 classroom observation training conducted by personnel from DIR, CIRCLE, MPR, and the Frank Porter Graham Center included the following sessions:

- background about ERF and the evaluation
- pre-K education and early academic development
- the Early Childhood Environmental Rating Scale-Revised (ECERS-R) instrument
- the Teacher Behavior Rating Scale (TBRS)
- live classroom observations
- quality assurance
- administrative procedures, including travel, responsibilities, and compensation
- final certification

The training for assessors and observers was repeated in spring 2005 and was similar to the fall training, except that the spring observer training was completed in five days. Table B.8 presents the number of assessors and observers who were trained or cross-trained during fall 2004 and spring 2005. In both the fall and spring, we did not extend field data-collection contracts to roughly 10 percent of the individuals hired for training to conduct child assessments and classroom observations, because they did not complete training satisfactorily. Classroom observers were required to attain an inter-rater agreement level of .90 with a trainer in order to be certified to begin working.

Table B.8. Number of persons trained as assessors and observers

	Classroom observers trained	Child assessors trained	QC observers trained	QC assessors trained	Cross-trained QCO/QCA	Cross-trained CO/CA
Fall 2004	17	47	6	7		
Spring 2005	15	45	1	2	4	8

Data Collection

Assessments and observations. For fall 2004, child assessments and classroom observations were conducted from October through December. For spring 2005, child assessments and classroom observations were conducted from March through June. Data-collection procedures were the same at all sites, regardless of whether the site received ERF funding.

Four DIR field supervisors were assigned specific sites and were responsible for scheduling child assessments and classroom observations. The field supervisors maintained ongoing contact with appropriate site and preschool personnel to ensure that parental consent forms had been completed and returned and that observers and assessors would be able to collect data as agreed.

Typically, one observer conducted up to three classroom observations per site. During the first two weeks of classroom observations, quality-assurance staff monitored at least two classroom observations performed by each observer at a site; this monitoring ensured that the reliability established during training had not decreased. The number of classroom observations completed by observers during one round of data collection ranged from 1 to 23, with a mean of 11 observations completed by observers during each data-collection period.

Child assessors worked as 3-member teams. Whether the team members worked simultaneously at one school or at several schools at once depended upon the number of children to be assessed in a preschool and the geographic location of the selected preschools in the site. The number of assessments completed by assessors during each round of data collection ranged from 1 to 114, with a mean of 31 assessments completed by each assessor during each round of data collection.

Surveys of teachers and preschool/center directors. For the fall data collection, survey data were obtained from teachers and preschool/center directors from October 2004 through January 2005. During spring 2005, we collected survey data from teachers and preschool/center directors from March 2005 through June 2005. We sent questionnaires for teachers and preschool/center directors to each site for distribution by grantee project directors or the preschool/center directors; the questionnaires were self administered. In addition to the surveys, teachers also completed SCBE forms for each of their students participating in the study. We sent grantee project directors and preschool center directors mailing materials to return documents to DIR.

Teachers and preschool/center directors were invited to call DIR's toll-free help line if they had questions about or difficulties with completing the surveys, the SCBEs, or returning the materials to DIR. The field supervisors made numerous calls to preschool/center directors and teachers to secure the return of completed surveys and SCBEs.

Parent survey. We contacted parents or guardians of students participating in the study by telephone to complete the parent survey. We made all call attempts from the telephone center at DIR and used a survey that was programmed for computer-assisted telephone interviewing (CATI) by using Sawtooth's WINCATI software.

All interviewers were trained and certified before conducting the survey. DIR interviewer training included:

- an introduction to ERF
- general interviewing techniques
- how to contact sample members for interviewing
- procedures for assuring respondent confidentiality
- a question-by-question review of the survey
- how to use face sheets and set disposition codes
- how to respond to frequently asked questions

To contact parents or guardians, interviewers first used the telephone number recorded by the parent on the returned parental consent form. Initially, the parent listed on the parent consent form was the first person contacted to complete the survey. However, if that person was not available, interviewers were instructed to ask for another parent or guardian of the child in the sample. If interviewers were unable to contact parents or guardians at that number, they made efforts to obtain updated telephone contact information. To increase survey response rates, follow-up postcards with DIR's toll-free number were sent to parents and guardians to encourage them to complete the survey. All parents and guardians who completed the survey were sent $10 gift cards as a way to thank them for participating in the study. Parent interviews were conducted for fall 2004 from October through January 2005. In spring 2005, parent surveys were conducted from April through July 2005. Final dispositions of parent survey attempts are shown in Table B.9.

Table B.9. Final disposition codes—spring parent survey

	Funded sites	Unfunded sites	Total sites
Parent surveys completed (spring)	574	603	1,177
% of eligible students with parent data	61.4	61.4	61.4
% refused	5.0	5.7	5.4
% unable to locate or contact	33.5	33.9	33.7

In-depth interviews with grantees. We conducted in-depth telephone interviews between May and July 2005 with project directors of the 28 ERF grantees for FY 2003 who participated in the study. Often, other staff members who participated in implementing the ERF grant joined the project directors on the call. These hour-long interviews provided background about the context in which the ERF grants were implemented.

Attendance data. In the spring of 2005, we sent grantee project directors and preschool center directors forms to document student attendance during the 2004–2005 school year. Attendance data collected for each student included the number of days attended during the fall and spring semesters and the date that students began school if later than the start date for the 2004–2005 school year.

Data Processing, Including Entry and Quality Assurance

A quality-assurance assessor or observer accompanied child assessors and classroom observers on their earliest data-collection assignments and reviewed the procedures used and forms completed in the initial child assessments and classroom observations. This initial quality-assurance check provided an opportunity for refresher training, if needed, and identified staff members whose field practices did not reflect the practices that were taught and modeled during training. After initial quality-assurance reviews, assessors and observers were expected to edit their own work for completeness, accuracy, and legibility. Each week, assessors and observers shipped data they collected by overnight delivery to DIR. At DIR, research assistants logged in and reviewed data for completeness.

After DIR's research assistants checked the data, field supervisors conducted thorough quality-assurance reviews of the data returned by observers and assessors from their sites. Field supervisors also contacted assessors and observers to resolve questions about data entered on the classroom observation and child assessment forms that they submitted. All quality-assurance problems were resolved by field supervisors in consultation with the data-collection manager before materials were sent to CIRCLE for data entry.

Supervisors in DIR's CATI center monitored parent telephone interviews to ensure that surveys were administered completely and properly and that all data were recorded correctly. Supervisors used an on-line telephone monitoring system to simultaneously hear interviewers ask questions and view their survey screens as they entered data from respondents during interviews. In this way, supervisors could verify that interviewers administered questions and coded responses properly.

Field supervisors also reviewed all teacher and preschool-director surveys and SCBE rating forms. DIR's data-entry clerks entered data from teacher and preschool-center director surveys into a database.

Classroom observation, child assessment, and the SCBE rating forms were sent to CIRCLE for scanning and creating raw data files.

Raw data files produced by DIR and CIRCLE were used for analyses. MPR also used these raw data files to create additional analysis files. These data files were reviewed to identify and correct errors, inconsistencies, or erroneous entries.

Methods for Calculating ERF's Cost Allocation per Child

Data provided by the ERF programs were used to estimate the annual per-student cost for the FY 2003 ERF grantees. The number of children "planned" to be served by ERF and the amount of the grantees' 3-year ERF award were included in these estimates. Calculations of the number of children "planned" to be served by ERF were based on estimates of the total number of children (of all ages) in the ERF-funded sites as reported in phone interviews conducted by DIR and MPR site recruiters with ERF project directors during the spring and summer of 2004 and on estimates of the number of students to be served as reported in the grant applications.

The two sources (interviews with project directors and grant application estimates) provided comparable estimates of the total number of children to be served annually through ERF funds. When aggregated, the numbers provided by project directors totaled 9,196 students, and estimates obtained from grant applications totaled 9,083 students. At the individual grantee level, there were fairly wide discrepancies in the estimates of the number of students to be served. However, these grantee-level differences offset each other, resulting in similar overall estimates.

The dollar value of the 3-year grant application was assumed to be equally divided across each of the three years of funding. That annual amount was then used in conjunction with the number of children served in ERF-supported classrooms to compute the following items:

- Average cost per student served across the grantees (weighted average)
- Median cost per student served across the grantees
- Average cost per student served for the 30 grantees (unweighted average)

Table B.10 shows these results based on estimates obtained from project directors and grant applications.

Table B.10. ERF annual costs per student in FY 2003 funded cohort

	Estimated using project director's estimates of children to be served	Estimating using grant application estimates of children to be served
Average cost per student	$2,714	$2,748
Median cost per student	$3,549	$2,856
Average of the grantees	$3,648	$3,143

The estimated average cost per student served in ERF-supported classrooms ranged from $2,500 to $3,500. Two caveats are appropriate in examining these per student costs. First, the grants include funds for required local evaluations, and some portion of those costs should be excluded from estimates related to providing services. Second, this estimate assumes that ERF grantees received no in-kind or financial support from sources other than the ERF grant. There was no reliable source of information to determine other sources of support used by ERF-funded programs or the amount that grantees allocated for evaluation.

Appendix C. Assessment and Observation Measures Used for ERF Data Collection

This appendix describes the child-assessment and classroom-observation instruments that were used in the National Evaluation of ERF. We describe the criteria used to select the instruments, their use in other studies, and their psychometric properties. We selected the child assessments to align with the goals of the ERF program for the development of children's language and early literacy skills. We also included measures of children's social-emotional development to examine the effects of an early literacy focus on this aspect of development. We selected measures of general classroom quality, including teacher behaviors and classroom environment, that previous research has found to be positively correlated with young children's cognitive skills and emotional development (Vandell and Wolfe, 2000; NICHD Early Childhood Research Network, 2002, 2003, and 2006). Further, we selected classroom observation measures of teacher instructional practices and classroom environment that are closely related to ERF's emphasis on language and emerging literacy skills.

This study's Technical Working Group provided critical input and made important contributions to the final decisions on instrumentation.

Child-Assessment Instruments

A maximum of 45 minutes was allotted for administering the full child-assessment battery in order to limit the burden to the children being tested. Although we made decisions about specific language and literacy measures to include in the ERF battery according to skills deemed necessary for successful reading, we considered following additional factors:

- Time required to administer the instruments
- Training required for staff to administer the instruments
- Qualifications that examiners needed so that appropriate and adequate staff were trained and available
- Sensitivity of the measures to change as a result of the intervention
- Appropriateness of the measure for a diverse population including racial and ethnic minorities, language minorities, and economically disadvantaged children
- Costs of the measures for the sample sizes
- Comparability of the measures to other national evaluation studies (especially other current early literacy intervention studies)
- Psychometric qualities of the measures under consideration, including adequate reliability and validity, with minimal floor or ceiling effects for low-income preschool children
- Availability of a Spanish-language version of assessment

The reading research literature that informed the selection of measures to use in the ERF evaluation indicated that there were strong correlations between preschool children's acquisition of oral language skills (particularly vocabulary and grammar) and phonological awareness, print and letter knowledge, and reading ability (Whitehurst and Lonigan 2001; Pullen and Justice 2003). The final measures selected for child assessment provided a balanced evaluation of the skills necessary for successful reading. The measures used to assess children's language,

phonological processing, print and letter knowledge, and social-emotional development are presented in the following sections.

Language

Three measures—the Pre-LAS, the Auditory Comprehension Scale of the Preschool Language Scale-IV, and the Expressive One-Word Picture Vocabulary Test—were used in the National Evaluation of ERF to assess children's language skills during fall 2004. In spring 2005, only two of these measures—the Auditory Comprehension Scale of the Preschool Language Scale-IV and the Expressive One-Word Picture Vocabulary Test—were used.

Pre-LAS 2000 (Pre-LAS): The Pre-LAS is an interactive measure of oral-language proficiency and preliteracy skills for children of all languages. The English version of the Pre-LAS was used as a language assessment screener during fall 2004 data collection to guide assessors in determining whether children understood enough English to be administered the complete English version of the ERF battery. The screener, the Pre-LAS Oral Component (the "Simon Says" subtest), is designed for children ages 4–6. The "Simon Says" subtest evaluates receptive language (that is, listening) skills and the ability to follow simple oral instructions through total physical responses (for example, "Simon Says put your hand on your head").

The criterion for using an English- or Spanish-language assessment in the National Evaluation of ERF was consistent with the criteria used in two other national studies of early childhood programs, the Head Start FACES 2003 study (U.S. Department of Health and Human Services December 2006) and the Head Start Impact Study (U.S. Department of Health and Human Services May 2005). That is, if children answered 6 out of the 20 items correctly, they were assessed in English. During fall 2004, Spanish-speaking children who made 15 or more errors on the 20 total items were administered all assessments in Spanish. No children who could not be assessed in English needed to be assessed in a language other than Spanish.

Preschool Language Scale-IV (PLS-IV): The Auditory Comprehension Scale of the Preschool Language Scale-IV was used in the ERF evaluation to provide a measure of children's language comprehension skills. We used the PLS-IV to assess complicated forms of language (for example, structure, grammar, and syntax) and receptive vocabulary. According to the PLS-IV manual (Zimmerman, Steiner, and Pond 2002), stability coefficients (test-retest reliability at a mean of a 5.9-day interval between the two testing sessions) for the Auditory Comprehension Subscale for ages 4 years to 5 years 11 months range from .83 to .91. Reliability coefficients for internal consistency for the Auditory Comprehension Subscale for ages 4 years to 5 years 11 months range from .83 to .90. The Auditory Comprehension Subscale was normed on a nationally representative sample of children of various ages so that raw scores can be converted to age-adjusted, standardized scores with a mean of 100 and a standard deviation of 15.

According to the authors, the PLS-IV has convergent validity with the DENVER II. The DENVER II was developed to assess language-development skills, language disorders, and psycholinguistics. Children who earned a "normal" rating on the DENVER II all scored within one standard deviation of the mean on the PLS-IV (sample size = 37).

Expressive One-Word Picture Vocabulary Test (EOWPVT): The EOWPVT is an assessment of English-speaking expressive vocabulary and can be used for individuals between the ages of 24 months and 18 years 11 months. Children are asked to name objects, concepts, and actions. The author (Brownell 2000) reports that the measure is internally consistent: coefficient alpha based on intercorrelations among test items (median of .96) and split-half reliability (median of .98). The EOWPVT also has high test-retest reliability based on an average time lag of 20 days between test administrations (for ages 4–6 yrs, mean alpha = .95). Inter-rater reliability is also high (reliability of scoring = 100 percent; reliability of response evaluation = 99.4 percent). The EOWPVT-III was normed on a nationally representative sample of children of various ages so that raw scores can be converted to age-adjusted, standardized scores with a mean of 100 and a standard deviation of 15.

Correlations with other measures of expressive language, measures of other areas of language development, academic achievement, and general cognitive ability were found to range from .64 to .90.

Print Concepts and Letter Knowledge

Two measures from the Preschool Comprehensive Test of Phonological and Print Processing (Pre-CTOPPP)—the Elision subtest and Print Awareness subtest—were used in the National Evaluation of ERF during fall 2004 and spring 2005 to assess children's print processing and print and letter knowledge. The ERF evaluation used a research version of the test available in 2004, for which national norms are not available. However, a slightly revised version of the test with normed scores is now available from a publisher, ProEd, and is called the Test of Preschool Early Literacy (TOPEL).

Pre-CTOPPP Elision Subtest: The Pre-CTOPPP's Elision subtest (Lonigan, Wagner, Rashotte 2002) was used to evaluate phonological processing abilities in the ERF evaluation. It was designed for children as young as three years old as a downward extension of the Comprehensive Test of Phonological Processing (CTOPP—Wagner, Torgesen, and Rashotte 1999). Like the CTOPP, the Pre-CTOPPP provides assessment of all three areas of phonological processing: phonological sensitivity, phonological memory, and phonological access.

Standardized scores cannot be computed for the Pre-CTOPPP Elision subtest, because national norms for this version of the subtest are not available. National norms for the revised TOPEL Phonological Awareness subtest (which combines the Pre-CTOPPP Elision and Blending subtests) cannot be used directly to standardize the Pre-CTOPPP Elision scores, because of substantive differences in content, question order, stopping rules, and administration procedures between the two versions.

Data on the reliability of the pre-CTOPPP Elision subtest are not available for a nationally representative sample, but data are available from large-scale data collection in four federal early childhood studies. The Pre-CTOPPP Elision subtest had high reliability in the sample children assessed in this evaluation, with Cronbach's alpha equal to 0.7123. In addition, the subtest had high reliability in three ongoing federal studies, with Cronbach's alpha equal to 0.88 for four-year-olds in the Head Start Impact Study, Cronbach's alpha equal to 0.81 for three- and four-year-olds in the IES Even Start Classroom Literacy Interventions and Outcomes Study, and

Cronbach's alpha equal to 0.83 in Fall 2003 and 0.88 in Spring 2004 for four-year-olds in the IES Preschool Curriculum Evaluation Research Study.[85]

Pre-CTOPPP Print Awareness Subtest: The Pre-CTOPPP's Print Awareness subtest (Lonigan, Wagner, Rashotte 2002) was used as a measure of children's print and letter knowledge skills in the ERF evaluation. The Print Awareness subtest contains the following types of items: print concepts, letter discrimination, word discrimination, letter-name identification, and letter-sound identification.

National norms are not available for the Pre-CTOPPP Print Awareness subtest used for the ERF evaluation. However, norms from the revised TOPEL Print Knowledge version of the test can be used to derive age-adjusted, standardized scores for the research version of the Print Awareness subtest. The two versions contain the same questions but in a different order and with different stopping rules. Because the National Evaluation of ERF administered all items of the Pre-CTOPPP Print Awareness subtest with no stopping rules, we applied the TOPEL scoring rules retroactively to the data to obtain comparable raw scores for the TOPEL Print Knowledge test and then translated those scores into standardized scores by using information from the test's publisher. The TOPEL Print Knowledge subtest has high internal consistency reliability (.95) and high test-retest reliability (.89) (Lonigan, Wagner, et al. 2007).

Social-Emotional Behavior

Social Competence and Behavior Evaluation (SCBE): To assess children's social-emotional development, we used the 30-item Social Competence and Behavior Evaluation (SCBE-30; LaFreniere and Dumas 1996), which was modified from the longer 80-item version of the SCBE (La Freniere, Dumas, Capuano, and Dubeau 1992)—also available in Spanish (Dumas, Martinez, and La Freniere 1998). The 30-item teacher version has three subscales—Social Competence, Anxiety-Withdrawal, and Anger-Aggression. SCBE-30 was designed for use with children from 2.5 years old to about 6 and has been successfully validated and used in numerous studies in a number of countries (La Freniere and Dumas 1996; La Freniere et al. 2002) and intervention studies (La Freniere and Capulano 1997). The internal consistency coefficients reported for the SCBE's subscales range from .80 to .92 (La Freniere and Dumas 1996). These scales have been used in studies of young children's adjustment (Denham, Caverly et al. 2002; Denham and Burton, in press; La Freniere and Dumas 1996; La Freniere et al. 2002).

Classroom-Observation Measures

We obtained measures of the classroom environment and instructional practices through direct observation of the classroom and teacher. We allotted approximately four hours for observations in each preschool classroom. We completed observations of up to three classrooms per site in the fall and spring. The observation protocols included the Teacher Behavior Rating Scale (TBRS), developed by the Center for Improving the Readiness of Children for Learning and Education

[85] Cronbach's alpha coefficients are from unpublished tabulations using child assessment data from the Head Start Impact Study (U.S. Department of Health and Human Services, 2005), and the forthcoming Even Start Classroom Observations and Interventions and Preschool Curriculum Evaluation Research studies being conducted by the Institute of Education Sciences.

(CIRCLE) at the University of Texas-Houston, and a subset of items from the Early Childhood Environment Rating Scale-Revised (ECERS-R) (Harms, Clifford, & Cryer 1998). The TBRS was developed to evaluate the early literacy and language qualities in preschool classrooms, but it also includes subscales that measure the general quality of the classroom and the sensitivity of teacher behavior. We included 11 ECERS-R items that compose the subscale, Teaching and Interactions, formed by a factor analysis of the instrument (Clifford, Barbarin et al. 2005), which produced a single score focused on the quality of teaching and interactions in the classroom environment.

Teacher Behavior Rating Scale

The TBRS has been used to evaluate the early literacy and language qualities of classrooms in numerous studies. It was developed with attention to the research literature about the classroom-learning opportunities and materials that contribute to children's early literacy skills. The TBRS has measured changes in the early literacy environment of the classroom over time in response to intervention and has related changes in the early literacy environment to growth in children's performance on well-accepted measures of early literacy skills (Landry, Swank, Smith, Assel, and Gunnewig, 2006).

The TBRS has been updated and modified over the last several years. Most recently, for the Preschool Curriculum Evaluation Research (PCER) project, the items were revised so that they would separately measure the frequency of a behavior (or quantity of materials) and the quality of the behavior (or of materials). Examination of the data for that evaluation indicated that the internal consistency remained high for the subscales (ranging from .69 to .97 in one evaluation, .63 to .93 in the other). Investigation of the PCER data also indicated that the correlations between quantity and quality assessments were fairly high, .72 to .97, and the coefficient alphas for the combined quality and quantity measures were also high, .82 to .95.

For the ERF national evaluation, the TBRS was further revised to allow four rather than three response categories for each item.[86] Accordingly, the version of TBRS used in ERF has not yet been used in any study with published findings. A different version of the TBRS was used in the Preschool Curriculum Evaluation Research (PCER) program, a multi-site efficacy evaluation of 14 preschool curricula being conducted by the Institute of Education Sciences. The TBRS version used in PCER is closest to the one used in ERF, but PCER used only the subscales that specifically measure the language, early literacy, and early-math aspects of the environment. Several subscales that measure the general quality of the classroom environment were not included in the PCER evaluation: teacher sensitivity, classroom community, quality and organization of activity centers, lesson plans, portfolios, dynamic assessments, and team teaching; however, these subscales were included in the version of TBRS used for ERF.

For the National Evaluation of ERF, inter-rater reliability was computed for the TBRS scales with a sample of 13 teachers who were observed independently by two different raters during fall 2004 data collection (see Table C.1). These coefficients are generally consistent with those

[86] For quantity items, the PCER version used *rarely/sometimes/often* as response categories, while the ERF version used *none/rarely/sometimes/often*; for quality items, the PCER version used *low/average/high*, while the ERF version used low/medium-low/medium-high/high.

obtained in the PCER evaluation. Thus, the reliability of the overall score and the subscales is generally acceptable for use to examine differences between groups.

Table C.1. ERF TBRS inter-rater reliability (n = 13 pairs)

Scale	Rxx
Book-Reading Behaviors	0.81928
Oral Language Use	0.88874
Phonological Awareness Activity	0.75595
Print and Letter Knowledge	0.87498
Written Expression	0.77145
Portfolios	1.00000
Dynamic Assessment	0.79377
General Teaching Behavior	0.82672
Classroom Community	0.74585
Teacher Sensitivity	0.88436
Lesson Plans	0.92370
Quality and Organization of Activity Centers	0.91801
Team Teaching Ability	0.98193
Math Concepts	0.89627
Total Score	0.92867

The validity of the TBRS has been established by showing significantly greater positive change in all dimensions measured by the TBRS for teachers receiving language and literacy interventions, compared to teachers who did not receive similar interventions (Landry, Swank, Smith, Assel, and Gunnewig; 2006) and in several other ongoing studies.

For the ERF evaluation, we formed subscales by first averaging quantity and quality items and then averaging across the composite items. As was true for the PCER evaluation, data from the ERF evaluation indicate that the correlations between quantity and quality items are high, .66 to .98 (see Table C.2). In the cases where the subscales were formed averaging quantity and quality, one cannot perfectly disentangle quantity from quality in the interpretation of middle-range scores. However, for subscales with very high item correlations (for example, .90 and above), the individual quantity and quality scores are very similar to the combined score.

Table C.2. Teacher behavior rating scale: correlations between quantity and quality items

Items	Correlation	Items	Correlation
General Quality of the Preschool Classroom			
Teacher Sensitivity		**Quality of Team Teaching**	
Item 1	.69	Item 1	Quality only
Item 2	.77	Item 2	.86
Item 3	.86	Item 3	.92
Item 4	.81	Item 4	Quality only
Average	.86	Item 5	Quality only
		Average	.87
Classroom Community		**Quality and Organization of**	
Item 1	.80	**Activity Centers**	
Item 2	.89	Item 1	.81
Item 3	Quality only	Item 2	Quality only
Item 4	Quality only	Item 3	Quality only
Item 5	.86	Item 4	Quality only
Average	.84	Item 5	Quality only
		Item 6	Quality only
		Item 7	.91
		Average	.81
Lesson Planning			
Item 1	.91		
Item 2	.87		
Item 3	.91		
Average	.93		

Table C.2. Teacher behavior rating scale: correlations between quantity and quality items—*Continued*

Items	Correlation	Items	Correlation
Classroom Language and Early Literacy Environment			
Oral Language Use by Lead Teacher		**Book-Reading Practices**	
Item 1	.66	Item 1	Quantity only
Item 2	.88	Item 2	.92
Item 3	.91	Item 3	.95
Item 4	.89	Item 4	.94
Item 5	.89	Item 5	.89
Item 6	.80	Item 6	.90
Item 7	.81	Item 7	.92
Average	.93	Item 8	.94
		Average	.95
Written Expression		**Child Portfolios**	
Item 1	.90	Item 1	Quantity only
Item 2	.81	Item 2	Quantity only
Item 3	.77		
Average	.98		
Print and Letter Knowledge		**Dynamic Assessment**	
Item 1	.86	Item 1	Quantity only
Item 2	.89	Item 2	Quantity only
Item 3	.92	Item 3	Quantity only
Item 4	.85		
Item 5	.86	**Math Concepts**	
Item 6	.88	Item 1	.85
Average	.93	Item 2	.84

NOTE: Some items have only a quality or only a quantity item but not both.
SOURCE: Correlations estimated from ERF classroom observation data.

In most cases, the original TBRS subscales were used for the ERF evaluation (see Table C.3). However, four of the TBRS subscales were modified to make greater use of the information available from the classroom observations:

- The ***Team Teaching Ability*** scale contains two items that measure the frequency and quality of the assistant teacher's language use in the classroom. These items provide an additional dimension to the overall helpfulness of the assistant teacher in the classroom. Moreover, in conjunction with the Oral Language Use scale, which measures the frequency and quality of the lead teacher's language use, these items provide a comprehensive view of the language stimulation provided by both adults in the classroom.

- The ***Phonological Awareness Activity*** scale contains indicators of whether specific phonological awareness activities were observed (for example, rhyming or syllable segmenting and blending), the number of classroom situations in which these activities were observed, and the quality of those activities, measured by children's engagement. The score of the Phonological Awareness Activity quantity subscale is

the average of *one variable* that captures the number of different classroom situations where these activities are observed (for example, circle time and mealtime) *and another variable* that captures the number and complexity of phonological awareness activities that were observed (thus, a higher score for sentence segmentation than for rhyming, and a higher score for doing 3 activities than 1). For the ERF evaluation, we replaced this subscale with a simple count of the number of phonological awareness activities observed because it is a more understandable measure of the frequency of these activities. The Phonological Awareness Activities quality subscale is typically formed by averaging the quality items that are observed. We followed this rule in forming the quality subscale for the ERF evaluation.

- The ***Print and Letter Knowledge*** scale contains 6 items that measure both teaching and the classroom environment. We divided this scale into subscales that measure teaching separately from the classroom environment so that progress in each area could be monitored.

- The ***Written Expression*** scale contains 3 items that measure both teaching and the classroom environment. We divided this scale into subscales that measure teaching and the classroom environment separately so that progress in each area could be monitored.

Internal consistency reliability coefficients for the original TBRS subscales and the subscales used for the ERF evaluation are provided in Table C.3.

Table C.3. Teacher behavior rating scale: original subscales and subscales used for ERF evaluation

Original Subscales		Subscales Used for ERF Evaluation	
Subscales and Items	Internal Consistency Reliability	Subscales and Items	Internal Consistency Reliability
General Quality of the Preschool Classroom			
Teacher Sensitivity	.89	**Teacher Sensitivity**	.89
1. Uses encouragement and positive feedback that provides child-or children-specific information regarding what they are doing well.		1. Uses encouragement and positive feedback that provides child- or children-specific information regarding what they are doing well.	
2. Uses *sensitivity behaviors* when responding to children's signals and needs (responds promptly and sensitively to children's verbal and nonverbal signals, values children's interests and needs (gets on child's eye level).		2. Uses *sensitivity behaviors* when responding to children's signals and needs (responds promptly and sensitively to children's verbal and nonverbal signals, values children's interests and needs (gets on child's eye level).	
3. Provides guidance that encourages children to regulate their behavior in learning and problem-solving situations vs. teacher "solving the problem" (includes all behavior, not just problem behaviors, e.g., "I don't know how; "I can't").		3. Provides guidance that encourages children to regulate their behavior in learning and problem-solving situations vs. teacher "solving the problem" (includes all behavior, not just problem behaviors, e.g., "I don't know how; "I can't").	
4. Engages children in literacy, language, or math activities using varied and playful techniques that make cognitive activities engaging (e.g., songs, books, games) *apart from the book read.*		4. Engages children in literacy, language, or math activities using varied and playful techniques that make cognitive activities engaging (e.g., songs, books, games) *apart from the book read.*	

Table C.3. Teacher behavior rating scale: original subscales and subscales used for ERF evaluation—*Continued*

Original Subscales		Subscales Used for ERF Evaluation	
Subscales and Items	Internal Consistency Reliability	Subscales and Items	Internal Consistency Reliability
Team Teaching Ability	.94	**Quality of Team Teaching**	.94
1. Teacher and assistant work together so that small groups of children receive ongoing instruction in center activities, small group activities, and read-alouds.		1. Teacher and assistant work together so that small groups of children receive ongoing instruction in center activities, small group activities, and read-alouds.	
2. During *small group work,* assistant scaffolds children's language, asks open-ended questions, and encourages conversation.		2. During *small group work,* assistant scaffolds children's language, asks open-ended questions, and encourages conversation.	
3. Assistant moves around classroom, scaffolding children's language, asking open-ended questions, and encouraging conversation (look for consistency *throughout the observation period).*		3. Assistant moves around classroom, scaffolding children's language, asking open-ended questions, and encouraging conversation (look for consistency *throughout the observation period).*	
4. The assistant supports the lead teacher by participating in classroom regulation of her own initiative (consider that appropriate classroom regulation should not cause disruption or interrupt teaching).		4. The assistant supports the lead teacher by participating in classroom regulation of her own initiative (consider that appropriate classroom regulation should not cause disruption or interrupt teaching).	
5. Overall, the assistant's presence in the classroom improves the teaching environment (e.g., positive presence for the children, engages the children, shows interest and enjoyment, and is prompt/sensitive in responding to children's needs).		5. Overall, the assistant's presence in the classroom improves the teaching environment (e.g., positive presence for the children, engages the children, shows interest and enjoyment, and is prompt/sensitive in responding to children's needs).	
		Oral Language Use by Assistant Teacher	.94
		2. During *small group work* assistant scaffolds children's language, asks open-ended questions, and encourages conversation.	
		3. Assistant moves around classroom scaffolding children's language, asking open-ended questions, and encouraging conversation (look for consistency *throughout the observation period).*	

Table C.3. Teacher behavior rating scale: original subscales and subscales used for ERF evaluation—*Continued*

Original Subscales		Subscales Used for ERF Evaluation	
Subscales and Items	Internal Consistency Reliability	Subscales and Items	Internal Consistency Reliability
Classroom Community	.86	**Classroom Community**	.86
1. Orients children for the expectations in the classroom through established rules and routines (e.g., what is expected and where things belong).		1. Orients children for the expectations in the classroom through established rules and routines (e.g., what is expected and where things belong).	
2. Encourages children to work with the teacher in establishing rules and routines (e.g., children may each have jobs in the class that are clearly defined as evidenced in charts with pictures or icons, and children can be seen practicing and doing these jobs around the classroom).		2. Encourages children to work with the teacher in establishing rules and routines (e.g., children may each have jobs in the class that are clearly defined as evidenced in charts with pictures or icons, and children can be seen practicing and doing these jobs around the classroom).	
3. Arranges and organizes space in a way that allows children to move around the room safely and facilitates interaction with their peers.		3. Arranges and organizes space in a way that allows children to move around the room safely and facilitates interaction with their peers.	
4. Designs a layout for the classroom so children are able to get materials on their own (e.g., shelves are clearly labeled, learning materials are at eye level, provides personal place for each child's belonging that is clearly labeled).		4. Designs a layout for the classroom so children are able to get materials on their own (e.g., shelves are clearly labeled, learning materials are at eye level, provides personal place for each child's belonging that is clearly labeled).	
5. Values children by displaying their work around the room (more children's work is seen displayed around the room than store-bought materials e.g., family or child photos, hand prints, children's books in library). Classroom should feel as if it is the children's place rather than the teacher's room.		5. Values children by displaying their work around the room (more children's work is seen displayed around the room than store-bought materials, e.g., family or child photos, hand prints, children's books in library). Classroom should feel as if it is the children's place rather than the teacher's room.	

Table C.3. Teacher behavior rating scale: original subscales and subscales used for ERF evaluation—*Continued*

Original Subscales		Subscales Used for ERF Evaluation	
Subscales and Items	Internal Consistency Reliability	Subscales and Items	Internal Consistency Reliability
Quality and Organization of Activity Centers	.90	**Quality and Organization of Activity Centers**	.90
1. Number of centers that cover critical learning activities and learning objectives *linked to the theme* including library & listening, construction (blocks), writer's corner, math/science, pretend & learn (dramatic play), creativity station (art), and ABC center.		1. Number of centers that cover critical learning activities and learning objectives *linked to the theme* including library & listening, construction (blocks), writer's corner, math/science, pretend & learn (dramatic play), creativity station (art), and ABC center.	
2. Materials, activities, and objectives follow the current theme and are linked to learning goals (exciting and obvious theme rates high; look for appropriate rotation of seasonal items, refreshing of materials).		2. Materials, activities, and objectives follow the current theme and are linked to learning goals (exciting and obvious theme rates high; look for appropriate rotation of seasonal items, refreshing of materials).	
3. Prepares children with specific information and discussion as to how to move children into centers, change centers, and use center materials for learning.		3. Prepares children with specific information and discussion as to how to move children into centers, change centers, and use center materials for learning.	
4. Centers have clear boundaries that allow children to easily distinguish between learning centers (e.g., centers are clearly labeled and are enclosed based on learning area; appropriate use of short shelves, bookcases, furniture, to create distinct areas of learning).		4. Centers have clear boundaries that allow children to easily distinguish between learning centers (e.g., centers are clearly labeled and are enclosed based on learning area; appropriate use of short shelves, bookcases, furniture, to create distinct areas of learning).	
5. Centers provide space that encourages child interaction (e.g., low shelves provide visibility; enough room in centers for multiple children; centers with noisy activities are located in an area separate from activities that require less noise).		5. Centers provide space that encourages child interaction (e.g., low shelves provide visibility; enough room in centers for multiple children; centers with noisy activities are located in an area separate from activities that require less noise).	
6. Tables in classrooms are arranged in a manner that supports centers (e.g., tables are arranged in close proximity to a center encouraging children to bring materials from a specific center to the table, rather than several tables being arranged in a row in the center of the room).		6. Tables in classrooms are arranged in a manner that supports centers (e.g., tables are arranged in close proximity to a center encouraging children to bring materials from a specific center to the table, rather than several tables being arranged in a row in the center of the room).	
7. Teacher effectively models use and care of center materials.		7. Teacher effectively models use and care of center materials.	

Table C.3. Teacher behavior rating scale: original subscales and subscales used for ERF evaluation—*Continued*

Original Subscales		Subscales Used for ERF Evaluation	
Subscales and Items	Internal Consistency Reliability	Subscales and Items	Internal Consistency Reliability
Lesson Plans	.93	**Lesson Planning**	.93
1. Shows strong thematic connection in written lesson plans (detailed information that ties theme-related materials and activities to learning objectives).		1. Shows strong thematic connection in written lesson plans (detailed information that ties theme-related materials and activities to learning objectives).	
2. Teacher is observed implementing and following through with activities from the lesson plan.		2. Teacher is observed implementing and following through with activities from the lesson plan.	
3. Lesson plan objectives are evident, based on materials located in centers and around the room (e.g., materials in dramatic play center reflect current theme, theme-related books are present, children's work related to theme or lesson plan is displayed around the room).		3. Lesson plan objectives are evident, based on materials located in centers and around the room (e.g., materials in dramatic play center reflect current theme, theme-related books are present, children's work related to theme or lesson plan is displayed around the room).	
Classroom Language and Early Literacy Environment			
Oral Language Use	.93	**Oral Language Use by Lead Teacher**	.93
1. Speaks clearly and uses grammatically correct sentences.		1. Speaks clearly and uses grammatically correct sentences.	
2. Models for children how to express their ideas in complete sentences.		2. Models for children how to express their ideas in complete sentences.	
3. Uses "scaffolding" language (nouns, descriptors, action words, linking concepts).		3. Uses "scaffolding" language (nouns, descriptors, action words, linking concepts).	
4. Uses "thinking" questions (open-ended, "why", "how") or comments to support children's thinking or activity or interest.		4. Uses "thinking" questions (open-ended, "why", "how") or comments to support children's thinking or activity or interest.	
5. Relates previously learned words and concepts to activity.		5. Relates previously learned words and concepts to activity.	
6. Encourages children's use of language throughout the observation period irrespective of type of activities.		6. Teacher encourages children's use of language throughout the observation period irrespective of type of activities.	
7. Engages children in conversations that involves child and teacher taking multiple turns (e.g., 3–5 turns).		7. Engages children in conversations that involves child and teacher taking multiple turns (e.g., 3–5 turns).	

Table C.3. Teacher behavior rating scale: original subscales and subscales used for ERF evaluation—*Continued*

Original Subscales		Subscales Used for ERF Evaluation	
Subscales and Items	Internal Consistency Reliability	Subscales and Items	Internal Consistency Reliability
Book-Reading Behaviors	.92	**Book-Reading Practices**	.92
1. Introduces the book through display of book cover, reading of title, author, and illustrator (no chart or display cards required).		1. Introduces the book through display of book cover, reading of title, author, and illustrator (no chart/display cards required).	
2. Encourages some discussion about one or more of these book features (refers to cover of book, title, author, or illustrator).		2. Encourages some discussion about one or more of these book features (refers to cover of book, title, author, or illustrator).	
3. Vocabulary words are discussed when preparing to read and/or reading books aloud (charts and displays are not required).		3. Vocabulary words are discussed when preparing to read and/or reading books aloud (charts and displays are not required).	
4. Vocabulary words are combined with pictures or objects when preparing to read or when reading books aloud.		4. Vocabulary words are combined with pictures or objects when preparing to read or when reading books aloud.	
5. Facial expressions and voice are used to capture children's attention by using different tones for characters (book) or modulating voice to emphasize words/facts (fiction or nonfiction).		5. Facial expressions and voice are used to capture children's attention by using different tones for characters (book) or modulating voice to emphasize words/facts (fiction or nonfiction).	
6. Teacher paces the reading to fit the type of book being read and to allow for children to be involved through comments and questions.		6. Teacher paces the reading to fit the type of book being read and to allow for children to be involved through comments and questions.	
7. Asks open-ended questions (e.g., "what if", "where have you seen", "how would") to encourage discussion of facts in the book (nonfiction), details, plot and/or characters (fiction), or topic and/or rhyming (poetry).		7. Asks open ended questions (e.g., "what if", "where have you seen", "how would") to encourage discussion of facts in the book (nonfiction), details, plot and/or characters (fiction), or topic and/or rhyming (poetry).	
8. Takes time to involve children in activities or discussions that extend books that are read (e.g., story maps/sequences, props, retells).		8. Takes time to involve children in activities or discussions that extend books that are read (e.g., story maps/sequences, props, retells).	

Table C.3. Teacher behavior rating scale: original subscales and subscales used for ERF evaluation—*Continued*

Original Subscales		Subscales Used for ERF Evaluation	
Subscales and Items	Internal Consistency Reliability	Subscales and Items	Internal Consistency Reliability
Phonological Awareness Activity	n.a.	**Number of Phonological Awareness Activities Observed**	n.a.
1. Number of different learning situations settings in which the teacher integrates phonological activities. Include: centers / book read / circle time / transitions / small group.		Number of activities listed in Item 2 that were observed.	
2. Provides phonological awareness activities from the developmental continuum: • Listening • Sentence segmenting • Syllable blending and segmenting • Onset-rime blending and segmenting • Rhyming • Phoneme blending, segmenting, and manipulation • Alliteration		**Quality of Phonological Awareness Activities** Average quality of child engagement in the activities observed in #2.	n.a.
3. Quality of child engagement in each of the phonological awareness activities in #2.			

Table C.3. Teacher behavior rating scale: original subscales and subscales used for ERF evaluation—*Continued*

Original Subscales		Subscales Used for ERF Evaluation	
Subscales and Items	Internal Consistency Reliability	Subscales and Items	Internal Consistency Reliability
Print and Letter Knowledge	.87	**Print and Letter Knowledge Learning Opportunities**	.90
1. Engages children in name and theme- or topic-related activities that promote letter/word knowledge, help learn to associate names of letters with shapes, and begin to make sound/letter matches.		1. Engages children in name and theme- or topic-related activities that promote letter/word knowledge, help learn to associate names of letters with shapes, and begin to make sound/letter matches.	
2. Provides opportunities for children to compare and discuss same/different letters, names, and words.		2. Provides opportunities for children to compare and discuss same/different letters, names, and words.	
3. Discusses concepts about print (text contains letters, words, sentences; reading progresses left to right, top to bottom, etc.).		3. Discusses concepts about print (text contains letters, words, sentences; reading progresses left to right, top to bottom, etc.).	
		Classroom Print Environment	.80
4. Provides a literacy connection (books/book extenders) in all centers that are linked to theme/topic.		4. Provides a literacy connection (books/book extenders) in all centers that are linked to theme/topic.	
5. The environment and centers have theme- or topic-related print (e.g., labels, charts, posters).		5. The environment and centers have theme- or topic-related print (e.g., labels, charts, posters).	
6. A letter wall is used as an interactive teaching tools (e.g., visible at eye level, has space for 3 to 5 words per letter and pictures for all words, consecutive ordering, organizes games and activities involving letter wall).		6. A letter wall is used as an interactive teaching tools (e.g., visible at eye level, has space for 3 to 5 words per letter and pictures for all words, consecutive ordering, organizes games and activities involving letter wall).	
Written Expression	.90	**Written Expression Learning Opportunities**	n.a.
1. Lead teacher models writing (e.g., experience charts, morning message, news of the day, child dictations).		1. Lead teacher models writing (e.g., experience charts, morning message, news of the day, child dictations).	
		Opportunities and Materials for Writing	.89
2. Provides children with a variety of opportunities and materials to engage in writing (e.g., journals, response to literature, etc.).		2. Provides children with a variety of opportunities and materials to engage in writing (e.g., journals, response to literature, etc.).	
3. Number of centers (excluding the writing center) where writing materials are provided.		3. Number of centers (excluding the writing center) where writing materials are provided.	

Table C.3. Teacher behavior rating scale: original subscales and subscales used for ERF evaluation—*Continued*

Original Subscales		Subscales Used for ERF Evaluation	
Subscales and Items	Internal Consistency Reliability	Subscales and Items	Internal Consistency Reliability
Portfolios	.66	**Child Portfolios**	.66
1. Dated documentation in portfolios of children's developmental progress with children's art work, samples of written expression, journals, children's notes, or children's dictations. Randomly select 5 portfolios and rate on basis of whether there are samples of work in 0–3 different areas contained in 0–5 different portfolios. Higher score for more types of work in larger number of sampled portfolios.		1. Dated documentation in portfolios of children's developmental progress with children's art work, samples of written expression, journals, children's notes, or children's dictations. Randomly select 5 portfolios and rate on basis of whether there are samples of work in 0–3 different areas contained in 0–5 different portfolios. Higher score for more types of work in larger number of sampled portfolios.	
2. Portfolios contain teacher-written observations in the form of anecdotal notes. In 5 randomly selected portfolios, rate on basis of whether there are 0–2 teacher notes in 0–4 portfolios. Higher score for more notes in more portfolios.		2. Portfolios contain teacher-written observations in the form of anecdotal notes. In 5 randomly selected portfolios, rate on basis of whether there are 0–2 teacher notes in 0–4 portfolios. Higher score for more notes in more portfolios.	
Dynamic Assessment	.72	**Dynamic Assessment**	.72
1. Dated documentation of children's developmental progress across a range of emergent literacy areas through the use of cognitive checklists/assessments. Portfolio items must be dated within the last 30 days.		1. Dated documentation of children's developmental progress across a range of emergent literacy areas through the use of cognitive checklists/assessments. Portfolio items must be dated within the last 30 days.	
2. Do you plan for instruction on basis of the individualized assessments/checklists?		2. Do you plan for instruction on basis of the individualized assessments/checklists?	
3. If yes, how do you use them? Planning small-group work / Grouping children by ability / Planning center activities / Developing IEP / Other application.		3. If yes, how do you use them? Planning small-group work / Grouping children by ability / Planning center activities / Developing IEP / Other application.	
Math Concepts	.86	**Subscale not analyzed separately in body of ERF Report, but items were included in TBRS Total Score**	n.a.
1. Involves children in organized *hands-on* activities that support one or more of the math strand concepts (i.e., counting, 1:1 correspondence, sorting, patterning, graphing). Shapes and measurements).			
2. Incorporates math in daily routines (e.g., attendance, lunch count, voting, graphics).			

Source: Internal consistency reliability estimated from ERF Classroom Observation data.

Early Childhood Environment Rating Scale—Revised (ECERS-R)

We used the ECERS-R (Harms, Clifford, and Cryer 1998) to evaluate classroom quality. The ECERS-R is a global measure of the preschool classroom environment, so its primary focus is not classroom language and literacy. The instrument has 43 items, of which, 36 are used to determine the overall quality score. Each item is scored on a scale of 1 to 7, in which, 1 = poor, 3 = minimally acceptable, 5 = good, and 7 = excellent. Reports of inter-rater agreement indicate that 86.1 percent of the time raters agree within one point on the scale, and no items had inter-rater agreement that was less than 70 percent (Harms, Clifford, and Cryer 1998).

We used the following subset of 11 items, which compose the subscale "Teaching and Interactions" (Clifford, Barbarin, Chang, Early, Bryant, Howes, Burchinal, and Painta 2005), to measure the quality of the preschool classroom environments in both ERF and non-ERF sites:

- Greeting/Departing
- Encouraging Children to Communicate
- Using Language to Develop Reasoning Skills
- Informal Use of Language
- Supervision of Gross Motor Activities
- General Supervision of Children
- Discipline
- Staff-Child Interactions
- Interactions among Children
- Free Play
- Group Time

These items were identified through factor analysis (Clifford, et al. 2005) and had coefficients of at least .4. This factor is similar to one constructed in previous studies (Clifford, Burchinal, Harms, Rossbach, and Lera 1996; Rossbach, Clifford, and Harms 1991).

Evidence for the validity of the ECERS-R has been demonstrated by comparing scores on the ECERS-R to other structural measures of classroom quality and child outcomes (Peisner-Feinberg and Burchinal 1997; Whitebook, Howes, and Phillips 1990). For the National Evaluation of ERF, we computed inter-rater reliability for the 11 ECERS items with a sample of 13 teachers who were observed independently by two different raters during fall 2004 data collection. The inter-rater reliability coefficient was .89, which is similar to the .915 reported in the ECERS manual (Harms, Clifford, and Cryer 1998).

Psychometric Information for Key Constructed Variables

Table C.4 presents key psychometric data for the constructed variables created for the impact analysis. The table is organized by measurement domain. We include the sample size, possible range of values for each variable, the actual range found in the ERF sample, the sample mean, standard deviation, and the internal consistency reliability (coefficient alpha). The psychometric data are presented for the full sample, that is, combining the program and control groups.

Table C.4. Descriptive information for composite variables constructed from classroom observations and child assessments, for the full sample

Measure	Sample size	Possible range		Range in ERF sample		Mean	Standard deviation	Internal consistency reliability[α]
		Minimum	Maximum	Minimum	Maximum			
Child Language Development								
EOWPVT: Expressive Vocabulary, raw score	1,624	0	99	1	99	39.22	15.35	NA
EOWPVT: Expressive Vocabulary, standard score	1,624	53	147	53	147	83.56	17.36	NA
PLS-IV: Auditory Comprehension, raw score	1,650	1	62	1	62	51.44	7.44	NA
PLS-IV: Auditory Comprehension, standard score	1,650	50	135	50	135	92.09	15.28	NA
Child Early Literacy Skills								
Pre-CTOPPP: Print Awareness, raw score	1,648	0	36	1	36	21.28	10.03	NA
Pre-CTOPPP: Print Awareness, standard score	1,656	58	144	62	144	100.02	16.96	NA
Pre-CTOPPP: Elision, raw score	1,646	0	18	0	18	9.21	4.19	NA
Child Social-Emotional Development								
SCBE: Social competence	1,574	0	50	7	50	31.87	9.54	.93
SCBE: Anxiety-withdrawal	1,574	0	50	0	41	10.78	6.68	.85
SCBE: Anger-aggression	1,574	0	50	0	48	9.56	8.60	.94
General Quality of the Preschool Classroom								
ECERS-R: Teaching and Interactions	169	1.00	7.00	1.64	7.00	5.78	1.03	.85
TBRS: Teacher Sensitivity	169	0.50	4.00	0.50	4.00	2.86	0.68	.89
TBRS: Quality of Team Teaching	151	0.71	4.00	0.80	4.00	2.68	0.96	.94
TBRS: Classroom Community	169	0.63	4.00	0.90	4.00	2.96	0.67	.86
TBRS: Quality and Organization of Activity Centers	167	0.78	4.00	0.86	4.00	2.64	0.78	.90
TBRS: Lesson Planning	168	0.50	4.00	0.50	4.00	2.71	1.01	.93
Language, Early Literacy, and Assessment Practices								
TBRS: Oral Language Use by Lead Teacher	169	0.50	4.00	0.50	4.00	2.61	0.77	.93
TBRS: Oral Language Use by Assistant Teacher	151	0.50	4.00	0.50	4.00	2.27	1.18	.94
TBRS: Book-Reading Practices	164	0.50	4.00	0.56	3.94	2.07	0.85	.92
TBRS: Number of Different Phonological Awareness Activities Observed	169	0.00	7.00	0.00	7.00	1.55	1.63	NA
TBRS: Quality of Phonological Awareness Activities	169	0.00	4.00	0.00	4.00	1.58	1.23	.80
TBRS: Print and Letter Knowledge Learning Opportunities	168	0.50	4.00	0.50	4.00	1.64	1.00	.90
TBRS: Classroom Print Environment	169	0.50	4.00	0.50	4.00	1.96	0.86	.80
TBRS: Written Expression Learning Opportunities	169	0.50	4.00	0.50	4.00	1.40	1.15	NA
TBRS: Opportunities and Materials for Writing	169	0.50	4.00	0.50	4.00	2.00	0.87	.84
TBRS: Child Portfolios	158	1.00	5.00	1.00	5.00	2.43	1.36	.66
TBRS: Dynamic Assessment	169	0.67	4.33	0.67	4.33	2.54	1.11	.72
TBRS: Total Score	167	0.62	4.00	0.94	3.89	2.34	0.65	.94

[a]Reliability was estimated by using Cronbach's coefficient alpha formula.
SOURCE: Child assessments and interviewer observations conducted in the fall and spring.

Appendix D. Supplementary Tables on the Impacts of ERF on Teachers and Classroom Environments

This appendix presents the impacts of ERF on teachers and classrooms in the fall of 2004. In addition, to supplement the information about the classroom language and literacy environment, this appendix presents the impacts of ERF on the proportion of classrooms in which specific phonological awareness activities were observed.

Impacts of ERF in Fall 2004

ERF had statistically significant impacts on some aspects of the classroom literacy environment in the fall, including the classroom print environment, writing materials, phonological awareness activities, and modeling writing for children.

Impacts on Teachers' Qualifications

We find no evidence of an impact of ERF on years of teaching experience, measured as either teaching preschool generally or teaching at the current school or center.

ERF had a positive impact on teachers' professional development in fall 2004 (see Table D.1). The program increased the number of hours of professional development that focused on language and early literacy topics by 48 hours (6 days) over the 12 months preceding the survey. ERF also had a positive impact on the mode of training. A higher proportion of ERF teachers than teachers in unfunded programs reported receiving professional development on language or literacy topics and on curriculum topics through mentoring or tutoring, the more intensive approach recommended by ERF. A larger proportion of ERF teachers than teachers in unfunded programs also reported receiving workshop training on language and literacy topics. Nearly half of all ERF teachers reported receiving mentoring in the previous year on language and literacy topics (using regression-adjusted percentages), and nearly 70 percent had attended workshop training.

Table D.1. ERF impacts on teachers' experience, training, and earnings, fall 2004

Domain/Outcome (range)	Unadjusted means		Regression-adjusted means				
	Funded	Unfunded	Funded	Unfunded	Estimated impact[a]	Effect size[b]	P-value of impact
Teaching Experience							
Years at current school or center (0–30)	5.56	6.47	5.89	5.22	0.68	0.12	0.684
Years at any preschool (0–36)	9.40	10.00	9.69	8.81	0.87	0.11	0.623
Professional Development ✓							
Professional development focusing on early language and literacy topics:							
Hours (1–160)	61.79	23.62	63.60	15.31	48.29	1.12	0.000*
Received professional development through:							
Mentoring or tutoring (%)	40.00	11.24	48.81	10.77	38.04	0.87	0.002*
Workshops (%)	54.44	49.44	68.82	37.55	31.27	0.63	0.003*
Professional development focusing on curriculum:							
Hours (0–160)	44.50	25.64	44.26	28.27	15.99	0.36	0.331
Received professional development through:							
Mentoring or tutoring (%)	34.44	11.24	35.66	10.31	25.35	0.62	0.045*
Workshops (%)	36.67	38.20	43.73	37.69	6.04	0.12	0.730
Number of teachers			90	89			
Number of sites			28	34			
Earnings							
Teachers' hourly earnings (6.05–60.00)	20.55	14.57	20.49	14.66	5.83	0.58	0.248
Number of preschools			41	41			
Number of sites			22	26			

*p-value (of adjusted difference in means) < 0.05; two-tailed test.
✓ Impact on domain is positive and statistically significant after adjustments for multiple comparisons (see Appendix A).
[a]All estimates except those for earnings were obtained from a regression model of the outcome variable on an indicator variable of ERF grant receipt; grant application score; and teacher's education, age, and an indicator variable of nonwhite, using SAS's PROC MIXED procedure for continuous outcome measures and SUDAAN logit for binary outcome measures. Missing values of covariates were mean-imputed by site. For earnings, the regression model included only an indicator variable of ERF grant receipt and grant application score without any teacher demographic controls.
[b]The effect size was calculated by dividing the estimated impact by the standard deviation of the outcome measure (that is, the impact expressed as a percentage of the standard deviation).
NOTE: All figures were estimated by using sample weights to account for the sample and survey designs. Standard errors of the impact estimates account for design effects from unequal weighting of the data and clustering at site level.
SOURCE: ERF fall teacher surveys and director surveys.

We found no statistically significant differences in the hourly earnings of teachers in ERF programs relative to those in unfunded programs in the fall. The impact estimate is small and not statistically distinguishable from zero.

Impacts on General Quality of Preschool Classrooms

ERF had no impacts on the domains reflecting the general quality of preschool classrooms in the fall. Impact estimates for measures of the quality of teacher-child interactions, the organization of the classroom environment, planning, and adequacy of supervision are small and do not meet the .05 threshold for statistical significance (see Table D.2).

Table D.2. ERF impacts on classroom outcomes: general quality of the preschool classroom, fall 2004

	Unadjusted means		Regression-adjusted means				
Domain/Outcome (range)	Funded	Unfunded	Funded	Unfunded	Estimated impact[a]	Effect size[b]	P-value of impact
Quality of Teacher-child Interactions							
Teaching and interactions (ECERS-R) (1.64–7.00)	5.70	5.42	5.74	5.30	0.43	0.41	0.213
Teacher sensitivity (TBRS) (0.75–4.00)	3.11	2.99	3.01	3.10	-0.09	-0.13	0.720
Quality of team teaching (TBRS) (0.80–4.00)	2.97	2.73	2.91	2.82	0.09	0.10	0.812
Organization of the Environment							
Classroom community (TBRS) (1.30–4.00)	3.18	2.96	3.14	2.96	0.18	0.28	0.475
Quality and organization of activity centers (TBRS) (0.86–4.00)	3.12	2.70	3.13	2.60	0.53	0.70	0.058
Planning							
Lesson planning (TBRS) (0.50–4.00)	3.06	2.50	2.94	2.69	0.24	0.25	0.487
Total Teacher Behavior Rating Scale							
Total TBRS score (1.00–3.67)	2.71	2.33	2.71	2.31	0.40	0.62	0.095
Adequacy of Supervision							
Child-staff ratio (1.83–18.00)	7.38	7.65	7.37	7.64	-0.27	-0.10	0.778
Number of classrooms			78	91			
Number of sites			28	37			

*p-value (of adjusted difference in means) < 0.05; two-tailed test.

[a]All estimates were obtained from a regression model of the outcome variable on an indicator variable of ERF grant receipt; grant application score; and teacher's education, age, and an indicator variable of nonwhite, using SAS's PROC MIXED procedure. Missing values of covariates were mean-imputed by site.

[b]The effect size was calculated by dividing the estimated impact by the standard deviation of the outcome measure (that is, the impact expressed as a percentage of the standard deviation).

NOTE: All figures were estimated using sample weights to account for the sample and survey designs. Standard errors of the impact estimates account for design effects due to unequal weighting of the data and clustering at site level.

SOURCE: ERF fall classroom observations.

Impacts on Classroom Support for Language and Early Literacy

In fall 2004, when the ERF program was expected to be fully implemented in the 2003 cohort of preschool classrooms, ERF had statistically significant, large impacts on important domains of the classroom early literacy environment, including phonological awareness activities, print and letter knowledge, and writing (see Table D.3). We found no discernable impacts on the oral language environment, book reading, or child screening and progress assessments in the fall.

Table D.3. ERF impacts on classroom outcomes: language, early literacy, and assessment practices, fall 2004

	Unadjusted Means		Regression-Adjusted Means				
Domain/Outcome (range)	Funded	Unfunded	Funded	Unfunded	Estimated impact[a]	Effect size[b]	P-value of impact
Oral Language Environment							
Oral Language Use by Lead Teacher (0.86–4.00)	2.99	2.83	2.98	2.83	0.14	0.20	0.583
Oral Language Use by Assistant Teacher (0.50–4.00)	2.66	2.40	2.58	2.49	0.09	0.08	0.843
Book Reading							
Number of Book Reading Sessions Observed (0–4)	1.65	1.48	1.66	1.34	0.32	0.28	0.449
Book Reading Practices (0.56–3.94)	2.34	2.01	2.38	1.85	0.53	0.62	0.098
Phonological Awareness Activities ✓							
Number of Different Phonological Awareness Activities Observed (0–7)	2.37	1.70	2.57	1.41	1.15	0.78	0.046*
Quality of Phonological Awareness Activities (0–4.00)	2.07	1.86	2.04	1.94	0.10	0.09	0.798
Print and Letter Knowledge ✓							
Learning Opportunities (0.50–4.00)	2.26	1.78	2.21	1.81	0.40	0.40	0.275
Classroom Print Environment (0.50–4.00)	2.38	1.89	2.40	1.77	0.62	0.76	0.025*
Written Expression ✓							
Learning Opportunities (0.50–4.00)	2.06	1.38	2.16	1.08	1.08	0.86	0.012*
Opportunities and Materials for Writing (0.50–4.00)	2.53	1.77	2.58	1.54	1.04	1.18	0.002*
Child Screening and Progress Assessments							
Child Portfolios (1.00–5.00)	2.79	2.21	2.96	1.96	1.00	0.67	0.077
Dynamic Assessment (0.67–4.33)	2.84	2.28	2.72	2.43	0.28	0.24	0.517
Number of classrooms			78	89			
Number of sites			28	37			

*p-value (of adjusted difference in means) < 0.05; two-tailed test.
✓ Impact on domain is positive and statistically significant after adjustments for multiple comparisons (see Appendix A).
[a]All estimates were obtained from a regression model of the outcome variable on an indicator variable of ERF grant receipt; grant application score; and teacher's education, age, and an indicator variable of nonwhite, using SAS's PROC MIXED procedure. Missing values of covariates were mean-imputed by site.
[b]The effect size was calculated by dividing the estimated impact by the standard deviation of the outcome measure (that is, the impact expressed as a percentage of the standard deviation).
NOTE: All figures were estimated using sample weights to account for the sample and survey designs. Standard errors of the impact estimates account for design effects due to unequal weighting of the data and clustering at site level.
SOURCE: ERF fall classroom observations.

ERF had a positive impact on phonological awareness activities. In particular, ERF increased the number of different phonological awareness activities observed during the 3-hour classroom observation. The number of phonological awareness activities increased by 1.15 on average, relative to what would have been observed in the absence of ERF. However, ERF had no statistically significant impact on the quality of these activities (measured by the level of child engagement).

ERF had positive impacts on print and letter knowledge and written expression. ERF classrooms scored higher on the availability of print in the classroom—labels, books, and letters displayed with pictures—compared with unfunded classrooms. ERF had no impact on print- and letter-knowledge learning opportunities. ERF classrooms provided significantly more writing materials and opportunities for writing compared with unfunded classrooms and significantly increased the written-expression learning opportunities relative to what we would expect in the absence of the program.

ERF had no impacts on either the oral language environment of the classroom or book reading in the fall. Estimated impacts on measures in these domains for the most part are small and do not reach the .05 threshold for statistical significance. ERF also had no statistically significant impacts on child screening and progress assessment, as measured by the recency, extensiveness, and completeness of child portfolios and dynamic assessments.

Impacts on Phonological Awareness Activities, Fall 2004 and Spring 2005

Table D.4 shows the impacts of ERF on the proportion of classrooms in the fall in which each phonological-awareness activity was observed. Because the outcome variables are binary and in some cases, the activity was observed infrequently, the impact estimates are unstable (see Appendix A for further discussion). Listening was observed in 43 percent of the funded classrooms and 57 percent of the unfunded classrooms (using regression-adjusted percentages). Rhyming, another common activity, was observed in 51 percent of funded classrooms and 44 percent of unfunded classrooms. Alliteration was observed more often in funded than unfunded classrooms; the impact of ERF was 41 percentage points. Sentence segmenting was also observed more often in funded than in unfunded classrooms. We would expect the percentage of classrooms conducting each activity to be less than 100 because many different activities could be occurring in each classroom during the 3-hour visit.

Table D.4. ERF impacts on phonological awareness activities, fall 2004

Domain/Outcome (range)	Unadjusted Means		Regression-Adjusted Means				
	Funded	Unfunded	Funded	Unfunded	Estimated Impact[a]	Effect Size[b]	P-value of Impact
Phonological Awareness Activities							
Listening (teacher draws attention to environmental sounds) (0–1)	52.6	53.8	43.08	57.21	-14.14	-0.28	0.433
Rhyming (identifying words with the same ending sound) (0–1)	47.4	44.0	51.27	44.82	6.45	0.13	0.697
Alliteration (note initial sounds in words (lazy lizard lounging)) (0–1)	43.6	27.5	61.97	20.94	41.03	0.86	0.001*
Onset-rime blending and segmenting (working with words that share sounds and varying the first letter or sound—c-at, b-at) (0–1)	25.6	14.3	43.51	10.96	32.54	0.80	0.066
Phoneme blending, segmenting and manipulation (*isolate sounds in words and replace with other sounds*) (0–1)	25.6	7.7	38.52	6.24	32.27	0.87	0.059
Sentence segmenting (clapping for each word in a sentence, deleting words in a sentence, using word cards) (0–1)	25.6	4.4	41.37	2.56	38.81	1.15	0.023*
Syllable blending and segmenting (clapping for each syllable, deleting syllables) (0–1)	16.7	18.7	12.32	23.95	11.63	-0.31	0.353
Number of Classrooms			78	91			
Number of Sites			28	37			

*p-value (of adjusted difference in means) < 0.05; two-tailed test.
[a]All estimates were obtained from a logit regression model of the outcome variable on an indicator variable of ERF grant receipt; grant application score; and teacher's education, age, and an indicator variable of nonwhite, using SUDAAN. Missing values of covariates were mean-imputed by site.
[b]The effect size was calculated by dividing the estimated impact by the standard deviation of the outcome measure (that is, the impact expressed as a percentage of the standard deviation).
NOTE: All figures were estimated using sample weights to account for the sample and survey designs. Standard errors of the impact estimates account for design effects due to unequal weighting of the data and clustering at site level.
SOURCE: ERF fall classroom observations.

Table D.5 shows the impacts of ERF on the proportion of classrooms in the spring in which each phonological awareness activity was observed. Listening was observed in 45 percent of funded and 28 percent of unfunded classrooms. Rhyming, another common activity, was observed more often in ERF classrooms than in unfunded classrooms. Other more challenging phonological awareness activities, such as blending and segmenting words, syllables, initial sounds, and phonemes, were observed in 37 percent or fewer ERF classrooms (using regression-adjusted percentages).

Table D.5. ERF impacts on phonological awareness activities, spring 2005

Domain/Outcome (range)	Unadjusted Means		Regression-Adjusted Means				
	Funded	Unfunded	Funded	Unfunded	Estimated Impact[a]	Effect Size[b]	P-value of Impact
Phonological Awareness Activities							
Listening (teacher draws attention to environmental sounds) (0–1)	39.7	33.0	45.37	28.46	16.91	0.35	0.295
Rhyming (identifying words with the same ending sound) (0–1)	64.1	28.6	70.39	26.16	44.23	0.89	0.002*
Alliteration (note initial sounds in words (lazy lizard lounging)) (0–1)	32.1	14.3	32.58	14.79	17.79	0.43	0.283
Onset-rime blending and segmenting (working with words that share sounds and varying the first letter or sound—c-at, b-at) (0–1)	26.9	4.4	32.69	3.77	28.93	0.81	0.101
Phoneme blending, segmenting and manipulation (*isolate sounds in words and replace with other sounds*) (0–1)	26.9	4.4	37.36	3.78	33.59	0.94	0.071
Sentence segmenting (clapping for each word in a sentence, deleting words in a sentence, using word cards) (0–1)	12.8	3.3	31.01	1.72	29.30	1.15	0.254
Syllable blending and segmenting (clapping for each syllable, deleting syllables) (0–1)	21.8	7.7	23.98	6.90	17.08	0.50	0.190
Number of Classrooms			78	91			
Number of Sites			28	37			

*p-value (of adjusted difference in means) < 0.05; two-tailed test.
[a]All estimates were obtained from a logit regression model of the outcome variable on an indicator variable of ERF grant receipt; grant application score; teacher's education, age, and an indicator variable of nonwhite, using SUDAAN. Missing values of covariates were mean-imputed by site.
[b]The effect size was calculated by dividing the estimated impact by the standard deviation of the outcome measure (that is, the impact expressed as a percentage of the standard deviation).
NOTE: All figures were estimated using sample weights to account for the sample and survey designs. Standard errors of the impact estimates account for design effects due to unequal weighting of the data and clustering at site level.
SOURCE: ERF spring classroom observations.

Appendix E. ERF Impacts on Teacher and Classroom Outcomes; Subgroups Analyses

This appendix presents subgroup impact estimates for the spring for a subset of the teacher and classroom outcomes examined in Chapter 6 on overall impacts. The outcomes chosen for this appendix include several key professional development outcomes, approximately half of the outcomes in the area of general preschool quality, and all of the outcomes in the language, early literacy, and assessment areas. In general, the pattern of positive impacts on professional development, the general quality of the preschool classroom, and the classroom language, early literacy, and assessment practices persists across most subgroups we examined, although the estimates are, in many cases, not statistically significant at conventional levels.

To better understand overall estimates of impacts on teacher training and classroom practice, we estimated impacts for subgroups of classrooms defined by specific, policy-relevant characteristics of teachers, classrooms, or preschools. The analysis examines impacts for teachers with and without a bachelor's degree; teachers with five or more years of teaching experience and teachers with fewer years of experience; whether the preschool received Head Start funding; and whether the preschool offered full-time or part-time classes. Although several limitations of the subgroup analysis (discussed in the following sections) mean that we should not draw conclusions about the program's effectiveness for the groups considered, nevertheless, the patterns of impacts across subgroups can provide indications of whether practices were changed across a broad spectrum of teachers classrooms and preschools or, alternatively, whether some subgroups appear to benefit to a greater or lesser degree.

One limitation of the subgroup analysis is that the study does not have the statistical power to estimate subgroup impacts with a high level of precision. A second limitation is that many of the subgroup characteristics that we examined are interrelated, and the analysis cannot control for correlations among these characteristics. For example, preschools with funding from Head Start may be more likely to have teachers without a bachelor's degree relative to preschools without Head Start funding. Also, when examining subgroups defined by teacher, classroom, or preschool characteristics that may not vary greatly within a site, we may not be comparing similar sets of sites. For example, only 34 of the 65 sites in the full sample have a selected classroom in which the teacher has less than a bachelor's degree. Only 27 of the 65 sites in the study included one or more preschools that receive Head Start funding. It is likely that teacher-education levels or Head Start funding is correlated with other aspects of the sites, preschools, and classrooms. Therefore, any differences in impacts that we observe across the subgroups may be related to aspects of these sites as well as to the subgroup differences being examined.

We note that when analyzing impacts for several subgroups, we are likely, simply by chance, to find impacts that are statistically significant at the 0.05 level in about 5 percent of the estimates. Therefore, in the discussion that follows, we focus primarily on *differences* in impacts across subgroups level (for instance, teachers with and without a bachelor's degree).

In the following text, we present estimated effect sizes and p-values from t-tests that measure the statistical significance of the subgroup impacts. We also present p-values from F-tests that measure the difference in impacts across subgroup levels (for example, across teachers with and without a bachelor's degree).

Impacts by Teacher Education

Current policy debates regarding quality standards for early-childhood programs focus on whether preschool teachers must have skills and knowledge that can best be provided by a bachelor's degree rather than by intensive professional development and teaching experience. Twenty-five state preschool programs require teachers to have a bachelor's degree, matching the minimum qualifications for teachers of kindergarten through grade 12 (Barnett et al. 2006). Policymakers are currently debating whether to require that 50 percent of Head Start teachers have a bachelor's degree by 2011. Given the level of policy interest in the relative skills of teachers with and without a bachelor's degree, we examined whether the impacts of ERF vary by whether the teacher has a bachelor's degree (or more education) or not.

We find that the impacts of ERF for teachers with and without a bachelor's degree are similar for many outcomes, and the difference between the impacts for teachers with and without a bachelor's degree is not statistically significant for any of the outcomes examined (see Table E.1). We estimate large, statistically significant impacts of ERF on all domains of language, early literacy, and assessment practices for teachers with a bachelor's degree and large but not statistically significant impacts on all domains except book reading for teachers without a bachelor's degree. Impact estimates for teachers without a bachelor's degree are imprecise because of the small sample size of this group.

Table E.1. ERF impacts on selected teacher and classroom outcomes, by level of teacher education, spring 2005

Outcome (range)	Teachers with a bachelor's degree		Teachers without a bachelor's degree		P-value of difference in impacts between subgroups
	Effect size[a]	P-value	Effect size[a]	P-value	
Teachers' Experience and Training					
Professional Development Hours—Early Language and Literacy	1.04	0.009 *	1.03	0.033 *	0.227
Received professional development through mentoring/tutoring	0.99	0.003 *	0.86	0.145	0.548
Professional Development Hours—Curriculum	0.45	0.254	0.52	0.248	0.167
Received professional development through mentoring/tutoring	0.74	0.055	1.29	0.052	0.337
Number of Teachers	125		65		
Number of Sites	55		36		
General Quality of the Preschool Classroom					
ECERS-R Teaching and Interactions	1.29	0.001 *	1.22	0.032 *	0.764
TBRS					
Teacher sensitivity	1.45	0.001 *	0.54	0.368	0.991
Classroom community	1.19	0.005 *	1.01	0.065	0.220
Total score	1.57	0.000 *	1.05	0.067	0.537
Language, Early Literacy, and Assessment Practices					
Oral Language Environment					
Oral Language Use by Lead Teacher (0.86–4.00)	1.27	0.005 *	1.04	0.070	0.128
Oral Language Use by Assistant Teacher (0.50–4.00)	0.91	0.050 *	0.98	0.148	0.693
Book Reading					
Number of Book Reading Sessions Observed (0–4)	0.33	0.478	-0.20	0.767	0.937
Book Reading Practices (0.56–3.94)	1.30	0.005 *	0.35	0.572	0.597
Phonological Awareness Activities					
Number of Different Phonological Awareness Activities Observed (0–7)	1.03	0.023 *	1.37	0.012 *	0.649
Quality of Phonological Awareness Activities (0–4.00)	0.58	0.232	1.05	0.047 *	0.108
Print and Letter Knowledge					
Learning Opportunities (0.50–4.00)	0.94	0.042 *	0.40	0.548	0.860
Classroom Print Environment (0.50–4.00)	0.79	0.069	0.80	0.166	0.316
Written Expression					
Learning Opportunities (0.50–4.00)	1.06	0.008 *	0.89	0.154	0.931
Opportunities and Materials for Writing (0.50–4.00)	1.60	0.000 *	0.86	0.143	0.805
Child Screening and Progress Assessments					
Child Portfolios (1.00–5.00)	0.78	0.124	0.97	0.118	0.903
Dynamic Assessment 0.67–4.33)	1.06	0.034 *	0.19	0.753	0.855
Number of Classrooms	99		49		
Number of Sites	52		34		

Notes from Table E.1

*p-value (of effect size or difference between subgroups) < 0.05, two-tailed test.
[a] All estimates were obtained from a regression model of the outcome variable on an indicator variable of ERF grant receipt; grant application score; teacher's age, and an indicator variable of nonwhite, using SAS's PROC MIXED procedure. Missing values of covariates were mean-imputed by site. The effect size was calculated by dividing the estimated impact by the standard deviation of the outcome measure (that is, the impact expressed as a percentage of the standard deviation).
NOTE: All figures were estimated using sample weights to account for the sample and survey designs. Standard errors of the impact estimates account for design effects due to unequal weighting of the data and clustering at site level.
SOURCE: ERF spring director and teacher surveys and classroom observations.

Impacts by Teacher Experience

Teachers with more teaching experience are likely to have more practical knowledge than less experienced teachers have about classroom management and how children learn, but their formal education is usually less recent. Preschools often employ a mix of new and experienced teachers; therefore, to address whether the kinds of skills emphasized by ERF make a greater difference for new teachers or for more experienced teachers, we examined the impacts of ERF according to whether the teacher had five or more years' preschool teaching experience or less than five years of experience.

We find that the impacts of ERF on professional development, measures of the general quality of the preschool classroom, and classroom language, literacy, and assessment practices are positive and typically large for both groups. The differences between the impacts for teachers with less than 5 years' experience and those with more experience are not statistically significant except for oral language use by the assistant teacher (see Table E.2). ERF improved the quality of oral language use by assistant teachers to a greater extent in classrooms with new teachers than in classrooms with experienced teachers, although ERF impacts on this outcome are positive for both groups.

Table E.2. ERF impacts on selected teacher and classroom outcomes, by years of teacher experience, spring 2005

Outcome (range)	Teachers with less than 5 years' preschool experience		Teachers with 5 or more years' preschool experience		P-value of difference in impacts between subgroups
	Effect size[a]	P-value	Effect size[a]	P-value	
Teachers' Experience and Training					
Professional Development Hours—Early Language and Literacy	1.02	0.031 *	1.15	0.003 *	0.769
Received professional development through mentoring/tutoring	0.28	0.350	1.19	0.000 *	0.273
Professional Development Hours—Curriculum	0.18	0.740	0.47	0.225	0.167
Received professional development through mentoring/tutoring	0.76	0.085	0.85	0.027 *	0.254
Number of Teachers	62		128		
Number of Sites	43		61		
General Quality of the Preschool Classroom					
ECERS-R Teaching and Interactions	1.49	0.003 *	0.98	0.018 *	0.988
TBRS					
Teacher sensitivity	0.80	0.153	0.99	0.025 *	0.887
Classroom community	1.35	0.015 *	1.15	0.008 *	0.369
Total score	0.99	0.039 *	1.59	0.000 *	0.944
Language, Early Literacy, and Assessment Practices					
Oral Language Environment					
Oral Language Use by Lead Teacher (0.86–4.00)	0.98	0.082	1.29	0.002 *	0.290
Oral Language Use by Assistant Teacher (0.50–4.00)	1.60	0.004 *	0.54	0.259	0.007*
Book Reading					
Number of Book Reading Sessions Observed (0–4)	0.34	0.571	0.00	0.994	0.235
Book Reading Practices (0.56–3.94)	0.78	0.130	1.12	0.005 *	0.315
Phonological Awareness Activities					
Number of Different Phonological Awareness Activities Observed (0–7)	1.05	0.028 *	1.15	0.015 *	0.298
Book Reading Practices (0.56–3.94)	0.93	0.071	0.65	0.131	0.374
Print and Letter Knowledge					
Learning Opportunities (0.50–4.00)	0.43	0.402	1.09	0.018 *	0.532
Classroom Print Environment (0.50–4.00)	0.54	0.336	0.95	0.025	0.359
Written Expression					
Learning Opportunities (0.50–4.00)	0.56	0.224	1.22	0.005 *	0.996
Opportunities and Materials for Writing (0.50–4.00)	1.29	0.018 *	1.68	0.000 *	0.415
Child Screening and Progress Assessments					
Child Portfolios (1.00–5.00)	0.84	0.108	0.83	0.055	0.215
Dynamic Assessment (0.67–4.33)	0.15	0.786	0.65	0.137	0.992
Number of Classrooms	51		118		
Number of Sites	36		60		

Notes from Table E.2

*p-value (of effect size or difference between subgroups) < 0.05, two-tailed test.
[a] All estimates were obtained from a regression model of the outcome variable on an indicator variable of ERF grant receipt; grant application score; teacher's education, age, and an indicator variable of nonwhite, using SAS's PROC MIXED procedure. Missing values of covariates were mean-imputed by site. The effect size was calculated by dividing the estimated impact by the standard deviation of the outcome measure (that is, the impact expressed as a percentage of the standard deviation).
NOTE: All figures were estimated using sample weights to account for the sample and survey designs. Standard errors of the impact estimates account for design effects due to unequal weighting of the data and clustering at site level.
SOURCE: ERF spring director and teacher surveys and classroom observations.

Impacts by Whether a Preschool Received Head Start Funding

Preschools in the study sample received funding from many different sources, including private fees, local agencies, state-education and early-childhood programs, and federal programs such as Even Start and Head Start. The largest source of federal funding for preschools is the Head Start program. The Head Start program has placed a strong emphasis over the past decade on improving the quality of programs, particularly through increasing the educational requirements of teachers and strengthening language and early literacy instruction in the classroom. These recent policy emphases led us to examine whether ERF introduced into a Head Start program had a greater or lesser effect on classroom practice than ERF in preschools not funded by Head Start. We compared the impacts of ERF in preschools that received Head Start funding with preschools that received no Head Start funding.

We found that the impacts of ERF on teacher and classroom outcomes for those with and without Head Start funding are, for the most part, positive and similar in magnitude. The difference between the impacts for classrooms with and without Head Start funding is not statistically significant for any outcome except one (see Table E.3). The one statistically significant difference that emerges between the Head Start and non-Head Start classrooms is the impact of ERF on written-expression learning opportunities. ERF had no impact on written-expression learning opportunities in classrooms with Head Start funding but had an impact (effect size = 1.54; p-value = 0.000) on this outcome in classrooms without Head Start funding.

Table E.3. ERF impacts on selected teacher and classroom outcomes, by Head Start funding or not, spring 2005

Outcome (range)	Preschools with Head Start funding			Preschools without Head Start funding			P-value of difference in impacts between subgroups
	Effect size[a]	P-value		Effect size[a]	P-value		
Teachers' Experience and Training							
Professional Development Hours—Early Language and Literacy	1.06	0.074		1.06	0.011	*	0.855
Received professional development through mentoring/tutoring	1.04	0.000	*	0.52	0.164		0.352
Professional Development Hours—Curriculum	0.37	0.492		0.56	0.178		0.610
Received professional development through mentoring/tutoring	1.06	0.000	*	0.44	0.314		0.147
Number of Teachers	63			100			
Number of Sites	27			47			
General Quality of the Preschool Classroom							
ECERS-R Teaching and Interactions	0.50	0.377		1.46	0.000	*	0.247
TBRS							
Teacher sensitivity	1.03	0.072		1.03	0.029	*	0.914
Classroom community	0.94	0.079		1.23	0.006	*	0.304
Total score	1.63	0.001	*	1.36	0.002	*	1.000
Language, Early Literacy, and Assessment Practices							
Oral Language Environment							
Oral Language Use by Lead Teacher (0.86–4.00)	1.19	0.033	*	1.10	0.007	*	0.758
Oral Language Use by Assistant Teacher (0.50–4.00)	1.32	0.029	*	0.73	0.161		0.135
Book Reading							
Number of Book Reading Sessions Observed (0–4)	–0.32	0.599		0.38	0.435		0.217
Book Reading Practices (0.56–3.94)	0.50	0.378		1.20	0.008	*	0.112
Phonological Awareness Activities							
Number of Different Phonological Awareness Activities Observed (0–7)	1.38	0.032	*	1.35	0.003	*	0.537
Quality of Phonological Awareness Activities (0–4.00)	1.52	0.005	*	0.72	0.094		0.078
Print and Letter Knowledge							
Learning Opportunities (0.50–4.00)	0.53	0.453		1.04	0.012	*	0.122
Classroom Print Environment (0.50–4.00)	0.94	0.167		0.80	0.087		0.444
Written Expression							
Learning Opportunities (0.50–4.00)	–0.02	0.980		1.54	0.000	*	0.000*
Opportunities and Materials for Writing (0.50–4.00)	1.39	0.003	*	1.46	0.001	*	0.765
Child Screening and Progress Assessments							
Child Portfolios (1.00–5.00)	0.52	0.403		1.26	0.011	*	0.398
Dynamic Assessment 0.67–4.33)	1.08	0.108		0.44	0.383		0.257
Number of Classrooms	44			96			
Number of Sites	25			49			

Notes from Table E.3

*p-value (of effect size or difference between subgroups) < 0.05, two-tailed test.
[a] All estimates were obtained from a regression model of the outcome variable on an indicator variable of ERF grant receipt; grant application score; teacher's education, age, and an indicator variable of nonwhite, using SAS's PROC MIXED procedure. Missing values of covariates were mean-imputed by site. The effect size was calculated by dividing the estimated impact by the standard deviation of the outcome measure (that is, the impact expressed as a percentage of the standard deviation).
NOTE: All figures were estimated using sample weights to account for the sample and survey designs. Standard errors of the impact estimates account for design effects due to unequal weighting of data and clustering at site level.
SOURCE: ERF spring director and teacher surveys and classroom observations.

Impacts by Whether Preschool Is Full-Time or Part-Time

ERF might have greater impacts on children's language and early literacy skills if children experience the program for a longer preschool day. However, the effects of a longer ERF day on children could be reduced if ERF is not implemented well in full-time programs compared to part-time programs. To inform the analysis of ERF impacts on children by program intensity, we examined the impacts of ERF on professional development and classroom-learning environments by whether the classroom meets full-time (defined as serving children six or more hours per day for five days per week) or part-time (defined as serving children fewer than six hours per day or fewer than 5 days per week).

We found that ERF had differential impacts on professional development and on a measure of organization of the classroom environment in full-time compared to part-time programs (see Table E.4). ERF had a positive impact on hours of professional development focusing on curriculum among teachers in full-time programs but had a negative impact on this outcome among teachers in part-time programs. Neither impact estimate is statistically significant at conventional levels, but the difference in the impact estimates is statistically significant ($p = 0.036$). ERF had a positive impact on the proportion of teachers in both groups who received professional development on language and literacy topics through mentoring, but the impact on teachers in part-time programs is larger and statistically significant. ERF had a large, positive impact on classroom community in full-time classrooms but had no statistically discernable impact on this outcome for part-time classrooms.

Although this pattern of differential ERF impacts on professional development and classroom organization is mixed, the pattern of ERF impacts on other measures of general classroom quality, the classroom language and literacy environment, and child assessment practices is more consistent for the two groups. The impacts of ERF on teacher-child interactions, oral language use, book reading, phonological awareness, print and letter knowledge, written expression, and child assessments are consistently positive, and most are of similar magnitude for full-time and part-time classrooms.

Table E.4. ERF impacts on selected teacher and classroom outcomes, by whether preschool is full day or part day, spring 2005

Outcome (range)	Full-day (6 or more hours) Effect size[a]	Full-day (6 or more hours) P-value	Part-day (fewer than 6 hours) Effect size[a]	Part-day (fewer than 6 hours) P-value	P-value of difference in impacts between subgroups
Teachers' Experience and Training					
Professional Development Hours—Early Language and Literacy	1.18	0.002 *	0.43	0.434	0.661
Received professional development through mentoring/tutoring	0.57	0.174	1.45	0.000 *	0.007*
Professional Development Hours—Curriculum	0.60	0.111	-0.55	0.320	0.036*
Received professional development through mentoring/tutoring	0.75	0.057	0.95	0.106	0.223
Number of Teachers	116		63		
Number of Sites	49		28		
General Quality of the Preschool Classroom					
ECERS-R Teaching and Interactions	0.92	0.015 *	1.56	0.033 *	0.815
TBRS					
Teacher sensitivity	0.87	0.038 *	1.02	0.203	0.772
Classroom community	1.33	0.002 *	–0.32	0.679	0.023*
Total score	1.38	0.001 *	1.09	0.113	0.572
Language, Early Literacy, and Assessment Practices					
Oral Language Environment					
Oral Language Use by Lead Teacher (0.86–4.00)	1.15	0.005 *	0.52	0.487	0.101
Oral Language Use by Assistant Teacher (0.50–4.00)	0.88	0.060	0.31	0.683	0.142
Book Reading					
Number of Book Reading Sessions Observed (0–4)	0.06	0.884	0.85	0.291	0.691
Book Reading Practices (0.56–3.94)	0.86	0.036 *	0.99	0.244	0.370
Phonological Awareness Activities					
Number of Different Phonological Awareness Activities Observed (0–7)	1.09	0.010 *	0.87	0.254	0.224
Quality of Phonological Awareness Activities (0–4.00)	0.95	0.015 *	0.29	0.718	0.303
Print and Letter Knowledge					
Learning Opportunities (0.50–4.00)	0.70	0.100	1.09	0.115	0.855
Classroom Print Environment (0.50–4.00)	0.86	0.049 *	0.60	0.419	0.344
Written Expression					
Learning Opportunities (0.50–4.00)	0.92	0.022 *	1.82	0.016	0.882
Opportunities and Materials for Writing (0.50–4.00)	1.52	0.000 *	1.94	0.009 *	0.857
Child Screening and Progress Assessments					

Outcome (range)	Full-day (6 or more hours) Effect size[a]	P-value	Part-day (fewer than 6 hours) Effect size[a]	P-value	P-value of difference in impacts between subgroups
Child Portfolios (1.00–5.00)	1.01	0.031 *	1.46	0.038	0.538
Dynamic Assessment 0.67–4.33)	0.50	0.296	0.05	0.951	0.736
Number of classrooms	107		48		
Number of sites	50		28		

Notes from Table E.4

*p-value (of effect size or difference between subgroups) < 0.05, two-tailed test.
[a] All estimates were obtained from a regression model of the outcome variable on an indicator variable of ERF grant receipt; grant application score; teacher's education, age, and an indicator variable of nonwhite, using SAS's PROC MIXED procedure. Missing values of covariates were mean-imputed by site. The effect size was calculated by dividing the estimated impact by the standard deviation of the outcome measure (that is, the impact expressed as a percentage of the standard deviation).
NOTE: All figures were estimated using sample weights to account for the sample and survey designs. Standard errors of the impact estimates account for design effects due to unequal weighting of the data and clustering at site level.
SOURCE: ERF spring director and teacher surveys and classroom observations.

Appendix F. ERF Impacts on Child Outcomes; Subgroups Analyses

The ERF evaluation estimated impacts for several subgroups defined by characteristics of children and the preschools they attended. The characteristics were gender, race and ethnicity, primary language spoken at home, parental education, whether the preschool received Head Start funding, and whether the preschool offered full-time or part-time classes. One limitation of this line of analysis is that the study does not have the statistical power to estimate subgroup impacts with a high level of precision. A related limitation is that we cannot control for the co-occurrence of characteristics considered. For example, one ethnic group may have a preponderance of the children whose primary language is other than English, and we cannot disentangle the effects of the two characteristics. Notwithstanding these important limitations, an examination of the patterns of impacts across subgroups informs our understanding of ERF's effects. For example, it indicates whether particular subgroups might derive greater or lesser benefits from ERF or, alternatively, whether all groups appear to benefit to a similar extent.

While the subgroup analysis can provide a general sense of the pattern and magnitude of impacts for the different population subgroups of interest, it is important to keep in mind that when analyzing impacts for several different subgroups, we are likely to find impacts that are statistically significant at the 5 percent level in about 5 percent of the estimates, simply by chance alone. Therefore, in the discussion that follows, we focus primarily on *differences* in impacts across subgroup levels (for instance, boys versus girls, or jointly across black, white, and Hispanic children), and where relevant, we discuss the robustness of these differences in impacts to adjustments for the multiple outcomes being examined across subgroups.

In general, there are very few significant differences in outcomes across subgroup levels, and the pattern of impacts observed for the full sample generally persists across most of the subgroups that we examined. In the print and letter knowledge domain, effect sizes of impacts on print awareness generally range from .30 to .55 for most subgroups, although these estimates are generally not statistically significant. In the phonological awareness domain, impact estimates on the Elision subtest are generally less than .20 and are not statistically significant for any of the subgroups examined. In the oral language domain, effect sizes of estimated impacts on the expressive vocabulary subtest are generally less than .15 and are not statistically significant for most subgroups. Estimated impacts on the auditory comprehension subtest are between .20 and .50 across almost all population subgroups that we examined, but these estimates are typically not statistically significant at conventional levels. Impact estimates for social-emotional skills are also generally not statistically significant.

In this appendix, we present estimated effect sizes and p-values from t-tests that gauge the statistical significance of the subgroup impacts. We also present p-values from F-tests that gauge the difference in impacts across subgroup levels.

Impacts by Gender

Research on early childhood development typically considers the possibility of variations by gender, and gender differences in verbal ability are widely believed to exist, although a careful review of the extensive empirical evidence suggests little or no verbal advantage for girls (Hyde and Linn 1988). We examined ERF impacts by gender to evaluate whether the program is more effective for boys or for girls. We find that the impacts for boys and girls are similar, and the difference between the impacts for boys and girls is not statistically significant for any of the

outcomes examined (see Table F.1). We estimate effect sizes of .33 standard deviation on the print-awareness standard score for both boys and girls. Estimated impacts in the phonological awareness domain are small and not statistically significant for either group. In the oral language domain, the estimated effect size on auditory comprehension standard scores is between .26 and .28 for both groups but not statistically significant, and the estimated impact on expressive vocabulary is small and not statistically significant. For both boys and girls, estimated impacts on the social-emotional subscales are also generally small and not statistically significant.

Table F.1. ERF impacts on child outcomes by gender

	Boys		Girls		P-value of difference in impacts between subgroups
Outcome (range)	Effect Size[a]	P-value	Effect Size[a]	P-value	
Language and Literacy Skills					
Print and Letter Knowledge					
Print awareness, raw score (0–36)	0.36	0.115	0.50	0.019*	0.283
Print awareness, standard score (58–144)	0.33	0.076	0.33	0.104	0.816
Phonological Awareness					
Elision, raw score (0–18)	0.02	0.910	0.17	0.264	0.236
Oral Language					
Expressive vocabulary, raw score (0–99)	–0.10	0.541	0.08	0.581	0.212
Expressive vocabulary, standard score (53–147)	–0.11	0.534	0.13	0.395	0.140
Auditory comprehension, raw score (1–62)	0.26	0.138	0.29	0.130	0.458
Auditory comprehension, standard score (50–135)	0.26	0.156	0.28	0.101	0.599
Number of students	841		807		
Number of sites	65		65		
Social Competence and Behavior Evaluation (Scales Range from 0 to 50)					
Social competence	0.06	0.776	0.15	0.525	0.995
Anxiety-withdrawal	0.08	0.675	–0.05	0.806	0.564
Anger-aggression	–0.34	0.083	–0.16	0.445	0.560
Number of students	833		813		
Number of sites	65		65		

*p-value (of effect size or difference between subgroups) < 0.05, two-tailed test.
[a]All estimates were obtained from a regression model of the outcome variable on an indicator variable of ERF grant receipt; grant application score; and an indicator variable of nonwhite, using SAS's PROC MIXED procedure. Language and literacy skill models also control for indicator variables of fall assessment taken in Spanish and missing fall assessment data and age at spring assessment. SCBE models also control for an indicator variable of missing fall SCBE data and age at spring SCBE observation. Missing values of covariates are mean-imputed by site and gender.
[b]The effect size was calculated by dividing the estimated impact by the standard deviation of the outcome measure (that is, the impact expressed as a percentage of the standard deviation).
NOTE: All figures were estimated by using sample weights to account for the sample and survey designs. Standard errors of the impact estimates account for design effects due to unequal weighting of the data and clustering at site and classroom level.
SOURCE: ERF spring child assessments and SCBE evaluations.

Impacts by Race and Ethnicity

Because differential impacts across racial and ethnic groups might indicate that the program is narrowing or increasing racial and ethnic gaps in children's early-language and literacy skills, we examined whether ERF impacts vary by race and ethnicity. We find that patterns of impacts are similar across Hispanic, white non-Hispanic, and black non-Hispanic children (see Table F.2).[87]

Estimated impacts in the print- and letter-knowledge domain range from .36 to .59 for the three groups, and the difference in impacts across the three groups is not statistically significant. Estimated impacts in the phonological awareness domain tend to be small and are not statistically significant. In the oral-language domain, estimated impacts for auditory-comprehension standard scores are between .34 and .42 for all three groups but are not statistically significant, and estimated impacts for expressive vocabulary are small and not statistically significant. We find no statistically significant impacts on social-emotional outcomes for any of the racial and ethnic groups.

Impacts by Primary Language Spoken at Home

Groups of preschools applying for an ERF grant in 2003 were encouraged to serve English-language learners (ELLs), and accordingly, our sample of children in ERF preschools includes a significant proportion of children whose native language is not English. ELLs who are mastering basic English may have difficulty learning early literacy skills, and it is possible that ERF could be less effective for this group. Alternatively, an enhanced-language and early literacy environment may help ELLs make greater progress in expressive vocabulary and phonological awareness than children whose home language is English. To examine whether ERF impacts differed for ELLs versus others, we defined subgroups according to the parents' report of whether the primary language spoken to the child at home was English or some other language.

Patterns of results for the two groups are similar (see Table F.3). Estimated impacts in the print- and letter-knowledge domain range between .40 and .57 for both groups, and the difference in impacts across subgroup levels is not statistically significant. Estimated impacts in the phonological awareness domain are small and not statistically significant for either group. In the oral-language domain, the estimated effect size on auditory comprehension standard scores is between .33 and .49 for both groups but not statistically significant, and the estimated impact on expressive vocabulary is small and not statistically significant. For both groups, estimated impacts on the social-emotional subscales are in a favorable direction but are not statistically significant.

[87] Because not all sites contain black or Hispanic children, the set of sites included in the analysis differs slightly for each subgroup.

Table F.2. ERF impacts on child outcomes by race/ethnicity

Outcome (range)	Hispanic		White, non-Hispanic		Black, non-Hispanic		P-value of difference in impacts between subgroups
	Effect size[a]	P-value	Effect size[a]	P-value	Effect size[b]	P-value	
Language and Literacy Skills							
Print and Letter Knowledge							
Print awareness, raw score (0–36)	0.43	0.135	0.57	0.028*	0.49	0.069	0.703
Print awareness, standard score (58–144)	0.36	0.106	0.59	0.022*	0.37	0.146	0.944
Phonological Awareness							
Elision, raw score (0–18)	0.11	0.619	0.03	0.916	0.30	0.198	0.328
Oral Language							
Expressive vocabulary, raw score (0–99)	0.09	0.666	0.13	0.601	–0.02	0.934	0.744
Expressive vocabulary, standard score (53–147)	0.13	0.547	0.14	0.561	–0.03	0.917	0.693
Auditory comprehension, raw score (1–62)	0.32	0.213	0.36	0.123	0.24	0.346	0.558
Auditory comprehension, standard score (50–135)	0.34	0.165	0.42	0.102	0.33	0.240	0.894
Number of Students	679		423		467		
Number of Sites	54		56		52		
Social Competence and Behavior Evaluation (Scales range from 0 to 50)							
Social competence	0.34	0.227	0.24	0.339	–0.16	0.570	
Anxiety-withdrawal	–0.46	0.052	0.06	0.817	0.17	0.543	
Anger-aggression	–0.19	0.397	–0.32	0.239	–0.31	0.290	
Number of students	691		411		450		
Number of sites	53		55		50		

*p-value (of effect size or difference between subgroups) < 0.05, two-tailed test.
[a]All estimates were obtained from a regression model of the outcome variable on an indicator variable of ERF grant receipt; grant application score; and an indicator variables of female, using SAS's PROC MIXED procedure. Language and literacy skill models also control for indicator variables of fall assessment taken in Spanish and missing fall assessment data and age at spring assessment. SCBE models also control for an indicator variable of missing fall SCBE data and age at spring SCBE observation. Missing values of covariates are mean-imputed by site and gender.
[b]The effect size was calculated by dividing the estimated impact by the standard deviation of the outcome measure (that is, the impact expressed as a percentage of the standard deviation).
NOTE: All figures were estimated using sample weights to account for the sample and survey designs. Standard errors of the impact estimates account for design effects due to unequal weighting of the data and clustering at site and classroom level.
SOURCE: ERF spring child assessments and SCBE evaluations.

Table F.3. ERF impacts on child outcomes by primary language spoken to child at home

Outcome (range)	English		Other language		P-value of difference in impacts between subgroups
	Effect size[a]	P-value	Effect size[a]	P-value	
Language and Literacy Skills					
Print and Letter Knowledge					
Print awareness, raw score (0–36)	0.57	0.014*	0.40	0.154	0.462
Print awareness, standard score (58–144)	0.46	0.025*	0.55	0.040*	0.779
Phonological Awareness					
Elision, raw score (0–18)	0.09	0.584	0.06	0.763	0.967
Oral Language					
Expressive vocabulary, raw score (0–99)	–0.04	0.835	0.14	0.518	0.504
Expressive vocabulary, standard score (53–147)	–0.02	0.899	0.21	0.354	0.349
Auditory comprehension, raw score (1–62)	0.27	0.117	0.42	0.104	0.293
Auditory comprehension, standard score (50–135)	0.33	0.121	0.49	0.069	0.609
Number of students	785		498		
Number of sites	64		56		
Social Competence and Behavior Evaluation (Scales range from 0 to 50)					
Social competence	0.18	0.430	0.16	0.572	
Anxiety-withdrawal	0.01	0.980	–0.44	0.098	
Anger-aggression	–0.38	0.068	–0.24	0.302	
Number of students	763		502		
Number of sites	64		55		

*p-value (of effect size or difference between subgroups) < 0.05, two-tailed test.
[a]All estimates were obtained from a regression model of the outcome variable on an indicator variable of ERF grant receipt; grant application score; and an indicator variables of female, using SAS's PROC MIXED procedure. Language and literacy skill models also control for indicator variables of fall assessment taken in Spanish and missing fall assessment data and age at spring assessment. SCBE models also control for an indicator variable of missing fall SCBE data and age at spring SCBE observation. Missing values of covariates are mean-imputed by site and gender.
[b]The effect size was calculated by dividing the estimated impact by the standard deviation of the outcome measure (that is, the impact expressed as a percentage of the standard deviation).
NOTE: All figures were estimated using sample weights to account for the sample and survey designs. Standard errors of the impact estimates account for design effects due to unequal weighting of the data and clustering at site and classroom level.
SOURCE: ERF spring child assessments and SCBE evaluations.

Impacts by Parental Education

Parents' education is correlated with children's cognitive and language development (Brooks-Gunn, Berlin, and Fuligni 2000; NICHD Early Child Care Research Network 2001). To determine whether ERF impacts differed by parental education, we defined subgroups according to whether or not at least one of the child's parents had attended college. We find no significant differences in impacts across these subgroups (see Table F.4). General patterns of impacts are similar to those for the full sample for these two subgroups. We find effect sizes in the range of .37 and .44 in the print- and letter-knowledge domain for both groups, although estimated impacts are not statistically significant for either group. Estimated impacts in the phonological-awareness domain are small and not statistically significant for either group. In the oral-language domain, the estimated effect size on auditory comprehension standard scores is about .33 for both groups but not statistically significant, and the estimated impact on expressive vocabulary is small and not statistically significant. For both groups, estimated impacts on the social-emotional subscales are in a favorable direction but are not statistically significant.

Table F.4. ERF impacts on child outcomes by parental education

Outcome (range)	No college		College		P-value of difference in impacts between subgroups
	Effect size[a]	P-value	Effect size[a]	P-value	
Language and Literacy Skills					
Print and Letter Knowledge					
Print awareness, raw score (0–36)	0.37	0.133	0.44	0.106	0.645
Print awareness, standard score (58–144)	0.40	0.053	0.11	0.668	0.086
Phonological Awareness					
Elision, raw score (0–18)	0.02	0.887	0.16	0.494	0.886
Oral Language					
Expressive vocabulary, raw score (0–99)	–0.11	0.655	0.11	0.639	0.488
Expressive vocabulary, standard score (53–147)	–0.07	0.781	0.14	0.556	0.583
Auditory comprehension, raw score (1–62)	0.29	0.154	0.46	0.044*	0.526
Auditory comprehension, standard score (50–135)	0.34	0.118	0.33	0.192	0.622
Number of students	762		441		
Number of sites	65		65		
Social Competence and Behavior Evaluation (Scales range from 0 to 50)					
Social competence	0.11	0.625	0.40	0.166	
Anxiety-withdrawal	–0.07	0.760	–0.20	0.402	
Anger-aggression	–0.26	0.167	–0.67	0.011*	
Number of students	755		436		
Number of sites	65		63		

*p-value (of effect size or difference between subgroups) < 0.05, two-tailed test.
[a]All estimates were obtained from a regression model of the outcome variable on an indicator variable of ERF grant receipt; grant application score; and an indicator variables of female, using SAS's PROC MIXED procedure. Language and literacy skill models also control for indicator variables of fall assessment taken in Spanish and missing fall assessment data and age at spring assessment. SCBE models also control for an indicator variable of missing fall SCBE data and age at spring SCBE observation. Missing values of covariates are mean-imputed by site and gender.
[b]The effect size was calculated by dividing the estimated impact by the standard deviation of the outcome measure (that is, the impact expressed as a percentage of the standard deviation).
NOTE: All figures were estimated using sample weights to account for the sample and survey designs. Standard errors of the impact estimates account for design effects due to unequal weighting of the data and clustering at site and classroom level.
SOURCE: ERF spring child assessments and SCBE evaluations.

Impacts by Whether Preschool Received Head Start Funding

Preschools in our study received funding from a variety of sources, as discussed in Chapter 4. The largest source of federal funding to preschools is the Head Start program, which provided funding to at least 47 of the 152 preschools in our sample (funding source data are missing for 21 preschools). The Head Start program focuses on improving the quality of its preschool program by increasing educational requirements for teachers and training all Head Start teachers on techniques for improving children's language and early literacy skills. We examined whether ERF implemented in preschools with a Head Start program had a greater or lesser effect on children than ERF implemented in preschools not funded by Head Start.

We note that when examining subgroups defined by a variable like Head Start funding (rather than a child-level variable such as gender), which varies little within a site, we are no longer comparing similar sets of sites. For instance, only 27 of the 65 sites in the full sample contain at least one preschool that receives Head Start funding; 49 of the 65 sites contain at least one preschool that receives no Head Start funding. It is, of course, likely that Head Start funding is correlated with other aspects of the sites, preschools, classrooms, and the children that they serve. Therefore, any differences in impacts that we observe across the two types of sites (those with and without preschools receiving Head Start funding) may be related to aspects of these sites rather than to their funding sources. Thus, it is especially important to interpret any differences cautiously.

Unlike the patterns for other subgroups examined, differences in impacts across children in preschools that received Head Start funding and those that do not are generally large, although these differences are statistically significant only for expressive vocabulary. For preschools that received no Head Start funding, the pattern of impacts is similar to what we observed for the full study sample: effect sizes up to .48 on print-awareness standard scores, effect sizes of .41 on auditory comprehension standard scores, and effect sizes of less of .07 on phonological awareness and expressive vocabulary; however, none of these impact estimates is statistically significant at conventional levels. Estimated impacts on social-emotional outcomes are in the preferred direction (positive for social competence and negative for anxiety-withdrawal and anger-aggression) but are not statistically significant.

The pattern of impacts differs for children in preschools receiving Head Start funding: we find small and negative but not statistically significant impacts in the print- and letter-knowledge and phonological awareness domains. In the oral language domain, we find small, negative, and not statistically significant impacts on auditory comprehension and large, negative, and statistically significant impacts on expressive vocabulary. The pattern of unfavorable results for children in Head Start preschools persists for the social-emotional outcomes. Although not statistically significant, the effect size on social competence is -.21, and the effect size on anxiety-withdrawal is .49, indicating an increase in anxious-withdrawn behavior among this group (see Table F.5).

Although the estimated impacts for children in preschools receiving Head Start funding are different in sign and magnitude from those for children in preschools not receiving Head Start funding, these differences are generally not statistically significant at conventional levels, with

the exception of the impacts on expressive vocabulary.[88] Nonetheless, the different pattern of results for children in preschools receiving Head Start funding compared to other children could suggest that ERF may not be as effective in preschools that receive some Head Start funding as in preschools that receive no Head Start funding. This lack of effectiveness in Head Start preschools could indicate that ERF is less effective among the particular population served by Head Start; that Head Start preschools implement ERF less effectively than other preschools; that Head Start is already positively affecting children's outcomes, which makes it difficult for ERF to improve children's early literacy skills over and beyond any gains already caused by Head Start; or that Head Start status could be confounded with other unobserved place-based factors.[89] We note that data presented in Table E.3 showed that impacts for teachers' professional development and for observed classroom practices related to language, early literacy, and assessment practices were similar in Head Start and non-Head Start preschools. The findings from Appendix E do not support the hypothesis that Head Start preschools implemented ERF less effectively than other preschools. Given the lack of statistically significant differences in child impacts and the similarity of classroom impacts across the two subgroups, strong conclusions about the relative effectiveness of ERF in preschools that receive Head Start funding versus preschools that receive no Head Start funding are not warranted.

[88] The difference in impacts across the two groups is statistically significant, even after adjusting for the multiple comparisons within the domain for these two subgroups by using the Benjamini-Hochberg procedure (Benjamini and Hochberg, 1995).

[89] Alternatively, the different pattern of results may be simply due to chance, as might be expected when estimating impacts for a large set of subgroups.

Table F.5. ERF impacts on child outcomes by funding source of center

Outcome (range)	Head Start funding		No Head Start funding		P-value of difference in impacts between subgroups
	Effect size[a]	P-value	Effect size[a]	P-value	
Language and Literacy Skills					
Print and Letter Knowledge					
Print awareness, raw score (0–36)	–0.18	0.577	0.57	0.055	0.194
Print awareness, standard score (58–144)	0.18	0.538	0.48	0.043	0.272
Phonological Awareness					
Elision, raw score (0–18)	–0.15	0.494	0.07	0.692	0.899
Oral Language					
Expressive vocabulary, raw score (0–99)	–0.83	0.015*	0.21	0.485	0.013*
Expressive vocabulary, standard score (53–147)	–0.79	0.016*	0.22	0.442	0.010*
Auditory comprehension, raw score (1–62)	–0.03	0.895	0.41	0.185	0.185
Auditory comprehension, standard score (50–135)	–0.08	0.730	0.39	0.157	0.136
Number of Students	495		873		
Number of Sites	27		49		
Social Competence and Behavior Evaluation (Scales range from 0 to 50)					
Social competence	–0.21	0.486	0.28	0.298	0.184
Anxiety-withdrawal	0.49	0.087	–0.28	0.160	0.092
Anger-aggression	–0.03	0.907	–0.33	0.163	0.462
Number of students	498		893		
Number of sites	27		49		

*p-value (of effect size or difference between subgroups) < 0.05, two-tailed test.
[a]All estimates were obtained from a regression model of the outcome variable on an indicator variable of ERF grant receipt; grant application score; and an indicator variables of female, using SAS's PROC MIXED procedure. Language and literacy skill models also control for indicator variables of fall assessment taken in Spanish and missing fall assessment data and age at spring assessment. SCBE models also control for an indicator variable of missing fall SCBE data and age at spring SCBE observation. Missing values of covariates are mean-imputed by site and gender.
[b]The effect size was calculated by dividing the estimated impact by the standard deviation of the outcome measure (that is, the impact expressed as a percentage of the standard deviation).
NOTE: All figures were estimated using sample weights to account for the sample and survey designs. Standard errors of the impact estimates account for design effects due to unequal weighting of the data and clustering at site and classroom level.
SOURCE: ERF spring child assessments and SCBE evaluations.

Impacts by Whether Preschool Is Full-Time or Part-Time

It is possible that ERF is more effective in full-time versus part-time preschools if the program's effectiveness varies with children's exposure. One hundred of the 152 preschools in our sample were classified as full-time, meaning that they served children at least six hours a day, five days a week. Estimated impacts are similar in magnitude across the two types of preschools—

estimated impacts on print and letter knowledge are slightly larger for children in full-time versus part-time preschools, but differences in impacts between the two groups are not statistically significant. There are no statistically significant impacts in any of the other outcome domains for either group, although the estimated effect size on auditory comprehension is .45 for children in part-time preschools (see Table F.6).

Table F.6. ERF impacts on child outcomes by whether the center is part-time versus full-time

	Part-time		Full-time		P-value of difference in impacts between subgroups
Outcome (range)	Effect size[a]	P-value	Effect size[a]	P-value	
Language and Literacy Skills					
Print and Letter Knowledge					
Print awareness, raw score (0–36)	0.32	0.335	0.52	0.032*	0.872
Print awareness, standard score (58–144)	0.34	0.284	0.51	0.019*	0.831
Phonological Awareness					
Elision, raw score (0–18)	0.17	0.505	0.01	0.959	0.691
Oral Language					
Expressive vocabulary, raw score (0–99)	0.05	0.874	–0.01	0.953	0.910
Expressive vocabulary, standard score (53–147)	0.14	0.670	0.01	0.958	0.934
Auditory comprehension, raw score (1–62)	0.51	0.057	0.11	0.574	0.122
Auditory comprehension, standard score (50–135)	0.45	0.152	0.17	0.409	0.233
Number of students	425		932		
Number of sites	29		50		
Social Competence and Behavior Evaluation (Scales range from 0 to 50)					
Social competence	0.53	0.157	–0.13	0.599	0.065
Anxiety-withdrawal	–0.38	0.311	0.11	0.579	0.112
Anger-aggression	0.10	0.729	–0.12	0.600	0.403
Number of students	444		935		
Number of sites	29		50		

*p-value (of effect size or difference between subgroups) < 0.05, two-tailed test.
[a]All estimates were obtained from a regression model of the outcome variable on an indicator variable of ERF grant receipt; grant application score; and an indicator variables of female, using SAS's PROC MIXED procedure. Language and literacy skill models also control for indicator variables of fall assessment taken in Spanish and missing fall assessment data and age at spring assessment. SCBE models also control for an indicator variable of missing fall SCBE data and age at spring SCBE observation. Missing values of covariates are mean-imputed by site and gender.
[b]The effect size was calculated by dividing the estimated impact by the standard deviation of the outcome measure (that is, the impact expressed as a percentage of the standard deviation).
NOTE: All figures were estimated using sample weights to account for the sample and survey designs. Standard errors of the impact estimates account for design effects due to unequal weighting of the data and clustering at site and classroom level.
SOURCE: ERF spring child assessments and SCBE evaluations.

Appendix G. Supplemental Descriptive Tables for Teacher Outcomes and Classroom Practice

This Appendix provides descriptive tables comparing the funded and unfunded classrooms on the variables discussed on the professional development, instructional practice, and classroom environment variables presented in Chapter 5 only for the Early Reading First classrooms. The tables should not be interpreted as causal estimates of program impact. In a regression discontinuity design, simple comparisons of group means can provide misleading estimates of impacts because those means are not conditioned on the proper functional form of the grant application score. Chapter 6 and the supplemental tables in Appendices D and E provide regression-based estimates of the program impact on these variables that condition on the application score.

Table G.1. Hours of professional development in language and literacy topics received in the past 12 months, by ERF funding status

	Overall	Funded classes	Unfunded classes	P-value[1]
Hours (median)	25.0	55.0	12.0	
Hours (mean)	42.8	71.5	16.1	0.01
Standard deviation	65.7	84.7	14.5	
Sample size	178.0	86.0	92.0	

[1] P-value based on Student's t-test.
SOURCE: Spring teacher surveys.

Table G.2. Topics in which teachers received professional development in the past 12 months (percent of teachers, by topic and ERF funding status)

Topic	Overall %	Funded classes %	Unfunded classes %	P-value[1]
Language Development and Early Literacy				
Phonemic & phonological awareness	81.4	100.0	64.7	0.01
Literacy-rich environments	83.0	97.8	69.6	0.01
Concepts of print writing & prewriting	79.9	96.7	64.7	0.01
Oral language	76.3	96.7	57.8	0.01
Facilitating emergent literacy	79.4	95.7	64.7	0.01
Alphabetic knowledge	72.7	92.4	54.9	0.01
Oral comprehension & cognition	67.0	88.0	48.0	0.01
Child Assessment				
Assessment	82.0	90.2	74.5	0.01
Child Development and Behavior				
Early childhood growth & development	65.5	76.1	55.9	0.01
Classroom management	67.5	76.1	59.8	0.01
Other Topics				
Other	46.4	56.5	37.3	0.01
Distribution of the number of topics in which teachers received professional development				
0	4.1	0.0	7.8	
1 to 4	13.9	1.1	25.5	
5 to 8	24.2	21.7	26.5	
9 or 10	57.7	77.2	40.2	
Mean # of topics (SD)	8.0 (3.32)	9.6 (1.7)	6.5 (3.7)	0.01*
Sample Size	194	92	102	

[1] P-value based on Student's t-test; all other p-values are based on Pearson chi-square test.
SOURCE: Spring teacher surveys.

Table G.3. Mean number of professional development topics, by method of training and ERF funding status

Training method	Overall Mean (SD)	Funded classes Mean (SD)	Unfunded classes Mean (SD)	P-value[1]
In-service	6.10 (4.03)	7.60 (3.48)	4.75 (4.04)	< 0.01
Mentor or tutor	2.81 (4.19)	4.73 (4.54)	1.09 (2.96)	< 0.01
Workshops	3.01 (4.01)	4.52 (4.42)	1.65 (3.01)	< 0.01
CE courses	1.68 (3.40)	2.48 (4.00)	0.95 (2.55)	< 0.01
National meetings	0.97 (2.49)	1.20 (2.81)	0.77 (2.16)	0.24
Other	0.40 (1.49)	0.55 (1.76)	0.26 (1.19)	0.18
Sample Size	194	92	102	

[1] P-value based on Student's t-test.
SOURCE: Spring teacher surveys.

Table G.4. Teacher professional development through formal education

	Percentage			
	Overall	Funded	Unfunded	P-value[1]
Percentage of teachers currently enrolled in teacher-related training or education	35.1	42.4	28.4	0.01
Child development associate (CDA)	2.6	4.3	1.0	
Teaching certificate program	3.1	2.2	3.9	
Special education teaching degree	0.5	0.0	1.0	
Associate's degree	2.1	0.0	3.9	
Bachelor's degree	6.7	5.4	7.8	
Graduate degree	11.9	17.4	6.9	
Other	8.2	13.0	3.9	
Not currently enrolled	64.9	57.6	71.6	
Sample size	194	92	102	

[1] P-value based on Pearson chi-square test.
SOURCE: Spring teacher surveys.

Table G.5. Sources of funding for professional development, by number of topics and ERF funding status, percent of teachers

Funding source	Overall %	Funded classes %	Unfunded classes %	P-value[1]
ERF				
No topics	—	17.4	—	
One topic	—	0.0	—	
Multiple topics	—	82.6	—	—
School district				
No topics	50.5	43.5	56.9	
One topic	7.7	6.5	8.8	
Multiple topics	41.8	50.0	34.3	0.09
Head Start				
No topics	66.5	68.5	64.7	
One topic	2.6	4.3	1.0	
Multiple topics	30.9	27.2	34.3	0.22
State preschool				
No topics	81.4	80.4	82.4	
One topic	2.6	2.2	2.9	
Multiple topics	16.0	17.4	14.7	0.84
Teacher				
No topics	89.7	87.0	92.2	
One topic	3.1	4.3	2.0	
Multiple topics	7.2	8.7	5.9	0.46
Other				
No topics	78.9	82.6	75.5	
One topic	9.8	10.9	8.8	
Multiple topics	11.3	6.5	15.7	0.13
Sample Size	194	92	102	

[1] All p-values based on Pearson chi-square test.
— Not available.
SOURCE: Spring teacher surveys.

Table G.6. Number of curricula per classroom, by ERF funding status

	Overall %	Funded classrooms %	Unfunded classrooms %	P-value
Percent of classrooms using:				
A single curriculum	45.4	39.1	51.0	
A combination of curricula	53.6	60.9	47.0	0.08[1]
No curriculum	1.0	0.0	2.0	
Average number of curricula used (SD)	1.77 (1.12)	1.88 (1.00)	1.68 (1.22)	0.20[2]
Sample Size	194	92	102	

[1] P-value is based on Pearson chi-square test.
[2] P-value is based on Student's t-test.
SOURCE: Spring teacher surveys.

Table G.7. Percentage of teachers reporting use of specific curricula, by ERF funding status

Curriculum	Overall %	Funded classrooms %	Unfunded classrooms %	P-value[1]
Creative Curriculum	52.1	45.7	57.8	0.09
High/Scope (Educating Young Children)	26.3	23.9	28.4	0.48
Building Language for Literacy	12.9	16.3	9.8	0.18
Doors to Discovery	10.3	15.2	5.9	0.03
Let's Begin with the Letter People	9.8	15.2	4.9	0.02
Opening the World of Learning	5.7	12.0	0.0	< 0.01
We Can!	4.6	8.7	1.0	0.01
DLM Early Childhood Express	5.7	7.6	3.9	0.27
Breakthrough to Literacy	3.1	6.5	0.0	< 0.01
Creating Child-Centered Classrooms	7.2	4.3	9.8	0.14
Scholastic Curriculum	3.6	3.3	3.9	0.81
CIRCLE	2.6	3.2	1.9	0.57
SRA Open Court Reading	3.6	2.2	4.9	0.31
Montessori	3.1	2.2	3.9	0.48
High Reach Learning	2.6	0.0	8.4	0.03
Other	24.2	21.7	26.5	0.44
Sample Size	194	92	102	

[1] P-values are based on Pearson chi-square test.
NOTE: Percentages exceed 100 because teachers may be using multiple curricula. "Other" includes all curriculum reported by four or fewer teachers.
SOURCE: Spring teacher surveys.

Table G.8. Number of assessments per classroom, by ERF funding status

	Overall %	Funded classrooms %	Unfunded classrooms %	P-value
No. of assessments per classroom:				
No assessment	4.6	2.2	6.9	
Single assessment	51.0	33.7	66.7	
Combination assessments	44.3	64.1	26.5	< 0.01[1]
Mean (SD)	1.64 (1.06)	2.11 (1.21)	1.23 (0.67)	< 0.01[2]
Sample Size	194	92	102	

[1] P-value is based on Pearson chi-square test
[2] P-value is based on Student's t-test.
SOURCE: Spring teacher surveys.

Table G.9. Instruments used to assess children's progress and needs within the previous 30 days, by ERF funding status

Assessment Instruments	Overall %	Funded classrooms %	Unfunded classrooms %	P-value[1]
Peabody Picture Vocabulary Test	17.0	33.7	2.0	< 0.01
Child Observation Record	23.7	26.1	21.6	0.46
Creative Curriculum Continuum	28.9	21.7	35.3	< 0.01
Preschool Individual Growth & Development Inventory	12.4	21.7	3.9	< 0.01
Phonological Awareness Literacy Screening	8.8	17.4	1.0	< 0.01
Teacher Rating of Oral Language & Literacy	6.2	12.0	1.0	< 0.01
Work Sampling	5.7	12.0	0.0	< 0.01
Desired Results	9.3	9.8	8.8	0.82
Brigance Inventory of Early Development	4.1	6.5	2.0	0.11
Learning Accomplishment Profile—Diagnostic (LAP-D)	6.7	4.3	8.8	0.21
State- or School District-designed	4.1	4.3	3.9	0.88
Galileo	3.6	2.2	4.9	0.31
Expressive One Word Picture Vocabulary Test	5.2	0.9	0.0	< 0.01
Get Ready to Read	2.6	0.0	4.9	0.03
Other[2]	26.3	28.3	24.5	0.55
Sample Size	194	92	102	

[1] P-values are based on Pearson chi-square test.
[2] "Other" includes all assessments reported by four or fewer teachers.
SOURCE: Spring teacher surveys.

Table G.10. General quality of the preschool classroom, based on ECERS-R and TBRS subscales

	Funded classrooms Mean / (SD)			Unfunded classrooms Mean / (SD)		
	Fall	Spring	Diff.	Fall	Spring	Diff.
ECERS-R Teaching and Interactions Subscale Score	5.653 (1.074)	5.776 (1.026)	+0.123	5.432 (1.116)	5.093 (1.033)	–0.339
General Teaching Behavior	3.143 (0.560)	3.137 (0.523)	–0.006	2.975 (0.631)	2.725 (0.599)	–0.250
Classroom Community	3.175 (0.593)	3.194 (0.558)	+0.019	2.960 (0.662)	2.753 (0.690)	–0.207
Teacher Sensitivity	3.107 (0.676)	3.067 (0.623)	–0.040	2.993 (0.715)	2.689 (0.687)	–0.304
Lesson Plans	3.060 (0.811)	3.051 (0.903)	–0.009	2.504 (1.020)	2.409 (1.006)	–0.095
Quality and Organization of Activity Centers	3.123 (0.674)	2.929 (0.725)	–0.194	2.698 (0.761)	2.379 (0.739)	–0.319
Team Teaching Ability	2.975 (0.834)	2.992 (0.881)	+0.017	2.729 (0.997)	2.397 (0.939)	–0.332
Math Concepts	2.333 (1.041)	2.353 (1.008)	+0.020	2.346 (0.929)	1.824 (0.858)	–0.522
Total TBRS Score	2.714 (0.608)	2.645 (0.646)	–0.069	2.331 (0.586)	2.072 (0.528)	–0.259
Sample size	78	78		91	91	

SOURCE: Fall and spring classroom observations.

References

Ackerman, Deborah J. and W. Steven Barnett (2006). Increasing the Effectiveness of Preschool Programs. *Preschool Policy Brief* (11). New Brunswick, NJ: National Institute for Early Education Research (NIEER).

Aiken, L. S., S. G.West, D. E. Schwalm, J. Carroll, and S. Hsiung (1998). "Comparison of a randomized and two quasi-experiments in a single outcome evaluation: Efficacy of a university-level remedial writing program," *Evaluation Review*, 22(2), 207–244

Barnett, W. Steven (2004). Better Teachers, Better Preschools: Student Achievement Linked to Teacher Qualifications. *Preschool Policy Matters* (2). New Brunswick, NJ: National Institute for Early Education Research (NIEER).

Barnett, W. Steven, Karen Schulman, and Rima Shore. (2004). Class Size: What's the Best Fit? *Preschool Policy Matters* (9). New Brunswick, NJ: National Institute for Early Education Research (NIEER).

Benjamini, Yoav and Yosef Hochberg, (1995). "Controlling the false discovery rate: a practical and powerful approach to multiple testing," *Journal of the Royal Statistical Society*. Series B (Methodological), 57(1), pp. 289–300.

Bjorklund, A. and R. Moffitt (1987). "The Estimation of Wage Gains and Welfare Gains in Self-Selection Models." *Review of Economics and Statistics*, 69.

Black, Dan, Jose Galdo, and Jeffrey Smith (June 2005). "Evaluating the Regression Discontinuity Design Using Experimental Data," unpublished paper.

Brooks-Gunn, Jeanne, Lisa J. Berlin, and Alison S. Fuligni (2000). "Early Childhood Intervention Programs: What about the Family?" in Jack P. Shonkoff and Samuel J. Meisels (Eds.), Handbook of Early Childhood Intervention, second edition. New York: Cambridge University Press, pp. 549–588.

Brownell, R. (2000). Expressive One-Word Picture Vocabulary Test Manual. Academic Therapy Publications, Novato, CA.

Buddelmeyer, Hielke and Emmanuel Skoufias (2003). "An Evaluation of the Performance of Regression Discontinuity Design on PROGRESA," IZA Discussion Paper No. 827.

Clifford, Richard M., Oscar Barbarin, Florence Chang, Diane Early, Donna Bryant, Carollee Howes, Margaret Burchinal, Robert Pianta. (2005). "What is Pre-Kindergarten? Characteristics of Public Pre-Kindergarten Programs." *Applied Developmental Science,* 9(3):126–143.

Clifford, R., Margaret Burchinal, T. Harms, H. Rossbach. (1996). Factor structure of the Early Childhood Environment Rating Scale (ECERS): An international comparison (unpublished paper). FPG Child Development Institute, University of North Carolina at Chapel Hill.

Denham, Susan A. and Rosemary Burton (2005). *Social and Emotional Prevention and Interventional Programming for Preschoolers*. New York, NY: Springer Publishing Company.

Denham, Susan A., Sarah Caverly, Michelle Schmidt, Kimberly Blair, Elizabeth DeMulder, Selma Caal, Hideko Hamada, and Teresa Mason (2002). "Preschool Understanding of Emotions: Contributions to Classroom Anger and Aggression." *Journal of Child Psychology and Psychiatry* 43(7), 901-916.

Dumas, Jean E., Alfonso Martinez, and Peter J. LaFreniere (1998). "The Spanish Version of the Social Competence and Behavior Evaluation (SCBE) Preschool Edition: Translation and Field Testing." *Hispanic Journal of Behavioral Sciences*: 20(2): 255-269.

Duncan, S. E., and E.A. DeAvila. (1998). Pre-LAS 2000. Monterey, CA: CTB/McGraw-Hill.

Early, Diane M., Donna M. Bryant, Robert C. Piata, Richard M. Clifford, Margaret M. Burchinal, Sharon Ritchie, Carollee Howes, and Oscar Barbarin (2006). "Are Teachers Education, Major, and Credentials related to Classroom Quality and Children's Academic Gains in Pre-Kindergarten?" *Early Childhood Research Quarterly* 21(2): 175-195.

Fan, J. (1992). "Design-adaptive Nonparametric Regression." *Journal of the American Statistical Association*. 87: 998-1004

Frank Porter Graham Child Development Institute (2004) "Program Evaluation" *Early Developments* 8(3). University of North Carolina at Chapel Hill.

Hahn Jinyong, Petra Todd, and Wilbert van der Klaauw (2001). "Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design." *Econometrica* 69(1): 201–209.

Harms, Thelma, Richard M Clifford, and Debby Cryer. (1998). *Early Childhood Environment Rating Scale, Revised Edition.* New York: Teachers College Press.

Hart, Betty, and Todd R. Risley. (1995). *Meaningful Differences in the Everyday Experience of Young American Children*. Baltimore: Paul H. Brookes Publishing Co.

Heckman, James. (1997). "Instrumental Variables: A Study of Implicit Behavioral Assumptions Used in Making Program Evaluations." *Journal of Human Resources*, 32:3.

Heckman, J. and E. Vytlacil, (1999). "Local Instrumental Variables and Latent Variable Models for Identifying and Bounding Treatment Effects." *Proceedings of the National Academy of Sciences*, 96:8.

Hedges, Larry. (2004) "Effect Sizes in Multisite Designs Using Assignment by Cluster," Working Paper. Chicago, IL: University of Chicago.

Hyde, Janet S. and Marcia C Linn. (1988), "Gender Differences in Verbal Ability: A Meta-Analysis." *Psychological Bulletin,* 104:53-69.

Irish, Kate, Rachel Schumacher, and Joan Lombardi (2004). Head Start Comprehensive Services: A Key Support for Early Learning for Poor Children. Policy Brief (4), Center for Law and Social Policy.

LaFreniere, P. J., and F. Capuano (1997). "Preventive intervention as a means of clarifying direction of effects in socialization: Anxious-withdrawn preschoolers." *Development and Psychopathology,* 9, 551-564.

LaFreniere, Peter J., and Jean E. Dumas (1996). "Social Competence and Behavior Evaluation in Children Ages 3 to 6 Years: The Short Form (SCBE-30)." *Psychological Assessment*, 8(4):369–377.

LaFreniere, P.J., J. Dumas, D. Dubeau and F. Capuano (1992). "The development and validation of the preschool socio-affective profile." *Psychological Assessment: Journal of Consulting and Clinical Psychology*, *4* (4), 442-450.

LaFreniere, P.J., N. Masataka, M. Butovskaya, Q. Chen, M.A. Dessen, K. Atwanger, S. Shreiner, R. Montirosso, and A. Frigerio (2002). "Cross-Cultural Analysis of Social Competence and Behavior Problems in Preschoolers." *Early Education and Development*, 13 (2).

Landry, Susan H. (2005). Effective Early Childhood Programs: Turning Knowledge Into Action. University of Texas Houston Health Science Center.

Landry, Susan H., April Crawford, Susan B. Gunnewig, and Paul R. Swank. (2004). "Teacher Behavior Rating Scale (TBRS)," Center for Improving the Readiness of Children for Learning and Education, unpublished research instrument.

Landry, Susan H., Paul R. Swank, Karen E. Smith, Michael A. Assel, and Susan B. Gunnewig. (2006). "Enhancing Cognitive Readiness for Pre-School Children: Bringing a Professional Development Model to Scale," *Journal of Learning Disabilities*, 39(4): 306-325.

Lee, David and David Card. (2006). "Regression Discontinuity Inference with Specification Error." NBER Technical Working Paper 322.

Lonigan, C.J., R. K. Wagner, and C.A. Rashotte (2002). The Preschool Comprehensive Test of Phonological and Print Processing. Florida State University.

Lonigan, C. J., R. K. Wagner, J. K. Torgesen, and C. A. Rashotte (2007). Preschool Comprehensive Test of Phonological and Print Processing (Pre-CTOPPP). Austin, TX: PRO-ED.

Ludwig, Jens and Douglas L. Miller (2007). "Does Head Start Improve Children's Life Chances? Evidence from a Regression Discontinuity Design." *Quarterly Journal of Economics*, 122(1), 159-208.

McCrary, Justin (2005). "Manipulation of the Running Variable in the Regression Discontinuity Design." Unpublished Paper, University of Michigan.

Murray, D. M. (1998). *Design and analysis of group-randomized trials*. Monographs in Epidemiology and Biostatistics (Vol. 27). New York: Oxford University Press.

National Institute for Early Childhood Education Research (2006). *The State of Preschool 2006: State Preschool Yearbook*. Rutgers. The State University of New Jersey.

National Research Council, Committee on the Prevention of Reading Difficulties in Young Children (1998). *Preventing Reading Difficulties in Young Children,* edited by Catherine E. Snow, M. Susan Burns, and Peg Griffin. Washington, DC, National Academies Press.

NICHD Early Child Care Research Network (2006). "Child-Care Effect Sizes for the NICHD Study of Early Child Care and Youth Development," *American Psychologist*, 61(2):96–116.

NICHD Early Child Care Research Network (2003). "Does Quality of Child Care Affect Child Outcomes at Age 4 ½ ?" *Developmental Psychology,* 39(3):451–469.

NICHD Early Child Care Research Network (2002). "Child-Care Structure, Process, and Outcome: Direct and Indirect Effects of Child-Care Quality on Young Children's Development," *Psychological Science*, 13(3), 199–206.

NICHD Early Child Care Research Network (2001). "Parenting and Family Influences when Children are in Child Care: Results from the NICHD Study of Early child Care," in J. Borkowski, S. Ramey, and M. Bristol-Power (Eds.), *Parenting and the Child's World Influences on Intellectual, Academic, and Social-emotional Development*. Mahwah, NJ: Erlbaum.

NICHD Early Child Care Research Network (1999). "Child Outcomes When Child Care Center Classes Meet Recommended Standards for Quality," *American Journal of Public Health*, 89(7), 1072–1077.

Office of Federal Register, National Archives and Records Administration (March 11, 2003). "Early Reading First Program; Notice Inviting Local Applications for New Awards in Fiscal Year (FY) 2003." *Federal Register*, vol. 68, no. 47, pp. 11705-11711. Washington, DC: Office of the Federal Register.

Peisner-Feinberg, E., and M. Burchinal (1997). "Relations between preschool children's childcare experiences and concurrent development: The cost, quality, and outcomes study," *Merrill-Palmer Quarterly*, 43(3), 451–477.

Perlman, M., G.L. Zellman, and V.N. Le (2004). "Examining the psychometric properties of the Early Childhood Environment Rating Scale-Revised (ECERS-R)," *Early Childhood Research Quarterly,* 19, 398–412.

Pianta, Robert C., Carollee Howes, Margaret M. Burchinal, Donna M. Bryant, Richard M. Clifford, Diane M. Early, and Oscar Barbarian (2005). "Features of Pre-Kindergarten Programs, Classrooms, and Teachers: Do They Predict Observed Classroom Quality and Child-Teacher Interactions?" *Applied Developmental Science*, 9(3):144–159.

Porter, Jack. (2003) "Estimation in the Regression Discontinuity Model." Unpublished working paper, Harvard University.

Pullen, P., and L. M. Justice. (2003). "Capitalizing on the Preschool Years: Strategies for Increasing Literacy," *Intervention in School and Clinic*, 29(2):87–98.

Ramsey, Philip H. (2002) "Comparison of Closed Testing Procedures for Pairwise Testing of Means," Psychological Method, 7(4).

Raudenbush, S.W., & Bryk, A.S. (2002). *Hierarchical Linear Models: Applications and data analysis methods* (2nd Edition). Thousand Oaks, CA: Sage.

Rossbach, H., R. Clifford, and T. Harms. (1991). Dimensions of learning environments: Cross national evaluation of the Early Childhood Environment Rating Scale. Paper presented at the AERA Annual Conference, Chicago.

Rubin, Donald. (1987). *Multiple Imputation for Nonresponse in Surveys*. New York: John Wiley and Sons, Inc.

Trochim, W. (1984). Research Design for Program Evaluation: the Regression-Discontinuity Approach. Beverly Hills: Sage Publications.

U.S. Census Bureau (2005). Current Population Survey, Population Estimates Program, Population Division, Washington, DC.

U.S. Department of Education (December 2005). Revised Fiscal Year 2006 Performance Plan. Washington, DC.

U.S. Department of Education (December 2004). Revised Fiscal Year 2005 Performance Plan and Interim Adjustments to the Strategic Plan. Washington, DC.

U.S. Department of Education (2003). *Guidance for the Early Reading First Program*. Washington, DC.

U.S. Department of Health and Human Services (December 2006). FACES 2003 Research Brief: Children's Outcomes and Program Quality in Head Start. Administration for Children and Families, Washington, DC.

U.S. Department of Health and Human Services. (May 2005). Head Start Impact Study: First Year Findings. Administration for Children and Families, Washington, DC.

U.S. Department of Health and Human Services (May 2004), The Head Start Management Initiative, Administration for Children and Families, Washington, DC.

U.S. Department of Health and Human Services (April 2004), Head Start Program Fact Sheet Fiscal Year 2003. Administration for Children and Families, Washington, DC.

U.S. Department of Health and Human Services (May 2003). Head Start FACES 2000: A Whole-Child Perspective on Program Performance. Fourth Progress Report. Administration for Children and Families, Washington, DC.

U.S. Department of Health and Human Services (January 2002). A Descriptive Study of Head Start Families: FACES Technical Report I. Administration for Children and Families, Washington, DC.

Vandell, Debora L., and Barbara Wolfe (2000). Child Care Quality: Does It Matter and Does It Need to Be Improved? University of Wisconsin-Madison Institute for Research on Poverty Special Report no. 78.

Wagner, R. K., J. K. Torgesen, and C. A. Rashotte. (1999). Comprehensive Test of Phonological Processing (CTOPP). Austin, TX: PRO-ED.

Whitebook, M., C. Howes, and D. Phillips. (1990). Who Cares? Childcare teachers and the quality of care in America. Final Report of the National Child Care Staffing Study. Oakland, CA: Child Care Employee Project.

Whitehurst, G.J., and C.J. Lonigan (2001). "Emergent Literacy: Development from Pre-readers to Readers," *Handbook of Early Literacy Research*, edited by Neuman, S.B., and Dickinson, D.K., Guilford Press, New York, 11–29.

Zimmerman, I. L., V.G. Steiner and R.E. Pond (2002). Preschool Language Scale—fourth edition, Examiner's Manual. San Antonio, TX: The Psychological Corporation.