THE ABBOTT PRESCHOOL PROGRAM LONGITUDINAL EFFECTS STUDY (APPLES)

Interim Report
June 2007
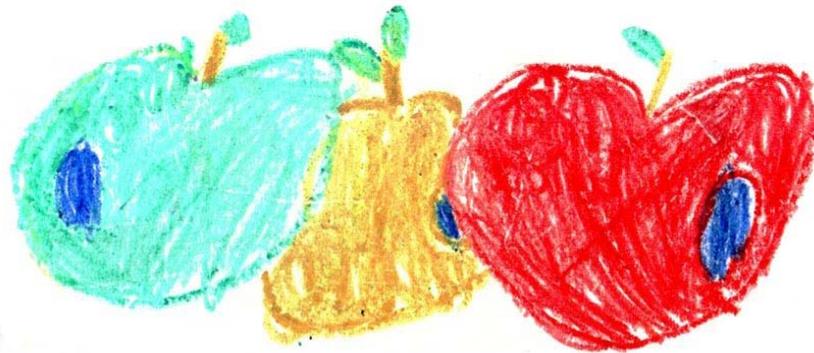
Ellen Frede, Ph.D.
National Institute for Early Education Research and
The College of New Jersey

Kwanghee Jung, Ph.D.
National Institute for Early Education Research

W. Steven Barnett, Ph.D.
National Institute for Early Education Research

Cynthia Esposito Lamy, Ed.D.
Robin Hood Foundation

Alexandra Figueras, M.S.
National Institute for Early Education Research

Acknowledgements

The Abbott Preschool Program Longitudinal Effects Study (APPLES)

Executive Summary

The New Jersey Supreme Court in *Abbott v. Burke* mandated, among other things, that the state establish high-quality preschool education for the highest-poverty school districts in the state. Since the Abbott Preschool Program began in the 1999-2000 school year, enrollment in the program has increased dramatically. In 2005-2006, the seventh year of implementation, the program served more than 40,500 3- and 4-year-old children in a mix of settings including public schools, private child care centers, and Head Start agencies.

In 2005-2006, the National Institute for Early Education Research (NIEER) undertook the subject of this report — a longitudinal study to determine if the learning gains found in early research at kindergarten entry continued into elementary school. While the study will continue to follow children beyond kindergarten, this first report follows them to the end of their kindergarten year. In partnership with the Early Learning Improvement Consortium (ELIC), a group of higher education institutions that conduct program evaluation in the Abbott Preschool Program, we also collected data on classroom quality statewide.

We measured the effects of attending the Abbott Preschool program on children's learning using two methods. The first is a rigorous regression-discontinuity design (RDD) that has been applied previously in the Abbott studies and in studies in other states. The RDD method provides greater assurance that estimates are unbiased, but can estimate effects only at kindergarten entry. We coupled the RDD results with results from a longitudinal cross-sectional design comparing children who did and did not attend the Abbott pre-K program. To the extent that estimates from the two approaches are similar, we can have greater confidence in the results of the longitudinal analyses as the study moves forward.

The findings of this study provide clear evidence of the following: 1) classroom quality in the Abbott Preschool Program continues, on the whole, to improve; 2) that children who attend the program, whether in public schools, private settings or Head Start, are improving in language, literacy, and math at least through the end of their kindergarten year; and 3) that children who attend preschool for two years at both age 3 and 4 significantly out-perform those who attend for only one year at 4 years of age or do not attend at all.

Significant findings are:

Classroom Quality

- There have been notable advances in classroom quality scores. In 2006, almost 90 percent of the classrooms scored above the mean score found in 2000. Areas most likely to be directly related to child learning, language and reasoning, activities,

interactions, and program structure scored in the good to excellent range. Less progress has been made in improving teaching practices specifically related to children's learning in math.

- Significantly, public school settings and child care center classrooms scored virtually the same across almost all measures of quality teaching practices. Some minor differences were found. Public school classrooms somewhat out-performed private settings in early literacy support while private settings scored somewhat higher on promoting oral language and items related to enjoyment of books.

Study Outcomes (Kindergarten Year)

- The RDD results show that substantial gains in learning and development occurred in language, literacy, and mathematics. The longitudinal study finds that these gains are largely sustained during the kindergarten year. Even children who did not attend preschool made some gains in the kindergarten year. For example, they gained nearly .25 of standard deviation and closed 18 percent of the gap between their scores and the national average in vocabulary, our broadest measure.

- However, the children who attended Abbott pre-K also continued to close the gap and those who attended for two years had closed over half the gap with the national average vocabulary score by the end of kindergarten.

- Similarly, in mathematics children who had one and two years of Abbott preschool education maintained nearly all of their initial advantage through to the end of kindergarten despite strong kindergarten gains for all children.

- Only in print awareness do the children who did not attend Abbott preschool programs catch-up by the end of kindergarten, and this raises concerns about the extent to which they fell behind on more advanced skills while working to acquire the basics.

Considerable attention and resources have been invested in the Abbott Preschool Program. According to NIEER's annual report on state-funded preschool, the Abbott program ranks as one of the highest quality state preschool programs in the nation, as the highest in providing access to 3-year-olds, and as the most well-funded (Barnett, Hustedt, Hawkinson, and Robin, 2006). As such, there is a great deal of interest in whether it is effective in helping children enter kindergarten with the knowledge, skills and dispositions that will lead to success in school. The results presented here provide clear evidence that by participating in a high-quality program, regardless of auspice, children are improving in literacy and math at least until the end of the kindergarten year.

This initial report, focusing on child data collected in the kindergarten year (Fall 2005 and Spring 2006) is the first in a series on the effects of the Abbott Preschool Program. Future reports will present results through the end of fourth grade and will include information on grade retention and special education placement as well as test results.

Introduction

As part of the landmark New Jersey Supreme Court school funding case, *Abbott v. Burke*, the Court established the Abbott Preschool Program. Beginning in the 1999-2000 school year, 3- and 4- year old children in the highest poverty districts in the state were able to receive a high-quality preschool education that would prepare them to enter school with the knowledge and skills necessary to meet the New Jersey Preschool Teaching and Learning Expectations: Standards of Quality (NJ Department of Education, 2004b) and the Kindergarten New Jersey Core Curriculum Content Standards (NJDOE, 2004a). Through a Department of Education (DOE) and Department of Human Services (DHS) partnership, Abbott preschool classrooms combine a DOE-funded six-hour, 180-day component with a DHS-funded wrap-around program that provides daily before- and after-care and summer programs. In total, the full-day, full-year program is available 10 hours per day, 245 days a year.

Enrollment in the Abbott preschool program has increased dramatically since its inception in 1999. During the 2005-2006 school year, the seventh year of Abbott preschool implementation, the 31 Abbott districts served more than 40,500 3- and 4-year-old children in preschool – 78 percent of a possible 52,160 children. The enrollment for the 2006-2007 school year is more than 39,678 children with a DOE budget of almost $500 million. Through contracts with the school districts, private child care providers and Head Start agencies, in addition to public schools, offer of Abbott Preschool:

- 37 percent of children are served in district-run classrooms,
- 7 percent are served in Head Start classrooms, and
- 56 percent are in private provider classrooms.
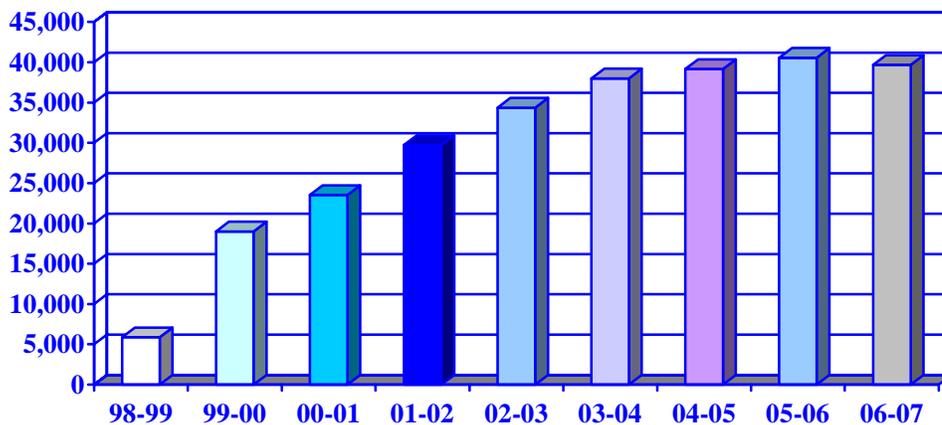
Figure 1:
Abbott Preschool Program Enrollment 1998-2007

Since 2002, the NJ DOE has implemented an assessment system for the New Jersey Abbott Preschool Program. To measure and assess progress statewide, the DOE formed the Early Learning Improvement Consortium (ELIC) by bringing together a group of the state's top early childhood education faculty. Drawing on research previously conducted by the Center for Early Education Research (Barnett, Tarr, Esposito Lamy and Frede, 2002), ELIC is responsible for collecting and reporting on data on children and classrooms.   Every fall from 2002 through 2005, assessments of kindergartners' skills were conducted to measure progress toward preparing children to succeed in school. In addition, members of ELIC conduct classroom observations on randomly selected Abbott preschool classrooms to measure progress in program quality. Findings have been reported yearly (Frede et al, 2004; Lamy, Frede and ELIC, 2005) and can be found at www.nj.gov/njded/ece.

In the 2004-2005 school year, ELIC reported that classroom quality scores had reached acceptable levels, and children were entering kindergarten with the gap in language and literacy skills substantially narrowed (Frede, et al, 2004; Lamy, Frede, and ELIC, 2005).   Given these trends, a longitudinal evaluation seemed warranted to determine if the learning gains continued into elementary school. This report describes the methods and results for assessing classrooms and investigating child outcomes by following subjects to the end of the kindergarten year.

**Research Methods and Results for Assessing
Abbott Preschool Classroom Quality**

Structured classroom observations have been conducted since the inception of the Abbott program in 1999 through 2006. The Center for Early Education Research at Rutgers University (the predecessor of NIEER) measured classroom quality in a subsample of 19 Abbott districts in 1999 through 2001. For the current report, change in aspects of classroom quality since before 2002 is measured across this subsample. Beginning in 2003, ELIC administered observations annually in all Abbott districts. Trained data collectors observed in randomly selected preschool classrooms using structured classroom observation instruments that assess materials, the environment, and teacher-child interactions.

Classroom Observation Instrumentation and Protocol

ELIC administers the following three instruments each year to measure classroom quality:

Early Childhood Environment Rating Scale - Revised (ECERS-R). Overall program quality is assessed by trained observers using a standardized measure of preschool classroom structure and process, the Early Childhood Environment Rating Scale – Revised (ECERS-R; Harms, Clifford & Cryer, 2005). This measure has been used extensively in the field and has well-established validity and reliability. The validity of the measure is supported by high correlations between both the scale items and ratings of items as highly important by a panel of nationally recognized experts, and between scale scores and ratings of classroom quality by experts. Internal consistency as measured by Cronbach's alpha is reported by the authors to be adequate, ranging from .81 to .91. Classroom quality is rated on a 7-point Likert scale, indicating a range of quality from inadequate (1) to excellent (7). The seven ECERS-R subscales are as follows: Space and Furnishings, Personal Care Routines, Language-Reasoning, Activities, Interaction, Program Structure, and Parents and Staff. Average subscale scores are calculated, as well as a total scale score averaged across all 43 items in the scale.

Table 1 provides the internal consistency of each of the subscales and the total scale using Cronbach's Alpha for this sample. Four of the subscale alphas show low internal consistency (Personal Care Routines, $\alpha = .54$, Space and Furnishings, $\alpha = .63$, Program Structure, $\alpha = .63$, and Parents and Staff, $\alpha = .63$). The total scale has excellent internal consistency ($\alpha = .90$).

Table 1.
Cronbach's Alpha for ECERS-R Subscales and Total Scale

Scale	Number of Items	Cronbach's Alpha
Space and Furnishings	8	.63
Personal Care	6	.54
Language and Reasoning	4	.70
Activities	10	.81
Interactions	5	.82
Program Structure	4	.63
Parents and Staff	6	.63
Total w/o parent and staff subscale	37	.89
Total Scale (all items)	43	.90

The Supports for Early Literacy Assessment (SELA). The extent to which the classroom environment is supportive of children's literacy development is measured with the Supports for Early Literacy Assessment (SELA; Smith, Davidson & Weisenfeld, 2001). This measure is revised for use by this project with the deletion of 4 items that overlap with the ECERS-R. The revised measure includes 16 items on a scale from 1 to 5, low quality (1) to high quality (5) for the support of early literacy development. Six subscales are: The Literate Environment, Language Development, Knowledge of Print/Book Concepts, Phonological Awareness, Letters and Words, and Parent Involvement. Internal consistency among scale items as measured by Cronbach's alpha on the current sample is good at .87.

The Preschool Classroom Mathematics Inventory (PCMI). The classroom support for the development of children's early mathematical skills is measured using the Preschool Classroom Mathematics Inventory (PCMI; Frede, Weber, Hornbeck, Stevenson-Boyd & Colon, 2005). This tool measures the materials and strategies used in the classroom to support children's early mathematical concept development, including counting, comparing, estimating, recognizing number symbols, classifying, seriating, geometric shapes, and spatial relations. The standards of the National Council of Teachers of Mathematics and the National Association for the Education of Young Children (2002) inform the measure, which is comprised of 11 items on a 5-point scale, from low quality (1) to high quality (5), and has two subscales, Materials and Numeracy and Other Mathematical Concepts. Internal consistency among the test items as measured by Cronbach's alpha is good at .86. The PCMI has been found to predict child progress on a standardized math assessment (Frede, Lamy and Boyd, 2005).

<u>Observer Training Procedures</u>

The members of ELIC hire and train observers who have specific expertise in early childhood education or a closely related field. Initial training in administering the observation protocols takes place in two full-day workshops. Trainees then observe in preschool classrooms alongside a trained observer to establish reliability on each observation instrument. The scores of the trainee and the reliable observer are then compared, item by item. The true score for each item is determined through discussion but is generally that of the trained observer. A reliability score for the trainee is computed by determining how many exact matches by item she/he has with the true score and how many are only one point above or below the true score. For the ECERS-R, the trainee must complete three observations with 80 percent or above exact matches or one-away from the true score and no less than 65 percent exact agreement.   The trainee must achieve 70 percent exact agreement for the PCMI and SELA for all three sessions. After five sessions, if the observer is not reliable, he or she is not included in data collection. Shadow scoring is repeated every six weeks. Reliability scores range from 80–98 percent with an average of 87 percent.

<u>Observation Protocol</u>

ELIC developed the following standard protocol for all observations. Observations should last no less than three hours and include greeting and at least one meal or snack. When scheduling observations, observers determine if it is likely to be a typical day by asking if field trips, assemblies or planned absences are scheduled. They do not reveal which teacher will be observed. However, if the teacher has an unplanned absence, the observation is conducted anyway and the interview takes place with the teacher assistant. Substitute teachers are noted in the data but are included in analyses because our effort is to capture children's experiences in Abbott preschool classrooms. Having a substitute is one of those experiences. The observers introduce themselves to the classroom staff and briefly explain what they will be doing. They try to be as unobtrusive as possible, and limit conversations with teachers and children to minimize their impact.

Sample for Classroom Observations

In the winter and early spring of 2005-2006, 316 Abbott Preschool classrooms were observed. Classrooms were first stratified by auspice and random selection was made proportionately. The final sample consisted of 104 public school administered classrooms, 176 private child care center classrooms, and 25 Head Start classrooms. The population consisted of all classrooms serving preschool-aged children in Abbott-funded classrooms including preschool handicapped classrooms that serve only children with disabilities. The sample included randomly selected handicapped classrooms in 10 public schools and one private preschool. In the following analyses, preschool handicapped and Head Start classrooms are only included in total scores since the sample size for these two programs is too small to validly represent them.

Table 2 presents information about the classroom teachers in the sample. As can be seen, virtually all teachers are female. On average they have been teaching about 6.5 years with a wide range of experience from first year teachers to veterans with 37 years of experience. More of the child care center teachers are new to teaching ($p < .05$). Twenty-four percent of the teachers speak Spanish with over 28 percent of child care center teachers fluent in Spanish. In addition, 1 percent of these child care teachers do not speak English. Some (5 percent of the total) of the substitute teachers reported not having obtained an undergraduate degree. A far greater number of public school teachers have post graduate degrees ($p < .05$). The majority (76 percent) of the teachers has early childhood certification and 35 percent have elementary certification. The overlap is teachers with dual certification. Those with elementary certification alone met the criteria to be "grandfathered" at the time the ECE certification was implemented.

Table 2. Teacher Demographics

|	Total (including Head Start) (N=308)	Public School (n=104)	Private Program (n=169)
Female (%)	97.4%	100%	96.4%
Years of experience			
Mean (years)	6.54	6.67	5.97
Range (years)	0-37	0-37	0-35
$\leq$ 3 years experience (%)	34.5	30.8	41.1*
3.1 – 7 years experience (%)	38.5	39.4	35.7
> 7 years experience (%)	27	29.8	23.2
Language fluency (%)			
English	98.7	100.0	97.6
Spanish	24	15.4	28.4*
Other languages (Arabic, Polish, French Creole, Greek, etc)	15.6	14.4	13.6
Highest degree earned (%)			
< BA	2.9	0	1.8
BA	78.2	71.2	88.2*
MA	18.2	27.9	9.5*
Doctoral Degree	0.6	1.0	0.6
Early Childhood Teaching Certification (%)	77.6	77.9	79.3
Elementary Teaching Certification (%)	36	52.9	25.4*

* $p < .05$

Classroom Observation Results

We report results by instrument for the total sample (including Head Start and preschool handicapped), the public school classrooms and the child care center classrooms. We then compare and analyze how scores have changed in the 19 districts that were observed in 1999-2000.

Early Childhood Environment Rating Scale–Revised

In 2005-2006, the average ECERS-R score across all sample classrooms was 4.81 with a standard deviation of.75 and a range from 2.55 to 6.6. Over 40 percent scored 5 or better, placing them in the good to excellent quality range. The vast majority of classrooms (86 percent) scored above the midpoint of 4. Only about 1 percent of the sample classrooms score below 3, indicating that they provided minimal to inadequate support for children's cognitive and social development. Figure 2 also shows the distribution of scores for public school classrooms and private child care center classrooms.

Percentage of Classrooms Scoring 1-7 on the ECERS-R
2005-2006

Figure 2. ECERS-R Score
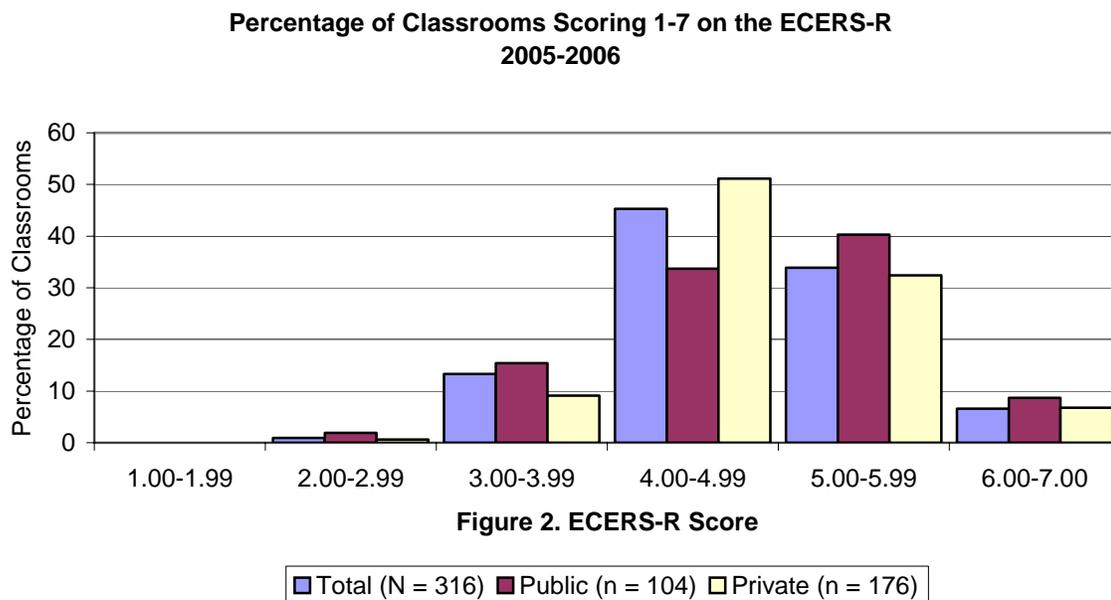
The seven subscales of the ECERS-R measure different aspects of classroom quality. Table 3 reports the ECERS-R average subscale scores and the average total score for total sample, and the scores for the sub-samples from public schools and private programs in 2006.

Table 3
ECERS-R Total and Subscale Scores across 30 Abbott districts, 2005-2006

ECERS-R Subscale	Total** (N=316)		Public School (n=104)	Private Program (n=176)
	M (*SD*)	Range	*M* (*SD*)	*M* (*SD*)
Space and Furnishings	4.72 (.90)	2.00-6.88	4.88 (.92)	4.73 (.87)
Personal Care	4.16 (1.22)	1.17-7.00	4.00 (1.29)	4.37 (1.11)*
Language and Reasoning	5.03 (1.09)	1.50-7.00	4.98 (1.09)	5.11 (1.09)
Activities	4.34 (.94)	1.80-6.80	4.28 (1.00)	4.48 (.89)
Interactions	5.93 (1.08)	1.00-7.00	5.88 (1.16)	5.91 (1.08)
Program Structure	5.02 (1.40)	1.00-7.00	4.88 (1.50)	5.15 (1.37)
Parents and Staff	5.19 (.89)	2.33-7.00	5.58 (.83)	5.04 (.87)*
Overall ECERS-R score (w/o Parents/Staff subscale)	4.75 (.79)	2.22-6.67	4.72 (.88)	4.84 (.72)
Overall ECERS-R score	4.81 (.75)	2.55-6.60	4.84 (.83)	4.87 (.69)

* $p < .05$
** Total includes Head Start classrooms and self-contained special education preschool classrooms.

The subscale mean scores range from a low of 4.16 for Personal Care Routines to a high of 5.93 for Staff/Child Interactions. The wide ranges in each subscale show that across all areas of classroom quality – basic environment and caregiving to intellectually challenging and intentional teaching practices – there are a small number of classrooms that score close to inadequate and more that are excellent or approaching excellence. Four subscales are particularly relevant to educational effectiveness of the program: Language and Reasoning, Activities, Interactions, and Program Structure. Abbott preschool classrooms score 5 or better on three of these subscales, placing them in the good to excellent quality range. Independent sample T-tests revealed that public schools and private programs are not significantly different except on two subscales, Personal Care Routines and Provisions for Parents and Staff ($p < .05$). The private programs scored significantly higher than the public schools on the Personal Care Routines subscale (4.37 vs. 4.00) and significantly lower on Parents and Staff (5.04 vs. 5.58). Neither of these subscales is likely to be directly related to child learning.

Figure 3 below provides a comparison of the distribution of ECERS-R total scores in 2000 and 2006 across the 19 districts for which these data are available in both years. The mean score in 1999-2000 across the 19 districts was 3.86. Thus, a substantial improvement has occurred in classroom quality since 1999-2000. In 1999-2000, scores ranged from 1.19 to 6.39. Almost 24 percent of the classrooms scored below 3 compared

to only 1 percent in 2006. At the other extreme, only 14 percent scored above a 5 in 1999-2000 compared with almost 40 percent in 2006.

**Percentage of Classrooms Scoring 1-7 on the ECERS-R**
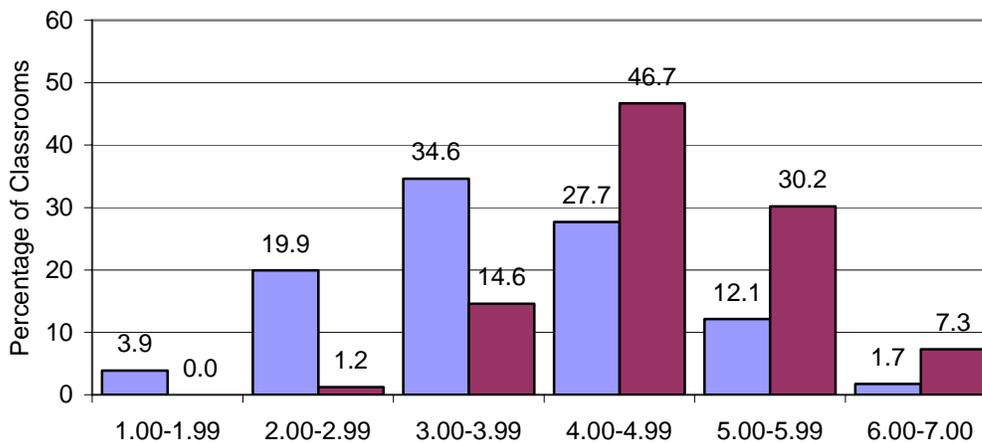
Figure 3. ECERS-R Score

□ 00 Total (N = 232) ■ 06 Total (N = 259)

In 2005-2006, the average ECERS-R mean scores for public school and private program classrooms (in 19 districts) are 4.68 ($SD = .90$) and 4.85 ($SD = .70$), respectively. Note that scores are somewhat lower than for the full sample of 316 across 30 districts. About 40 percent of the public school classrooms and about 38 percent of the private center classrooms score 5 or better. There is more variability in the public school classrooms. Figure 4 presents a comparison of public school scores from 2000 to 2006 and Figure 5 presents this comparison for private program classrooms. As can be seen, there is a steady and real shift in scores in the public school classrooms toward higher quality but the dramatic shift seen in the private programs is especially striking.

Appendix A provides ECERS-R results by subscale for the total sample and by auspice. In addition, comparisons between 2000 and 2006 are also reported.

Percentage of Classrooms Scoring 1-7 on the ECERS-R
**Public School 2000 vs 2006**
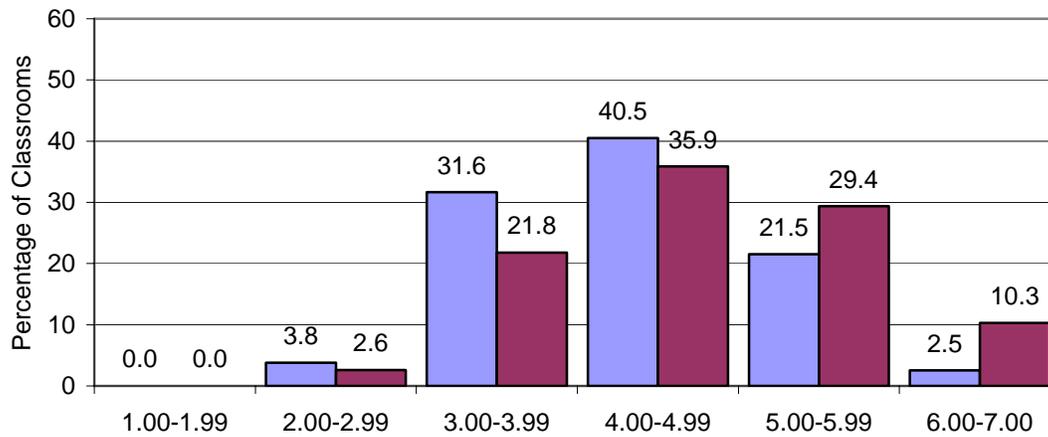
Figure 4. ECERS-R Score

00 Public (N = 79)   06 Public (N = 78)

Percentage of Classrooms Scoring 1-7 on the ECERS-R
**Private Programs 2000 vs 2006**
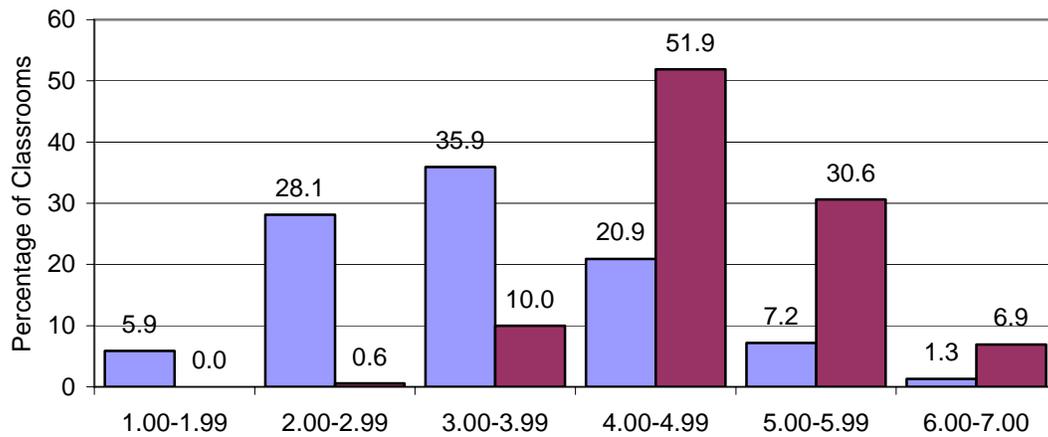
Figure 5. ECERS-R Score

00 Private (N = 153)   06 Private (N =160)

ECERS-R Factor Analysis

In addition to calculating the overall and subscale scores ECERS-R, a factor analysis was conducted. An initial principal component analysis yielded a solution that had 11 components with eigenvalues greater than 1 and explained 71 percent of the variance. In previous studies, factor analyses of ECERS-R yielded two factors (Burchinal et al., 2002).

We conducted confirmatory factor analysis with a two-component solution with varimax rotation. Note that in this factor analysis we do not include the "parents and staff" subscale items because it is scored based on teacher report and is not included in most other studies of ECERS-R. In addition, the item "provisions for children with disabilities" is not included because it was not applicable for most of the classrooms (75 percent). All items that were not distinct in their loading (item loaded on both factors and the difference is less than .20) or loaded below .40 were dropped from the consideration.

The first factor, Provisions for Learning, included 12 items and accounted for 17 percent of the variance. The Cronbach's alpha for this factor is .85. The items included are fine motor, schedule, art, blocks, space for privacy, gross motor equipment, furnishing for relaxation, room arrangement, dramatic play, nature/science, promoting acceptance of diversity, and block/pictures. The mean across 316 classrooms on this factor was 4.46 (SD = 1.01). See Figure 6 below to see the distribution of scores on this factor.

**Percentage of Classrooms Scoring 1-7
on the Provisions for Learning**
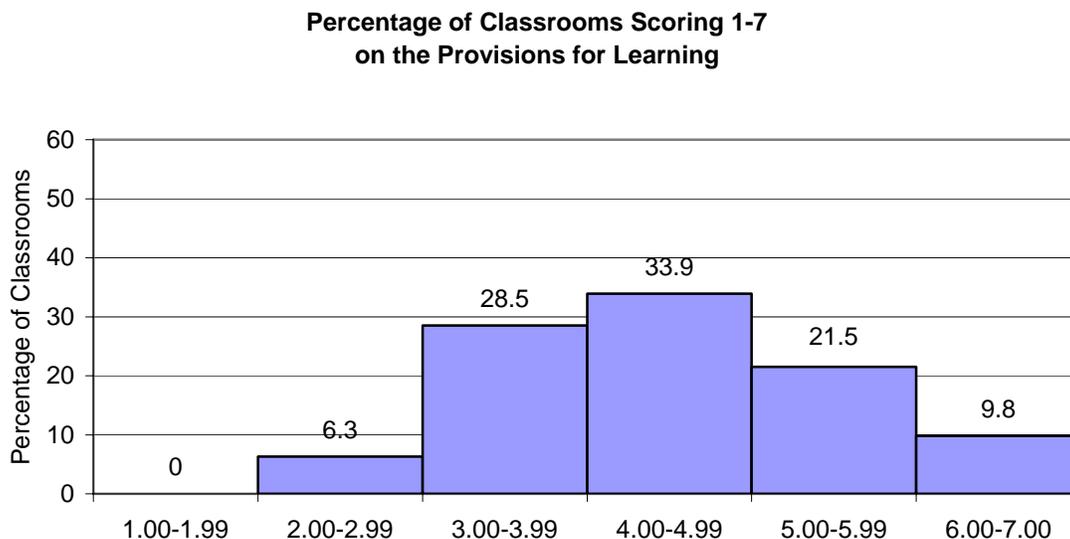
**Figure 6. ECERS-R Factor 1: Provisions for Learning**

The average scores for the Provisions for Learning factor for the public school and private program are 4.42 (SD = 1.06), and 4.60 (SD = .97), respectively. Only about 10 percent of the public school and 1 percent of the private program classrooms scored in the inadequate to minimal range (1 to 3), while about 34 percent of both public school

and private program classrooms scored in the good to excellent range (5 to 7). See Figure 7.

**Percentage of Classrooms Scoring 1-7
on the Provisions for Learning**

Figure 7. Provisions for Learning Public School Vs Private Program

Public ■ Private
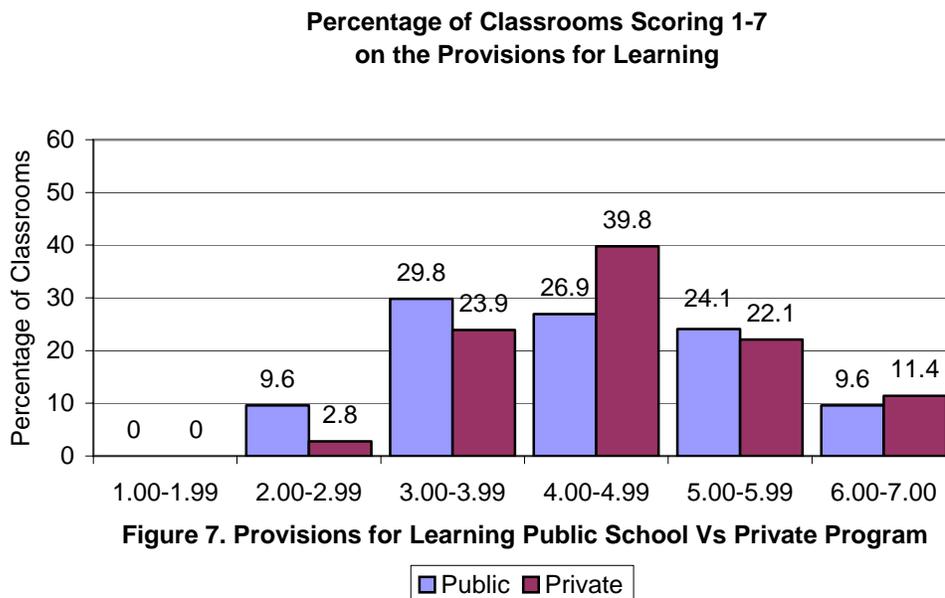
The second factor, labeled Teaching and Interaction, included 9 items and accounted for 15 percent of the variance. The Cronbach's alpha for this factor is .84. Included items are general supervision of children, informal use of language, staff-child interactions, interactions among children, discipline, supervision of gross-motor activities, using language to develop reasoning, encouraging children to communicate, and group time. The mean across classrooms on this factor was 5.63 (SD = .98). See Figure 8 below for the distribution of scores for this factor.

**Percentage of Classrooms Scoring 1-7
on the Teaching and Interaction**

Figure 8. ECERS-R Factor 2: Teaching and Interaction
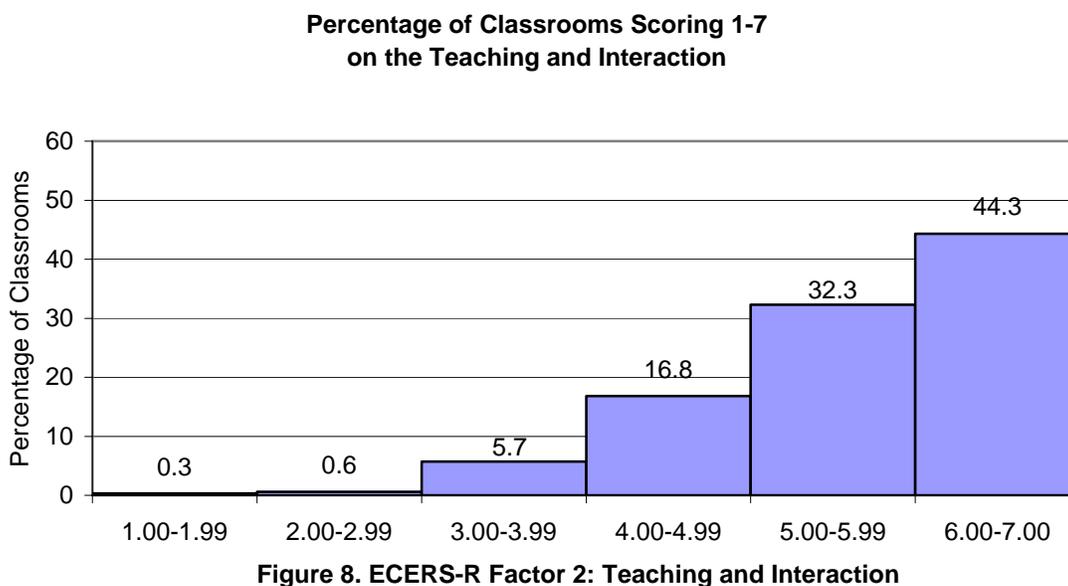
The average scores for the Teaching and Interaction factor for the public school and private program are 5.56 (SD = 1.04), and 5.67 (SD = .95), respectively. Only about 3 percent of the public and none of the private program classrooms score in the inadequate to minimal range (1 to 3), while about 76 percent of public school and about 77 percent of private program classrooms score in the good to excellent range (5 to 7). See Figure 9.

**Percentage of Classrooms Scoring 1-7
on the Teaching and Interaction**

Figure 9. Teaching and Interaction Public School Vs Private Program
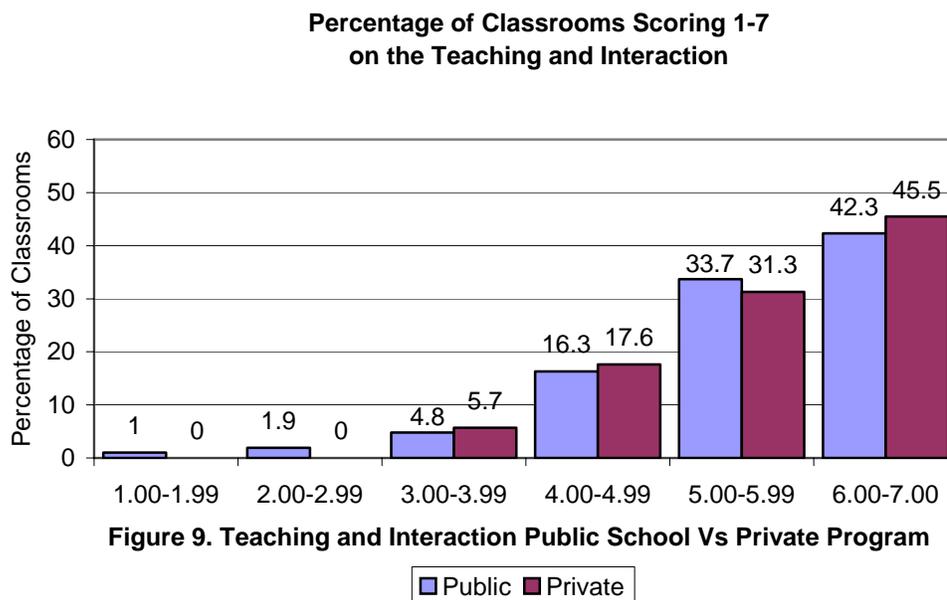
□ Public ■ Private

Table 4 below reports the ECERS-R Provisions for Learning and Teaching and Interaction factor scores for the total sample, and for the public school and private programs. Independent sample T-tests found no statistically significant differences in Provisions for Learning and Teaching and Interaction between public schools and private programs.

Table 4. ECERS-R Two Factor Scores

ECERS-R Subscale	Total * (*N*=316)		Public School (*n*=104)	Private Program (*n*=176)
	M (*SD*)	Range	*M* (*SD*)	*M* (*SD*)
Provisions for Learning	4.46 (1.01)	2.08-7.00	4.42 (1.06)	4.60 (.97)
Teaching and Interaction	5.63 (.98)	1.67-7.00	5.56 (1.04)	5.67 (.95)

* Total includes Head Start classrooms and self-contained special education preschool classrooms.

Supports for Early Literacy Assessment (SELA)

The SELA measures the classroom environment and teaching practices that lead to early literacy and language development. The average total SELA score across 316 classrooms was 3.46 with a standard deviation of .63 for the total sample, 3.55 (SD = .66) for public schools and 3.41 (SD = .59) for private programs. On a scale of 1 to 5, 1 representing very low quality and 5 representing high quality, or the ideal, this score indicates that most Abbott preschool classrooms can be characterized as providing good support for children's language and literacy development. Scores ranged from 1 to 5, with about 22 percent of classrooms at or near the ideal (a score of 4 to 5), over 75 percent scoring above the mid-point and less than 1 percent scoring in the low to poor quality range (a score of 1 to 2). See Figure 10 below.

Percentage of Classrooms Scoring 1-5 on the SELA
**2005 - 2006**

Figure 10. SELA Score
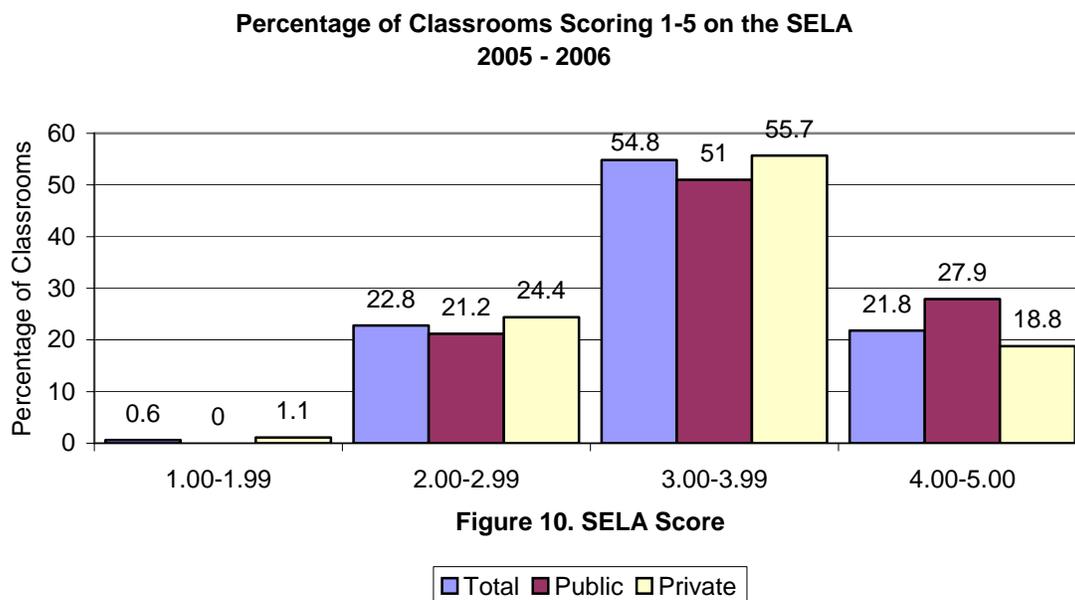
Table 5 below lists the SELA average item scores and the average total scores in 2005-2006. For the total sample, the three highest scoring items were for "Creating inviting places to look at books" (M = 4.34, SD = .79), "Writing materials are available and easy to use" (M = 4.31, SD = .80), and "Sharing books to build language" (M = 4.23, SD = .84). These items scored 4 or more indicating near the ideal. For these items, nearly 50 percent of classrooms score at the ideal. The lowest scoring item was "Drawing children's attention to the sounds they hear in words," with an average score of 2.40 (SD = 1.32). This item measures the extent to which the teacher draws attention to the sounds that children hear in words. About 34 percent of sample classrooms score a 1 on this item, while about 11 percent scored at the ideal.

For the public and private classrooms, the three highest scoring items were the same as the total sample. For private programs, one more item, "Using print for purpose," is near the ideal (M = 4.02, SD = .85). Independent sample T-tests revealed that public school classrooms scored significantly higher than private program classrooms on 6 of the 16 SELA items (p < .05), including functions and features of print, phonological awareness, helping children recognize letters, promoting children's interest in writing, promoting

home-based support, and activities to involve parents in children's literacy. T-tests also showed that private programs scored significantly higher on activities promoting oral language development than public schools did.

Table 5. SELA Scores Across 30 Abbott Districts, 2005-2006

SELA Item	Total (*N*=316)		Public School (*n*=104)	Private Program (*n*=176)
	M (*SD*)	Range	*M* (*SD*)	*M* (*SD*)
Using print for a purpose	3.97 (.89)	1.0 - 5.0	3.89 (.97)	4.02 (.85)
Creating inviting places to look at books	4.34 (.79)	1.0 - 5.0	4.33 (.79)	4.42 (.72)
Inviting interest in a wide variety of books	3.91 (.95)	1.0 - 5.0	3.86 (1.01)	3.99 (.93)
Writing materials are available and easy to use	4.31 (.80)	1.0 - 5.0	4.38 (.78)	4.30 (.83)
Literacy items and props in pretend area	3.35 (1.03)	1.0 - 5.0	3.48 (1.01)	3.35 (1.04)
Encouraging children using oral language	3.68 (1.10)	1.0 - 5.0	3.73 (1.10)	3.67 (1.06)
Introduce new words	2.99 (1.13)	1.0 - 5.0	3.09 (1.18)	2.94 (1.08)
Activities promoting oral language development	3.72 (1.03)	1.0 - 5.0	3.53 (.98)	3.81 (.99)*
Sharing books	4.23 (.84)	1.0 - 5.0	4.30 (.82)	4.19 (.87)
Functions and features of print	3.20 (1.14)	1.0 - 5.0	3.38 (1.20)	3.07 (1.11)*
Phonological awareness	2.40 (1.32)	1.0 - 5.0	2.73 (1.35)	2.27 (1.32)*
Helping children recognize letters	3.01 (1.17)	1.0 - 5.0	3.22 (1.17)	2.85 (1.16)*
Promoting children's interest in writing	3.27 (1.18)	1.0 - 5.0	3.53 (1.14)	3.13 (1.18)*
Promoting home based support	3.24 (1.05)	1.0 - 5.0	3.41 (.97)	3.11 (1.02)*
Activities to involve parent in children's literacy	2.69 (1.14)	1.0 - 5.0	2.97 (1.19)	2.51 (1.03)*
Promoting native language	2.95 (1.25)	1.0 - 5.0	2.86 (1.28)	2.96 (1.24)
Overall SELA Score	3.46 (.63)	1.63-4.81	3.55 (.66)	3.41 (.59)

* p < .05
** Total includes Head Start classrooms and self-contained special education preschool classrooms.

Preschool Classroom Mathematics Inventory (PCMI)

Classroom support for the development of children's early mathematical skills is measured using the Preschool Classroom Mathematics Inventory (PCMI; Frede, Weber, Hornbeck, Stevenson-Boyd & Colon, 2005), and is comprised of 11 items on a 5-point scale, from low quality (1) to high quality (5). The average PCMI total scale score across the 316 sample classrooms was 2.29 (*SD* = .58), indicating that the average Abbott preschool provides limited support for children's mathematical skill development. The average PCMI score for public schools was 2.37 (*SD* = .60) and for private programs was 2.29 (*SD* = .57). The vast majority of classrooms scored between 1 and 3 on this scale. See Figure 11 and Table 6 for distribution of scores. Independent sample T-tests found that public schools scored higher on one PCMI item "Teachers encourage counting" compared to private programs (*p* < .05). No other statistically significant differences were found on the PCMI scores.

Percentage of Classrooms Scoring 1-5 on the PCMI
Spring 2006

Figure 11. PCMI Score
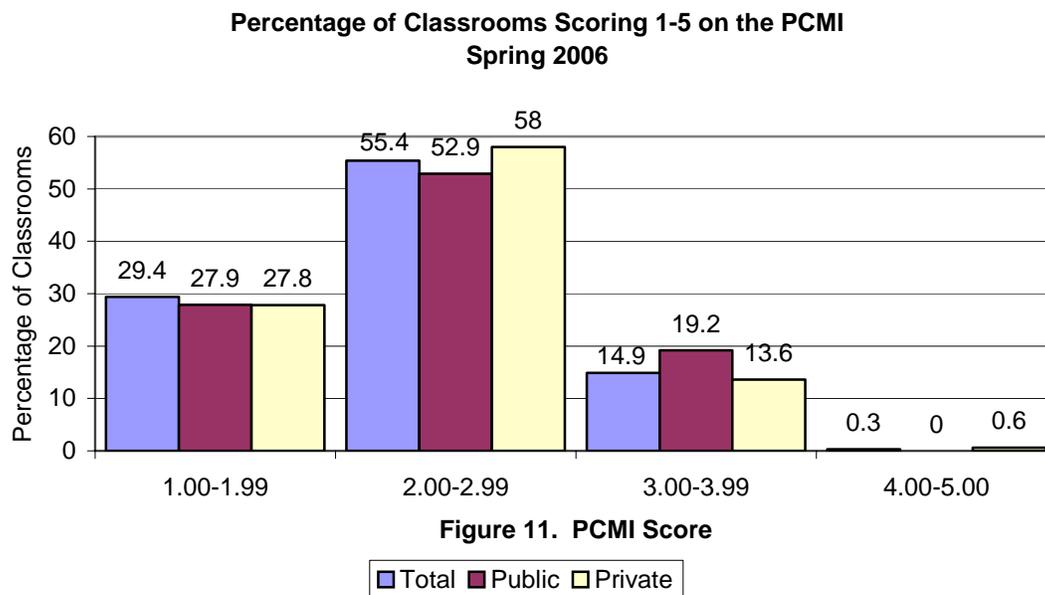
Table 6. PCMI Scores Across 30 Abbott Districts, 2005-2006

PCMI Item	Total (N = 316)		Public School (n = 104)	Private Program (n = 176)
	M (*SD*)	Range	*M* (*SD*)	*M* (*SD*)
Materials for counting	3.81 (.93)	2.0 - 5.0	3.88 (.97)	3.82 (.89)
Materials for measuring	3.28 (1.03)	1.0 - 5.0	3.40 (1.06)	3.32 (.99)
Materials for classifying	2.91 (.96)	1.0 - 5.0	2.89 (.93)	3.01 (.93)
Materials for geometry	3.46 (.94)	1.0 - 5.0	3.52 (.92)	3.47 (.93)
Teachers encourage one-to-one correspondence	1.70 (.94)	1.0 - 5.0	1.63 (.96)	1.76 (.97)
Teachers encourage counting	2.30 (.97)	1.0 - 5.0	2.48 (.99)	2.22 (.95)*
Teachers encourage estimation	1.49 (.87)	1.0 - 5.0	1.60 (.94)	1.47 (.87)
Teachers use math terminology	1.66 (.88)	1.0 - 5.0	1.75 (.95)	1.59 (.84)
Teachers measure and compare	1.48 (.89)	1.0 - 5.0	1.63 (.95)	1.44 (.86)
Teachers encourage classification	1.53 (.82)	1.0 - 5.0	1.63 (.88)	1.49 (.81)
Teachers encourage geometry	1.60 (.84)	1.0 - 5.0	1.68 (.92)	1.56 (.81)
Overall PCMI score	2.29 (.58)	1.18-4.09	2.37 (.60)	2.29 (.57)

* p < .05
** Total includes Head Start classrooms and self-contained special education preschool classrooms.

Analysis of Classroom Observation Scores by Age of Children in the Classroom

Across Abbott districts some classrooms are designed to serve only 4-year-olds the year before kindergarten or only the 3-year-olds two years before kindergarten. Some are designed for mixed-age groups. Mixed-age grouping occurs not only by design but also because of space issues. The classroom quality data also were analyzed by age group of the classroom. Most comparisons showed no statistically significant differences by age of children. However, the following differences were found:

- Three-year-old classrooms scored significantly better than mixed-age classrooms on three ECERS-R subscales (Interactions, Program Structure and Teaching and Interaction) and on the SELA item "Helping children recognize letters."

- On SELA, 4-year-old classrooms out-performed both other age classrooms on six items: "Sharing books to build language, knowledge, and a love of book-reading," "Calling attention to the functions and features of print," "Drawing children's attention to the sounds they hear in words," "Helping children recognize letters," "Promoting children's interest in writing," and "Activities to involve parent in children's literacy."
- On PCMI, the 4-year-old classrooms scored higher on "Teachers encourage children to count and/or write numbers for a purpose."

**Research Methods and Results for the
Investigation of Effects on Child Outcomes**

Beginning in the fall of 2005, researchers from NIEER with the help of their ELIC partners and with partial funding from The Pew Charitable Trusts, designed and implemented a two-step research process to determine the long-term effects of attendance in an Abbott preschool classroom. The first step was to conduct a study using regression discontinuity design (RDD) to estimate the effects of the program on children's abilities at kindergarten entry. The RDD approach produces an estimate that is more likely to be free of selection bias and, thus, is less likely to underestimate the program's effects. This estimate is then compared with differences between two groups of kindergarten children – one group attended preschool, the other did not. If the estimates in both analyses are similar, then we have confidence that the traditional treatment vs. control sample is not biased. Thus, the purpose of the design is to determine what the short-term, yearly, and long-term effects of the Abbott Preschool program are on children's academic skills from early in kindergarten through fourth grade and whether these children are less likely to have been retained in grade or placed in Special Education than children who did not attend. Because some children attended preschool for one year at age 4 and others attended preschool for two years at ages 3 and 4, we are able to estimate the effects of one year versus two years of preschool attendance. In later analyses, family characteristics, such as mother's education level, language spoken in the home, and a measure of family resources, will be included to investigate their influence on the effects of the preschool program.

The Research Model

The APPLES employed an RDD coupled with a longitudinal cross-sectional design. This combination has several strengths. Typically, state preschool program evaluations estimate program effects with a longitudinal cross-sectional design, comparing the test scores of children who attended a program with similar children who did not attend. As programs move toward offering all children services, it can be very difficult to find a comparable group of children who did not go to preschool. Even when programs target a special subgroup of children (e.g., low-income children or those with learning delays), a problem remains: those who are eligible for the program but who do not attend are *not* the same even before the program begins. Differences chiefly arise because of the differences between families, and possibly children, that are inherent when some parents choose to enroll their children and others do not. When samples are chosen after the program ends other differences are possible – for example, some children who did not attend the program will have moved into the district.

The RDD solution is to compare two groups of children who selected and were selected by the Abbott program, using the stringent age cutoff for enrollment eligibility to define groups. This concept is easier to understand when taking the extreme case: consider two children who differ only in that one was born the day before the age cutoff and the other the day after. When both are about to turn 5 years old the slightly younger child will enter the preschool program and the slightly older child will enter kindergarten

having already completed the preschool program. If both are tested at that time, the difference in their scores provides an unbiased estimate of the state preschool program's effect, under reasonable assumptions. Obviously, if only children with birthdays one day on either side of the age cutoff were included in a study, the sample size would be unreasonably small. However, the approach can be applied to wider age ranges around the cutoff. In fact, all children entering kindergarten from the Abbott preschool program, and all children beginning preschool in the same year can be included in analyses using the RDD.

The APPLES also involves an added longitudinal component using the more typical approach of comparing children in the same age cohort who did and did not attend an Abbott preschool program. By following this second sample of children across five years, we can estimate the impact of the program on children's learning through age 8. One of the keys to success in this second effort is the ability to check its estimates against those from the RDD approach and select the best analytical model based on that comparison. Thus, this is the first of a series of reports detailing the estimated effects of the Abbott Preschool program. Preliminary findings from the RDD and longitudinal analyses of Fall 2005 and Spring 2006 data are presented in the rest of this report.

Sampling Strategy

To select the sample of children, we use a methodology that has been effective in five other statewide evaluations (Lamy, Barnett, and Jung, 2005). We first gathered state-level and district-level information on the programs, including the location and number of programs, program type, the number of children attending the program, and number of classrooms. We then randomly selected the total number of Abbott-funded preschool classrooms from the 15 largest Abbott-funded districts. We used the largest 15 districts under the assumption that if the program were effective there, it would be effective in the remaining 16 smaller districts. Previous analyses have shown that quality and children's scores at kindergarten entry are higher on average in the smaller districts. From this list of the universe of programs, individual classrooms were randomly sampled. From each of the randomly sampled classrooms, approximately four children were selected. The kindergarten sample was selected without consideration of preschool participation thus ensuring that a proportionally appropriate number of children would not have attended Abbott preschool. These children would form the comparison group for the longitudinal study discussed in the next section.

Trained research staff from NIEER, William Paterson University, and The College of New Jersey visited each sampled program site, selected children into the sample using a procedure to ensure randomness, and conducted the child assessments as early as possible in the school year. A liaison at each site gathered information on the children's preschool status, usually from existing school records but occasionally from parent report, and was reimbursed $5.00 per selected child.

The Sample

As mentioned above, our RDD methodology requires two groups of children. The group of kindergartners who attended the Abbott preschool program the previous year is called the *Preschool* group or the experimental group. Children who received some form of early care other than the Abbott preschool program at age 4 were not included in this group. The second group of 4-year-old children attending the Abbott preschool program is called the *No Preschool* group, or the control group. This group is called the *No Preschool* group despite the fact that they are currently enrolled in the state-funded preschool program because they are at the very beginning of their preschool year and have not had the preschool "treatment" yet.

Data was gathered from 563 classrooms, with an average of 4.18 children per class. The total New Jersey sample size was 2,356 children. For the RDD there were 766 kindergarten children who attended Abbott preschool and 778 4-year-old children in Abbott preschool in 2005-2006. For the longitudinal design sample the same 766 kindergarteners who attended Abbott preschool are the treatment group and the 246 kindergarten children who did not attend any preschool are the comparison group. See Table 7 for further breakdown of the kindergarten sample.

Table 7. Preschool Attendance of Kindergarten Sample

Number of Years	**Any preschool program**	**Abbott preschool program only**
Preschool at 4 but not at 3 (one year)	498	461
Preschool at 3 and 4 (two years)	327	305
Total who attended preschool	825	766
No preschool	246	
Total Sample	1,071	

The longitudinal design plans to follow the sample children across five years. The initial longitudinal sample is comprised of kindergarten children in the 15 largest Abbott districts. The total sample size for this study is 1,071 kindergarten, 766 kindergarten children who attended Abbott pre-K and 246 kindergarten children who did not attend any pre-K program. Table 8 depicts the gender and ethnicity of the longitudinal sample by preschool attendance. There were no statistically significant differences among the groups.

Table 8. Longitudinal Sample Demographic Characteristics at Kindergarten Entry

	Total	**No Pre-K**	**1 Year of Abbott Pre-K**	**2 Year of Abbott Pre-K**
<u>N</u>	1,071	246	461	305
African American	40.4%	40.3%	41.9%	38.4%
Hispanic	51.4%	50.8%	49.2%	55.4%
White/ Other	8.1%	8.8%	8.9%	6.2%
Female	49%	45.2%	50.1%	50.8%

<u>Family Characteristics</u>

Data on the family characteristics of the sample children will be included as covariates in later analyses. Family characteristics data have been collected through short family interviews done by phone. Contact information for the families of the sample children were collected from school records. Other family data included maternal education level, primary language spoken in the home, number of siblings, and family income level. Preschool attendance status was also confirmed.

Instrumentation

Instrumentation for preschoolers and kindergartners in the first year of the study was identical. To allow for longitudinal comparison, instrumentation will remain consistent across all the years of the study. The battery of child assessments took an average of approximately 25 minutes per child and took place at the child's school program, in a room or area appropriate for testing. Lastly, children's special education and grade retention status will be monitored across the years of the study.

Receptive Vocabulary

Children's receptive vocabulary was measured using the Peabody Picture Vocabulary Test, 3rd Edition (PPVT-III; Dunn & Dunn, 1997) and for Spanish-speakers the *Test de Vocabulario en Imagenes Peabody* (TVIP; Dunn, Padilla, Lugo, & Dunn, 1986). The PPVT is predictive of general cognitive abilities and is a direct measure of vocabulary size. The rank order of item difficulties is highly correlated with the frequency with which words are used in spoken and written language. The test is adaptive (to avoid floor and ceiling problems), establishing a floor below which the child is assumed to know all the answers and a ceiling above which the child is assumed to know none of the answers. Reliability is good as judged by either split-half reliabilities or test-retest reliabilities. The TVIP is appropriate for measuring growth in Spanish vocabulary for bilingual students and for monolingual Spanish speakers.

All children in our sample were administered the PPVT, regardless of home language, to get some sense of their receptive vocabulary ability in English. All children who spoke some Spanish were also subsequently administered the TVIP. The testing session was then continued, with the additional measures administered in either English or Spanish, depending upon what the child's teacher designated as his or her best testing language. When running preliminary analyses, if there was a case where a child scored better on the TVIP than on the PPVT, but the assessor had continued testing in English (or vice versa), we excluded that case from the analyses.

Mathematical Skills

Children's early mathematical skills were measured with the Woodcock-Johnson Tests of Achievement, 3rd Edition (Woodcock, McGrew, & Mather, 2001) Subtest 10 Applied Problems. For Spanish-speakers the *Bateria Woodcock-Munoz Pruebas de Aprovechamiento – Revisado* (Woodcock & Munoz, 1990) *Prueba 25 Problemas Aplicados* was used. Subtests of the Woodcock-Johnson are reported to have good reliability. Raw scores are reported.

Print Awareness

Print awareness was measured using the Print Awareness subtest of the Preschool Comprehensive Test of Phonological and Print Processing (Pre-CTOPPP; Lonigan, Wagner, Torgeson, & Rashotte, 2002). The Pre-CTOPPP was designed as a downward extension of the Comprehensive Test of Phonological Processing (CTOPP; Wagner, Torgeson, & Rashotte, 1999), which measures phonological sensitivity in elementary school-aged children. Although not yet published, the Pre-CTOPPP has been used with middle-class and low-income samples and includes a Spanish version. As the Pre-CTOPP was developed recently, relatively little technical information is available about its performance and psychometric properties. Print Awareness items measure whether children recognize individual letters and letter-sound correspondences and whether they

differentiate words in print from pictures and other symbols. The percentage of items answered correctly out of the 36 total subtest items is reported.

The skills and knowledge measured by the Pre-CTOPPP (which are predictive of later literacy ability) are expected to be present by the end of the preschool year. Thus, it is not an appropriate test for the end of kindergarten and results are not reported past kindergarten entry. In later years, full batteries for reading will be administered.

Results

Regression Discontinuity Design. The estimated effects of the Abbott program are displayed in individual figures for each outcome measure. Each figure displays a regression line of the children's predicted test scores by age, measured by the number of days their birth date is from the program enrollment cut-off date. The discontinuity (gap or jump) in the regression line at the cut-off date is the estimated effect of the Abbott program. The estimated regression lines control for ethnic background, gender, and age, and take into account the effects of clustering by classroom in the sample.

Receptive Vocabulary

The estimated effect of state-funded preschool on children's receptive vocabulary as measured by the PPVT is statistically significant. Attending the Abbott program at age 4 is estimated to increase PPVT scores by about 4.57 raw score points. This represents an improvement of about 28 percent of the standard deviation for the control (*No Preschool*) group. The effects of the program can also be understood as 34.8 percent more growth over the year in children's average vocabulary scores. Age equivalence scores provide a measure of children's vocabulary knowledge using a normed estimate of the average age of children who score the same. Results indicate that the average improvement due to the Abbott program is approximately an additional 4.17 months of vocabulary development.

Figure 12. The Effect of the Abbott Program on Children's Receptive Vocabulary Scores
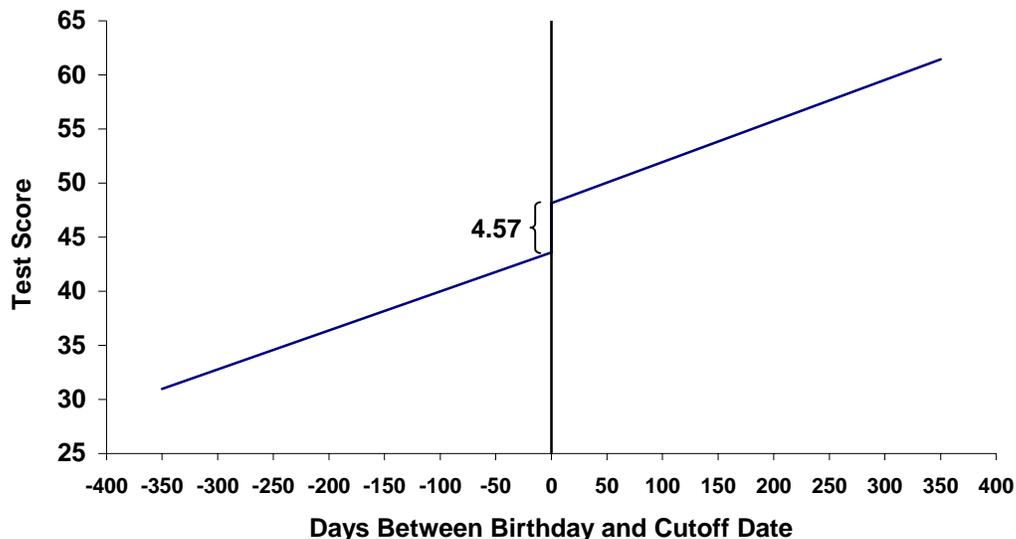
Figure 12 portrays a regression line of the children's predicted PPVT scores by the distance in days their birth date is from the program enrollment cut-off date. The discontinuity in the regression line at the cut-off date represents the estimated effect of the preschool program and is estimated to equal 4.57 raw score points.

<u>Math Skills</u>

The estimated effect of state-funded preschool on children's early math skills as measured by the Woodcock-Johnson-III Applied Problems subtest scores is statistically significant for the Abbott program. The increase in scores for Abbott children due to the program is about 1.36 raw score points. This represents an improvement of about 36 percent of the standard deviation for the control (*No Preschool*) group. The effects of the program can also be understood as 41.4 percent more growth over the year in children's average math scores.

Figure 13 below portrays a regression line of the children's predicted Applied Problems scores by the distance in days their birth date is from the program enrollment cut-off date. The discontinuity in the regression line at the cut-off date represents the estimated effect of the preschool program and is 1.36 raw score points.

Figure 13. The Effect of the Abbott program on Children's Early Math Scores
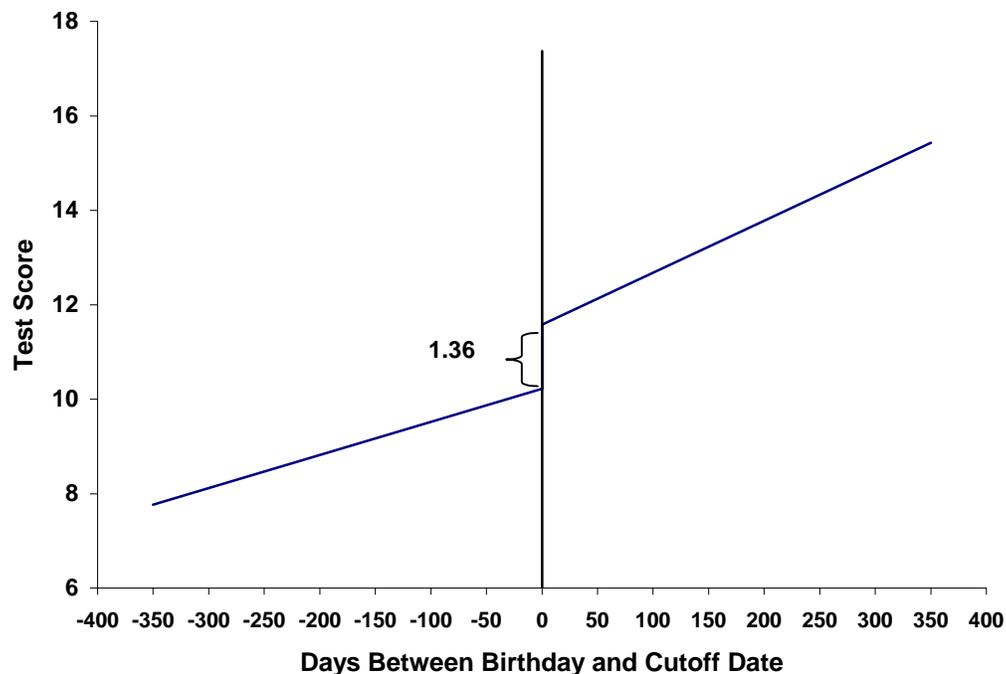
Print Awareness

The effect of state-funded preschool on children's Print Awareness scores is statistically significant for the Abbott program. The effect of the Abbott program on children's gains in Print Awareness scores is 14 percent more items answered correctly. This increase represents and improvement of approximately 56 percent of the control (*No Preschool*) group standard deviation on the Print Awareness subtest. The effect of the program can also be understood as 96 percent more growth over the year in children's average print awareness scores.

Figure 14 portrays a regression line of the children's predicted Print Awareness scores by the distance in days their birth date is from the program enrollment cut-off date. The discontinuity in the regression line at the cut-off date represents the estimated effect of the preschool program and is 14 percent more items answered correctly.

Figure 14. The Effect of the Abbott Program on Children's Print Awareness Scores
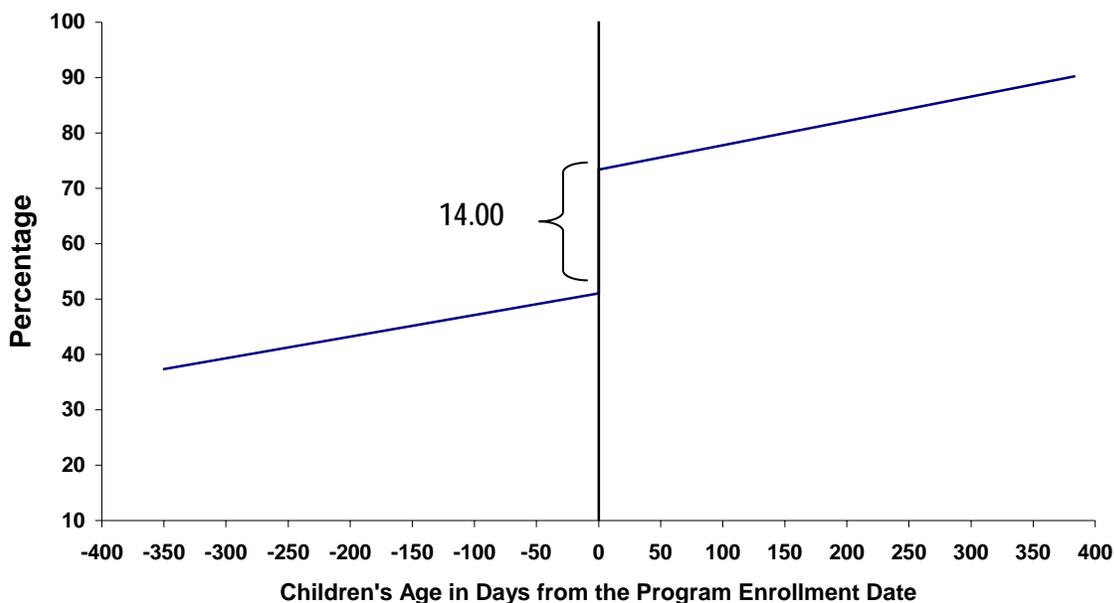
Longitudinal Study. Regression analyses of the longitudinal sample control for student ethnicity, gender, age, and school district. The effects of cluster sampling in kindergarten are taken into account. The results of analyses conducted on the longitudinal sample at kindergarten entry and at the end of kindergarten are reported below in the text and in tables that report scores for each group: no preschool, one year of preschool at age 4, and two years of preschool at ages 3 and 4.

Receptive Vocabulary

At the beginning of the kindergarten year, the estimated effect of state-funded preschool on children's receptive vocabulary as measured by PPVT is statistically significant. Attending one year of the Abbott preschool program at age 4 is estimated to increase PPVT scores by 3.82 standardized score points. This represents an improvement of about 21.4 percent of the standard deviation for the control group. Attending two years of the Abbott preschool program at ages 3 and 4 is estimated to increase PPVT scores by 7.41 standardized score points. This improvement is about 41.6 percent of the standard deviation for the control group. Figure 15 portrays these differences in children's gains in receptive vocabulary at the beginning of the kindergarten year.

Figure 15. Longitudinal Study Receptive Vocabulary at Kindergarten Entry by Years of Attendance (*N*=1,038)
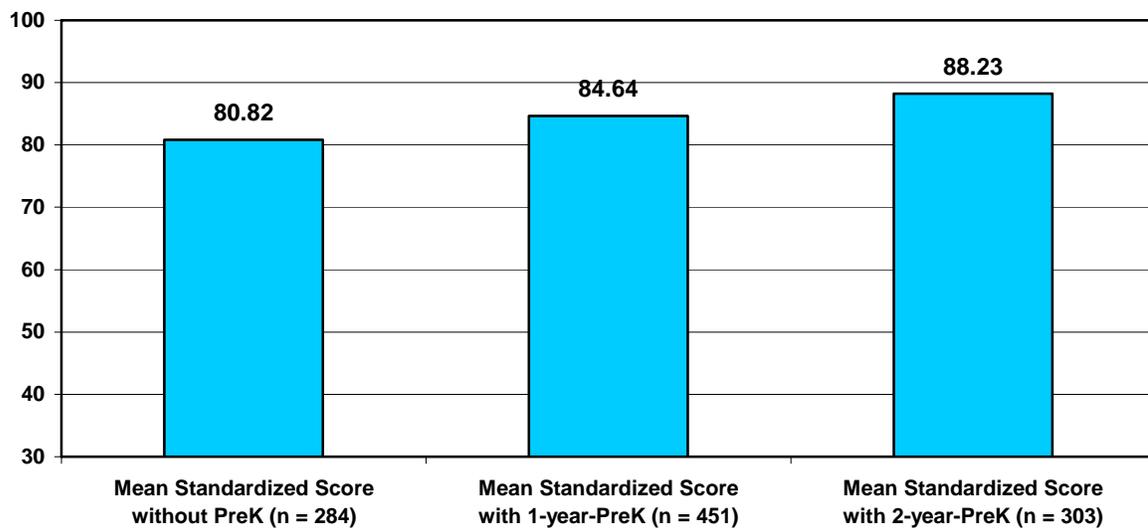
At the end of the kindergarten year, the estimated effect on children's receptive vocabulary continues to be statistically significant. Attending one year of the Abbott preschool program at age 4 is estimated to increase PPVT scores by 3.39 standardized score points. This represents an improvement of about 22.3 percent of the standard deviation for the control group. Attending two years of the Abbott preschool program at ages 3 and 4 is estimated to increase PPVT scores by 6.24 standardized score points, an

improvement of 41.1 percent of the standard deviation. Figure 16 portrays the difference in children's gains in receptive vocabulary at the end of the kindergarten year.

Figure 16. Longitudinal Sample Receptive Vocabulary at the End of Kindergarten by Years of Attendance (*N*=974)
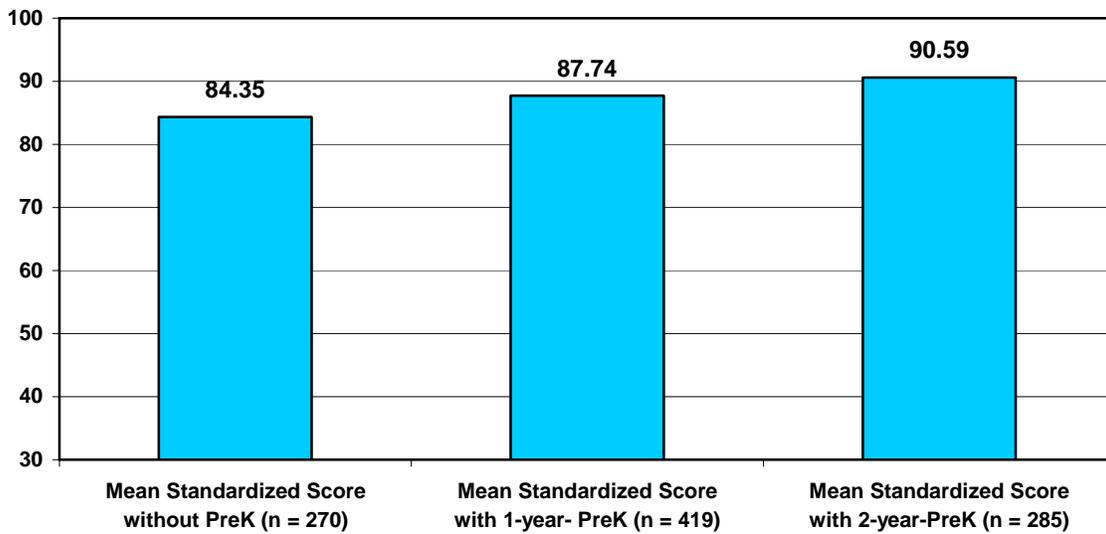
In order to compare the RDD and longitudinal results, effects of one and two years of Abbott preschool education on the PPVT in kindergarten entry in the longitudinal sample also were analyzed using raw scores. (The RDD approach is only properly applied to raw scores, as the standardized scores are already age-adjusted, whereas the RDD analysis relies on estimating the relationship between test score and age.) Attending one year of the Abbott preschool program at age 4 was estimated to increase PPVT scores by 4.06 raw score points. This effect is slightly smaller (11 percent) than the estimated 4.57 point gain found using the RDD approach.

Mathematical Skills

At the beginning of the kindergarten year, the effect of state-funded preschool on children's early math skills as measured by the Woodcock-Johnson-III Applied problems subtest is statistically significant. The improvement in children's scores due to one year of attendance at age four is about .86 raw score points. This represents an improvement of about 19.7 percent of the standard deviation for the control (*No Preschool*) group. The improvement in children's scores due to two years of attendance at ages 3 and 4 is about 1.47 raw score points. This is about 33.6 percent of the standard deviation for the control (*No Preschool*) group. Figure 17 below portrays the estimated impact of state-funded preschool programs on children's math scores at the beginning of the kindergarten year.

At the end of the kindergarten year, the effect on children's early math skills continues to be statistically significant. The improvement in children's scores due to one year at age 4 is about .61 raw score points. This represents an improvement of about 12.8 percent of the standard deviation for the control group. The estimated improvement in children's scores due to two years attendance at ages 3 and 4 is about 1.38 raw score points. This represents about 29 percent of the standard deviation for the control group. Figure 18 portrays the estimated impact of state-funded preschool programs on children's math score gains at the end of the kindergarten year.

Figure 17. Longitudinal Study Mathematics Scores at Kindergarten Entry by Years of Attendance (*N*=1,054)
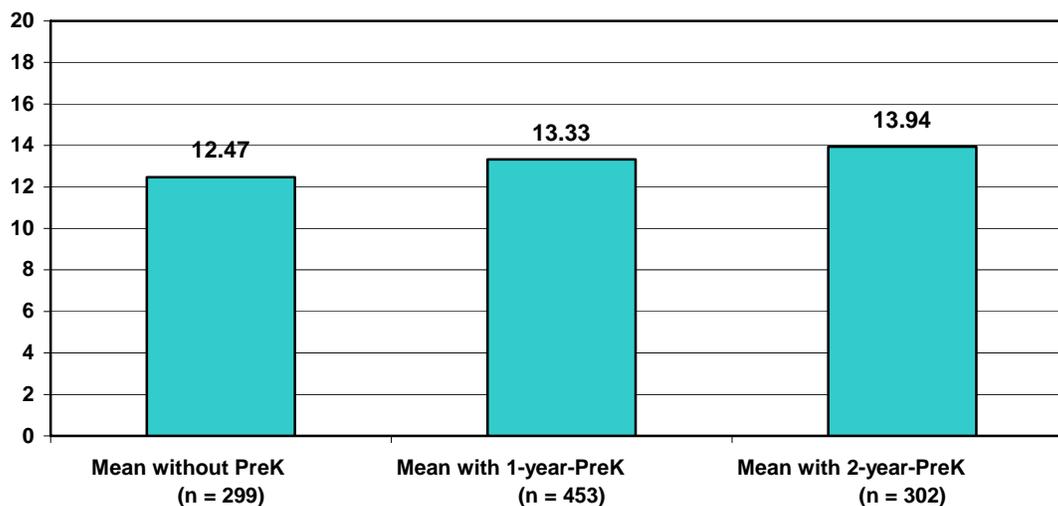
Figure 18. Longitudinal Study Mathematics Scores at the End of Kindergarten by Years of Attendance (*N*=973)
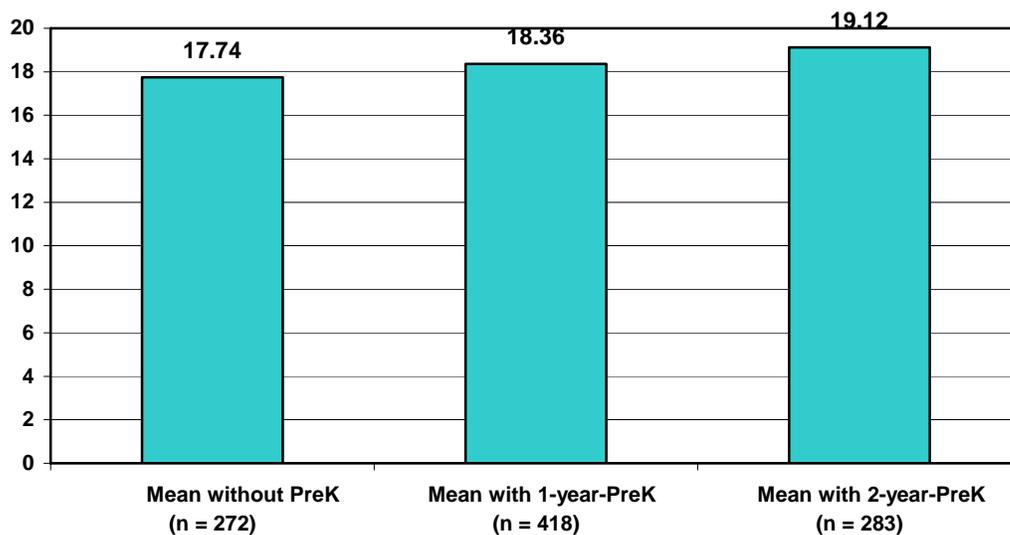
Print Awareness

At the beginning of the kindergarten year, the effect of state-funded preschool on children's Print Awareness scores is statistically significant. The effect of one year of attendance on children's gains in Print Awareness scores is 7.8 percent more items answered correctly. This increase represents an improvement of approximately 29.1 percent of the control group standard deviation on the Print Awareness subtest. The effect of two years of attendance at ages 3 and 4 on children's gains in Print Awareness scores is 8.4 percent more items answered correctly. This increase represents approximately 31.4 percent of the control group standard deviation on the Print Awareness subtest. Figure 19 depicts the estimated impact of the Abbott preschool programs on children's Print Awareness at the beginning of the kindergarten year.

At the end of kindergarten, the effect on children's Print Awareness scores continues to be statistically significant for children who had two years of Abbott preschool. The effect of one year of attendance on children's gains in Print Awareness scores is 1.41 percent more items answered correctly. This increase represents and improvement of approximately 10.5 percent of the control group standard deviation on the Print Awareness subtest but this difference is not statistically significant. The effect of two years of the attendance on children's gains in Print Awareness scores is 1.88 percent more items answered correctly. This increase represents approximately 14 percent of the control group standard deviation on the Print Awareness subtest. Figure 20 shows the estimated effect of the Abbott preschool programs on children's Print Awareness at the end of the kindergarten year.

Figure 19. Longitudinal Study Print Awareness Scores at the Kindergarten Entry by Years of Attendance (% correct) (*N*=1,053*)
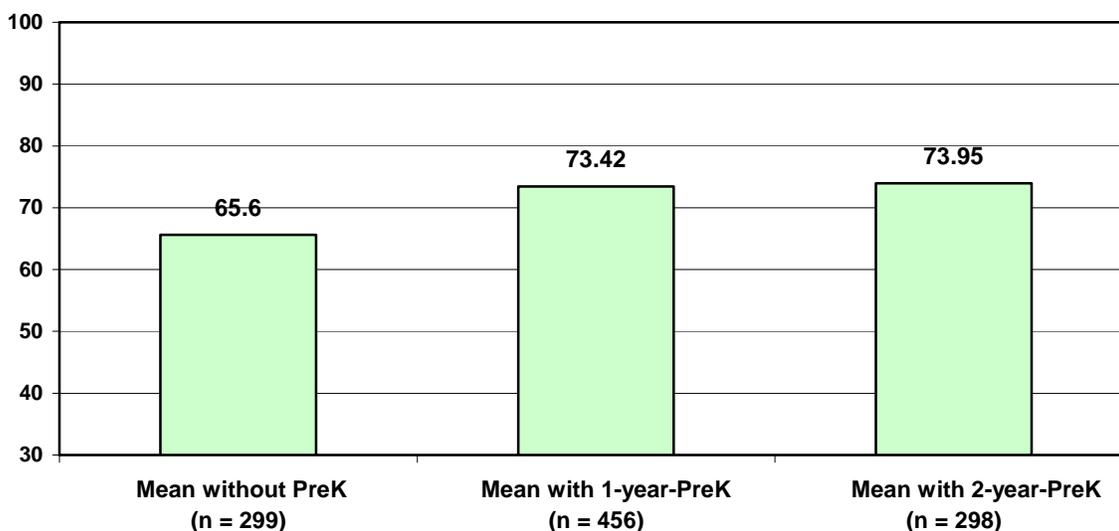
**Figure 20. Longitudinal Study Print Awareness Scores at the End of Kindergarten by Years of Attendance (% correct) (*N=979)*
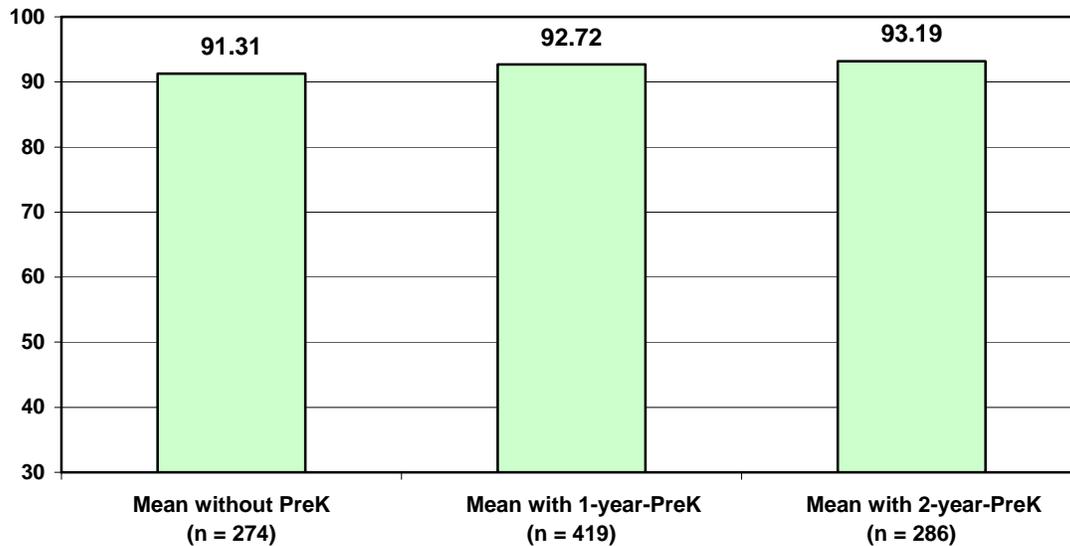
Summary and Discussion

Considerable attention and resources have been invested in the Abbott Preschool Program. According to NIEER's annual report on preschool education, the Abbott program has among the highest quality standards of any state preschool program in the nation, ranks first in providing access to 3-year-olds, and is the most well-funded (Barnett, Hustedt, Hawkinson and Robin, 2006). Thus, there is a great deal of interest in how effective Abbott classrooms are in helping children enter kindergarten with the knowledge, skills and dispositions that will lead to success in school. The results presented here provide clear evidence that by participating in a high-quality program regardless of auspice, children are improving in literacy and math at least until the end of the kindergarten year.

Classroom Quality

Taken as a whole, the advances in classroom quality scores are notable. In 2000, the average ECERS-R score was 3.86. In 2006, almost 90 percent of the classrooms score above that mean for 2000. The average score of 4.81 is similar to that found in other studies of publicly funded preschool in this country (Early et. al., 2007). In those areas most likely to be directly related to child learning – Language and Reasoning, Activities, Interactions, and Program Structure – classrooms on average scored in the good to excellent range. Scores on the Teaching and Interaction factor are quite high.

Average scores on SELA also reflect practices that are likely to lead to more learning, with the highest scores in supplying materials that support language and literacy

development and in teaching practices that enhance oral language development. Fully, 75 percent score a 3 or better. However, in the special case of language and literacy, the lower scoring items are mostly related to specific language and literacy skill development, including introducing new vocabulary, assisting children in developing print awareness and letter recognition, supporting phonological development (children's ability to hear the sounds in words) and promoting interest in writing. In addition, assisting parents in supporting their children's language and literacy development and supporting bilingual language development are also lower scoring items.

Results on PCMI, however, are not so heartening. The only scores above a 3 are on items that reflect the materials in the classroom. Given how well-funded the Abbott classrooms are, even these scores seem low and likely represent the same lack of understanding of mathematical learning and teaching that is clearly shown in the very low scores for teaching support. Six of the seven items that measure whether the teachers actively plan for and support mathematical learning have scores between 1 and 2. Thirty to 50 percent of the classrooms score a 1 on these items, meaning that no teaching is occurring to support these concepts. Clearly, math learning is enhanced when math is incorporated throughout the classroom activities (Arnold et al., 2002). However, a great deal of math reasoning is also constructed by the child while using math-related materials (Ginsburg, Inoue, & Seo, 1999) so the slightly better scores on mathematics materials are meaningful but not enough.

It should be noted that for some of the individual, low-scoring items on SELA and PCMI that measure fairly specific teaching strategies, it is difficult to know with current data whether it should be expected that teachers would use these techniques regularly in the 3.5-4 hour observation period. However, over the four years that these measures have been used in Abbott classrooms, there have always been a small percentage of classrooms that score above a 4 on the items and all classrooms have improved over time. This indicates that it is possible to meet the criteria and that professional development must continue to focus on these areas.

What is especially interesting in these findings is the fact that public school and child care center classrooms are now scoring the same across almost all measures of quality teaching practices. The items and subscales that show significant differences between auspice on ECERS-R and PCMI are unlikely to result in differences in child learning. It is difficult to know the impact of the six items from SELA on which the public school classrooms outperform the private classrooms. The differences are not large but as a whole these are the items most related to early literacy skill development. On the other hand, the private programs scored higher on promoting oral language and the same on items related to enjoyment of books. Oral language development and interactive book reading are important components of early literacy (Dickinson and Tabors, 2002) so the high scores in these areas are likely to profoundly influence children's learning.

Not surprisingly, we found that results on most of the basic "good" early childhood practices measured by ECERS-R did not vary by the age level of the classroom. The finding that classrooms designed to serve only 4-year-olds outscore the

3-year-old and mixed-age classrooms on some of the skills-based items on SELA and PCMI is not surprising since the introduction of these skills is more appropriate for the older children. These differences were determined with simple comparisons. Further investigation is needed if they remain after controlling for auspice, district, and teacher qualifications, especially given the fact that some districts design and plan their program to be mixed-age and in other locations it is a result of space issues where the teachers may not be expressly prepared for the two ages.

Another area where further analyses are needed is in determining if the ability of the teacher to speak Spanish influences the score on the SELA item Promoting native language acquisition. Given the larger number of Spanish-speaking teachers with private providers, it could be expected that the private provider classrooms would score significantly higher on this item. However, this difference between the auspices is not statistically significant.

Children's Outcomes in Kindergarten

Previous research indicated that the Abbott Preschool Program has beneficial effects on children's skills at kindergarten entry (Barnett, Jung, Lamy, Wang, and Cook, 2007; Frede et al, 2004; Lamy et al., 2005). The results presented here continue to provide strong evidence that the Abbott program has positive effects on children's learning in the areas of oral language, literacy and math skills and these positive effects are now found at the end of the kindergarten year. This signifies that Abbott produces the kinds of initial effects likely to lead to increased school success and continued advantages in reading and math skills. Meaningful effects were found on children's receptive vocabulary, math and print awareness skills. Children's early print awareness and receptive vocabulary skills have been found to predict later reading abilities in the early elementary grades (Snow, Burns, & Griffin, 1998). In addition, the effects found in this study are the first link in a chain that can produce the long-term school success and economic benefits found in other preschool education studies that have followed children into adulthood (Schweinhart et al., 2005; Campbell et al., 2002; Reynolds, Temple, Robertson, & Mann, 2002).

We conducted two separate studies of program effects at kindergarten entry to address concerns that the simple comparison of children who attended and did not attend Abbott preschool programs might be biased by unmeasured differences between the groups. This does seem to have been the case. The regression discontinuity design which attempts to control for these unmeasured differences provides estimates for the effects of one year of preschool education that are higher by 11 percent for language (PPVT), 37 percent for math, and 44 percent for print awareness. This indicates that the estimated effects in our longitudinal study underestimate the effects of preschool by meaningful amounts.

Thus, the longitudinal study addresses the question of the extent to which effects may fade out over time, but it must be understood that the longitudinal study somewhat underestimates the effects of the Abbott preschool program. At least for PPVT, the

underestimation appears to be fairly modest. Results of the study indicate that there are persistent effects on children's learning through the end of kindergarten, with only modest declines in the advantages from attending Abbott preschool programs for language and math. (Curiously, the largest decline was for the effect of one year of preschool education on math scores and the smallest decline for the effect of two years of preschool education on math scores.) Print awareness effects do not appear to last, apparently because nearly all children eventually acquire these skills during kindergarten. In this assessment at the end of kindergarten, we do not have measures of other aspects of literacy development beyond print awareness. It is highly likely that although children who attended preschool no longer outperform the non-preschool group on this set of early literacy skills, the fact that they acquired the skills earlier facilitated their development of subsequent skills which were not assessed in this battery. Tests at the end of first grade will include measures of literacy more broadly.

Very little research exists that compares the effects of one year versus two years of preschool attendance. Children who attended the Abbott Preschool Program for two years at ages 3 and 4 out-perform children who attended for only one year at age 4 and those who did not attend on all of the outcome measures with one exception. The gains in language and math from two years are quite large, nearly double for language and 70 percent larger for math. Children who had two years of preschool do not score significantly differently from those who had one year on the Print Awareness test. This is not a great surprise since this test is actually designed to assess preschool children's early literacy skills, and the majority of the children score well on it by the end of kindergarten. Caution must be used in interpreting these results. We cannot control for possible selection bias across the groups. Parents who know about and choose to send their children to preschool at age 3 may be different in immeasurable ways from those who only send them at 4. For this comparison we do not have the estimates from the more rigorous RDD to verify our results. The fact that this study is large scale and that it is fairly safe to assume that the quality of program for both years is similar adds to the importance of the findings.

This initial report, focusing on child data collected in the kindergarten year (Fall 2005 and Spring 2006) is the first in a series on the effects of the Abbott Preschool Program. Future reports will present results through the end of fourth grade and will include information on grade retention and special education placement as well as test results.

References

Arnold, D. H., Fischer, P. H., Doctoroff, G. L., & Dobbs, J. (2002). Accelerating math development in Head Start classrooms. *Journal of Educational Psychology, 94*, 762-770.

Barnett, W. S., Hustedt. J. T., Hawkinson, L. E., & Robin, K. B. (2006). *The state of preschool 2006: State preschool yearbook*. New Brunswick, NJ: National Institute for Early Education Research.

Barnett, W. S., Jung, K., Lamy, C., Wong, V., Cook, T. (2007, March). *Effects of five state prekindergarten programs on early learning*. Paper presented at the bi-annual Society for Research in Child Development, Boston, MA.

Barnett, W. S., Lamy, C., & Jung, K. (2005). *The effects of state prekindergarten programs on young children's school readiness in five states*. NIEER Policy Report. New Brunswick, NJ: National Institute for Early Education Research.

Barnett, W. S., Tarr, J., Esposito-Lamy, C., & Frede, E. (2002). *Fragile lives, shattered dreams: A report on implementation of preschool education in New Jersey's Abbott districts*. Rutgers University, New Brunswick, NJ: CEER.

Burchinal, M. R., Cryer, D., Clifford, R. M., & Howes, C. (2002). Caregiver training and classroom quality in child care centers. *Applied Developmental Science, 6*, 2-11.

Campbell, F. A., Ramey, C. T., Pungello, E. P., Sparling, J., & Miller-Johnson, S. (2002). Early childhood education: Young adult outcomes from the Abecedarian Project. *Applied Developmental Science, 6,* 42-57.

Dunn, L. M., & Dunn, L. M. (1997). *Peabody Picture Vocabulary Test-Third Edition (PPVT-3)*. Circle Pines, MN: AGS Publishing.

Dunn, L., Lugo, D., Padilla, E., & Dunn, L. (1986). *Test de Vocabulario en Imagenes Peabody*. Circle Pines, MN: American Guidance Service.

Early, D. M. et al. (2007). Teachers' education, classroom quality, and young children's academic skills: Results from seven studies of preschool programs. *Child Development, 78*(2), 558-580.

Frede, E., Lamy, C. E., & Boyd, J. S. (2005) *Not just calendars and counting blocks: Using the NAEYC/NCTM joint position statement "Early childhood mathematics: Promoting good beginnings" as a basis for measuring classroom teaching practices and their relationship to child outcome.* Paper presented at the annual National Association for the Education of Young Children conference, Washington, DC.

Frede, E., Lamy, C. E., Seplocha, H., Strasser, J., Jambunathan, S., Juncker, J., & Wolock, E. (2004). *A rising tide: Classroom quality and language skills in the Abbott Preschool Program: Year two preliminary update of the Early Learning Improvement Consortium.* Trenton, NJ: New Jersey Department of Education. **www.nj.gov/njded/ece**.

Frede, E., Weber, M., Hornbeck, A., Stevenson-Boyd, J., & Colon, A. (2005). *Preschool Classroom Mathematics Inventory*. Available from the first author at efrede@nieer.org.

Ginsburg, H. P., Inoue, N., & Seo, K. H. (1999). Young children doing mathematics: Observations of everyday activities. In J. Copley (Ed.) *Mathematics in the early years*. Washington, D.C: NAEYC.

Harms, T., Clifford, R., & Cryer, D. (2005). *Early Childhood Environment Rating Scale (ECERS-R)*, revised edition. New York, NY: Teacher College Press.

Lamy, C., Frede, E., & ELIC. (2005). *Giant steps for the littlest children: Progress in the sixth year of the Abbott preschool program.* Trenton, NJ: New Jersey Department of Education. www.nj.gov/njded/ece

Lonigan, C., Wagner, R., Torgeson, J., & Rashotte, C. (2002). *Preschool Comprehensive Test of Phonological & Print Processing (Pre-CTOPPP)*. Tallahassee, FL: Florida State University, Department of Psychology.

NAEYC & NCTM. (2002). *Early childhood mathematics: Promoting good beginnings*. A joint position statement of the National Association for the Education of Young Children (NAEYC) and the National Council for Teachers of Mathematics (NCTM). Available at: http://www.naeyc.org/about.positions/psmath.asp or http://www.nctm.org/about/content.aspx?id=6352.

NJ DOE (2002b). *New Jersey preschool teaching and learning expectations: Standards of quality.* Trenton: author.

NJ DOE (2002a). *New Jersey kindergarten core curriculum content standards.* Trenton: author.

Reynolds, A. J., Temple, J. A., Robertson, D. L., & Mann, E. A. (2002). *Age 21 cost-benefit analysis of the Title I Chicago Child-Parent Centers.* (Discussion Paper no. 1245-02). Madison, WI: Institute for Research on Poverty. Available on line at http://www.irp.wisc.edu/publications/dps/pdfs/dp124502.pdf.

Schweinhart, L. J., Montie, J., Xiang, Z., Barnett, W. S., Belfield, C. R., & Nores, M. (2005). *Lifetime effects: The High/Scope Perry Preschool study through age 40* (Monographs of the High/Scope Educational Research Foundation, 14). Ypsilanti, MI: High/Scope Educational Research Foundation.

Smith, S., Davidson, S., & Weisenfeld, G. (2001). *Supports for Early Literacy Assessment for early childhood programs serving preschool-age children*. New York: New York University

Snow, C., Burns, M. S., & Griffin, P. (Eds.). (1998). *Preventing reading difficulties in young children.* Washington, DC: National Academy Press.

Wagner, R., Torgeson, J., & Rashotte, C. (1999). *Comprehensive Test of Phonological Processing (CTOPP).* Austin, TX: Pro-Ed.

Woodcock, R. W., & Munoz, A. F. (1990). *Bateria Woodcock-Munoz Pruebas de Aprovechamiento – Revisados.* Itasca, IL: Riverside Publishing.

Woodcock, R. W., McGrew, K. S. & Mather, N. (2001). *Woodcock-Johnson Tests of Achievement.* Itasca, IL: Riverside Publishing.

Appendix A
Detailed ECERS-R Subscale Scores

<u>Space and Furnishing Subscale.</u> The average score for the Space and Furnishings subscale for the total sample is 4.72 (*SD* = .90). This subscale measures the quality of the indoor and outdoor space available to children during the day. The vast majority of classrooms (almost 80 percent) score above the midpoint 4 with over 40 percent scoring in the good to excellent range. Less than 1 percent score in the inadequate to minimal range (1 to 3). The average scores for Space and Furnishings subscale for the public classroom and private classroom are 4.88 (*SD* = .92) and 4.73 (*SD* = .87), respectively. About 47 percent of the public school classrooms and about 40 percent of the private program classrooms score in the good to excellent range (5 to 7).

Space and Furnishing Subscale
Percentage of Classrooms Scoring 1-7 on the ECERS-R
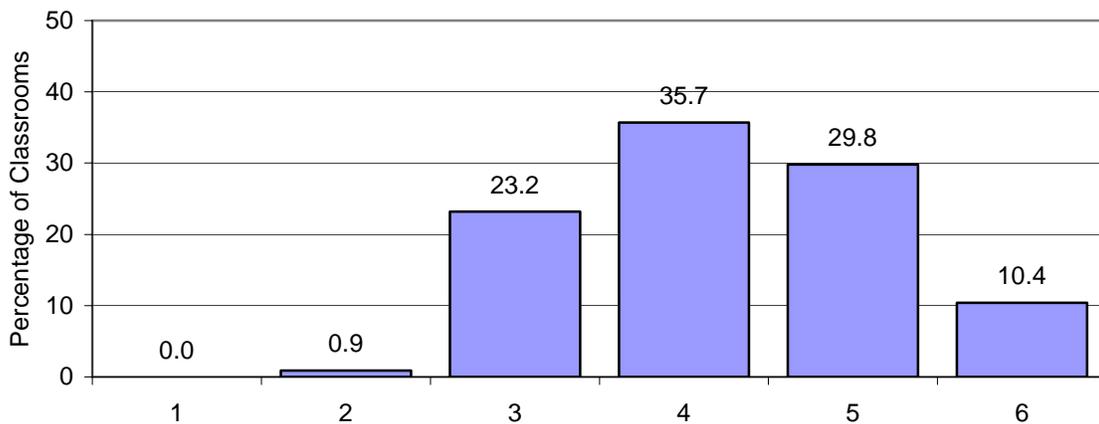
Space and Furnishings Subscale
Percentage of Classrooms Scoring 1-7
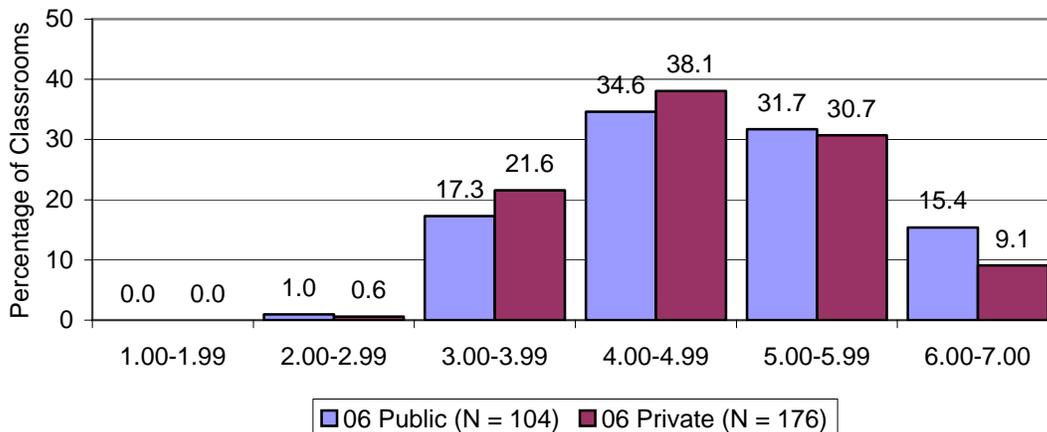
<u>Personal Care Subscale.</u> The ECERS-R subscale with the lowest average score is the Personal Care subscale, with an average score of 4.16 (SD = 1.22). This subscale measures the extent to which classroom environment is conducive to the children's health and safety. About 17 percent of sample classrooms score below 3, in the minimal to inadequate range. However, about 29 percent score a 5 or better. The personal care subscale is the lowest average subscale score in both public school (M = 4.00, SD = 1.29) and private program (M = 4.37, SD = 1.11). While 12 percent of the private program classrooms score below 3, about 20 percent of the public school classrooms score below 3 indicating inadequate personal care routines.

Personal Care Subscale
Percentage of Classrooms Scoring 1-7
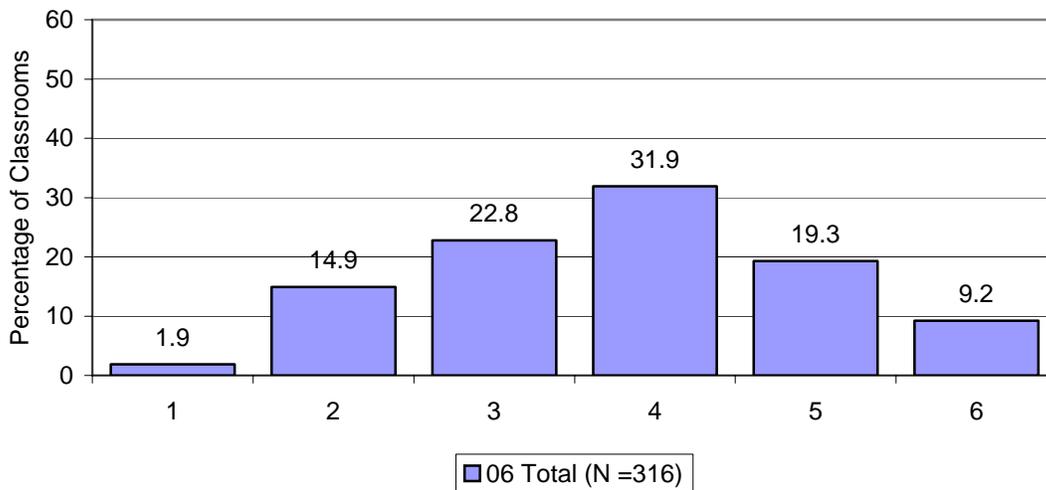
Personal Care Subscale
Percentage of Classrooms Scoring 1-7
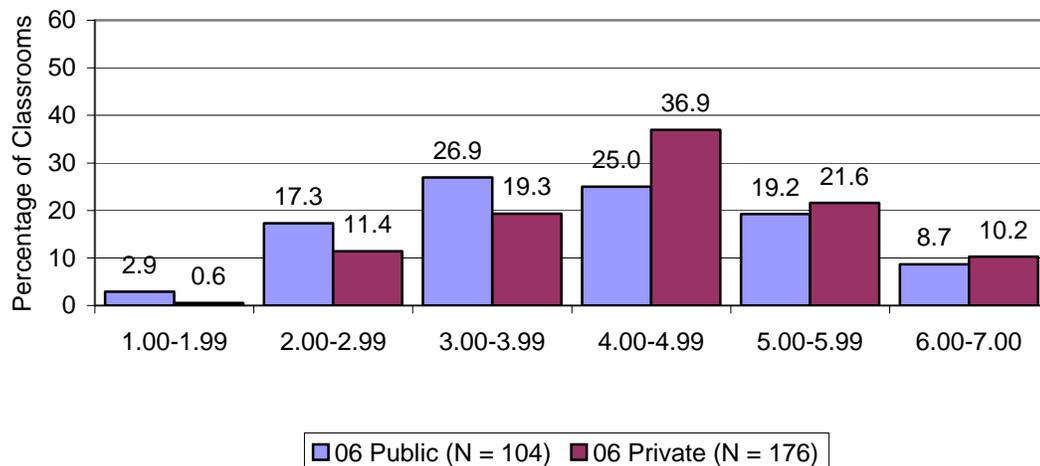
The Language and Reasoning Subscale. The average Language and Reasoning subscale score for the total classroom is higher than the total average score, at 5.03 (*SD* = 1.09). This subscale measures the quality of the selection of books and other language-related materials in the classroom, along with the quality of the communication between adults and children to support children's use of language and reasoning skills. About 60 percent of the classrooms score in the good to excellent range (5 to 7), and 85 percent score above the midpoint. Only about 3 percent score in the inadequate to minimal range (1 to 3). The average scores for Language and Reasoning subscale for the public school and private program are 4.98 (*SD* =1.09), and 5.11 (*SD* = 1.09), respectively. About 56 percent of the public school and 60 percent of the private program classrooms score in the good to excellent range (5 to 7).

Language and Reasoning Subscale -
**Percentage of Classrooms Scoring 1 - 7**
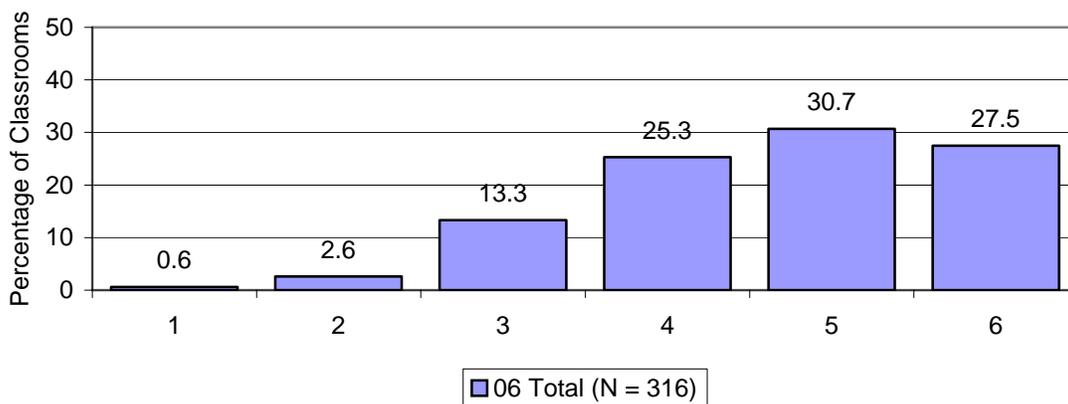
Language and Reasoning Subscale -
**Percentage of Classrooms Scoring 1 - 7**
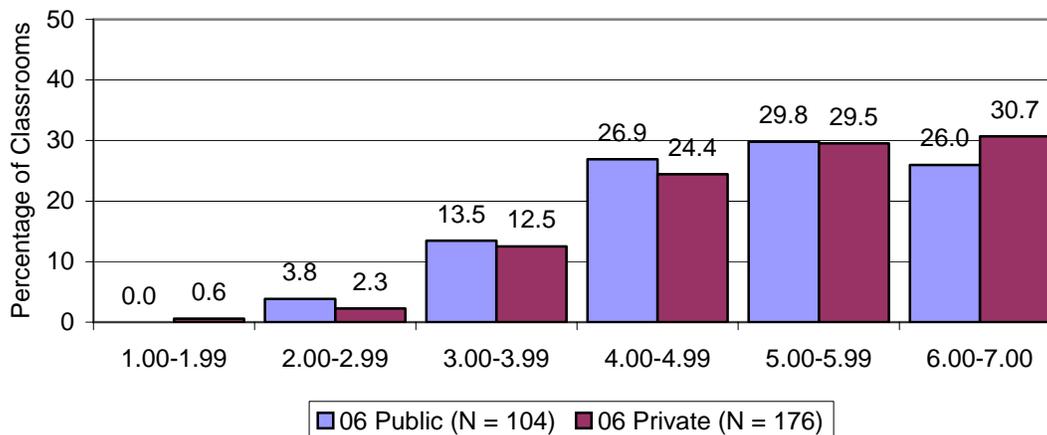
The Activities Subscale. The average Activities subscale score is lower than the total average score, at 4.34 (SD = .94). This subscale measures the quality and quantity of dramatic play, art, music, math, science, fine motor skills and computers. The frequencies of scores indicate that only about 6 percent of classrooms score in the inadequate to minimal range on this subscale, while about 27 percent of sample classrooms score in the good to excellent range. The average scores for Activities subscale for the public school and private program are 4.28 (SD = 1.00), and 4.48 (SD = .89), respectively. About 8 percent of the public schools and about 3 percent of the private programs show inadequate to minimal score range, while about 27 percent of public school and about 30 percent of private program score in the good to excellent range.

Activities Subscale - Percentage of Classrooms Scoring 1 - 7
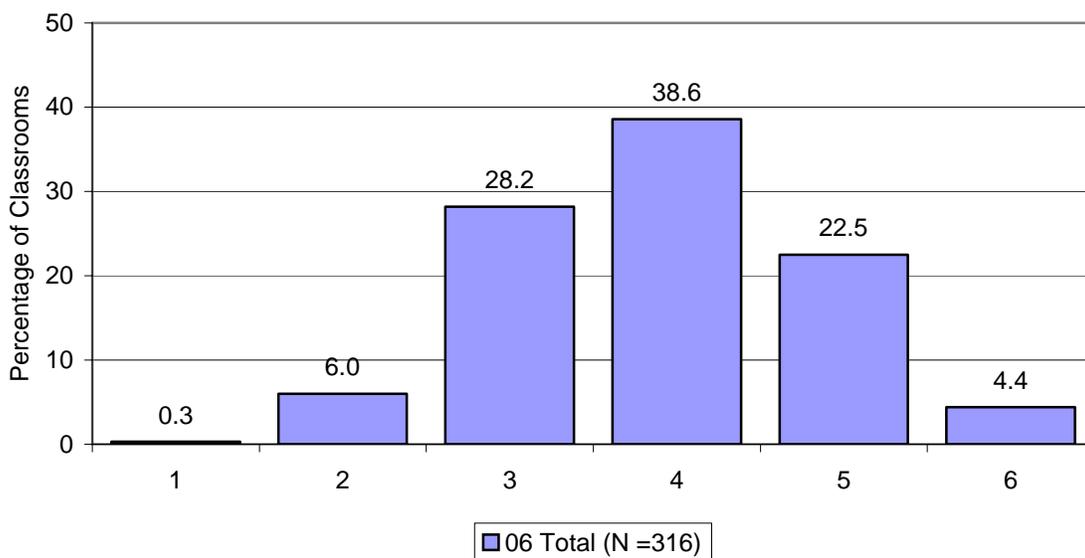

**Activities Subscale - Percentage of Classrooms Scoring 1 - 7**
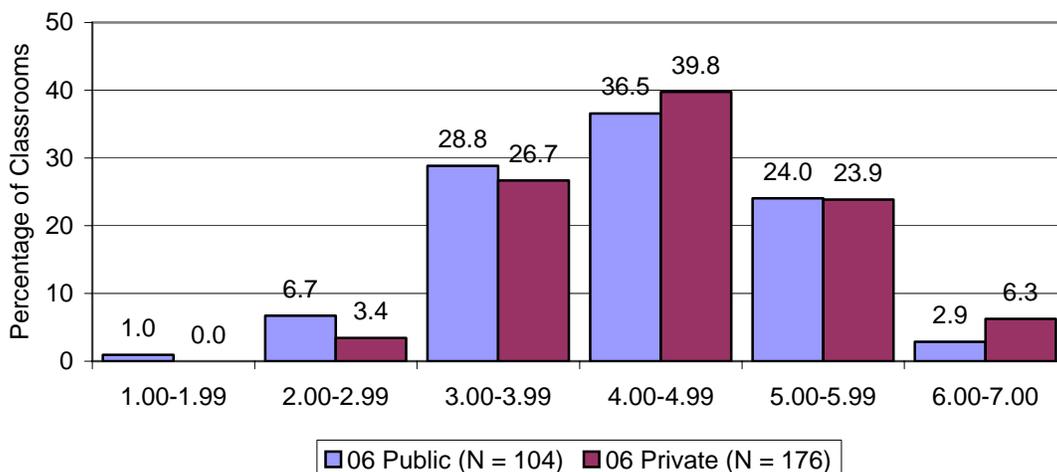
The Interaction Subscale

The ECERS-R subscale with the highest average score is the Interaction subscale with an average score of 5.93 ($SD = 1.08$). This score indicates that on average interactions between adults and children, and among children, show good to excellent quality in the Abbott preschool classrooms. Moreover, nearly 85 percent of sample classrooms score in the good to excellent range on this subscale, and only about 6 percent score below 4. Interaction subscale shows highest average subscale score in both public school (M = 5.88, SD = 1.16) and private program (M = 5.91, SD = 1.08). About 86 percent of the public school classrooms and about 82 percent of private program classrooms score in the good to excellent range (5 to 7).

**Interactions Subscale - Percentage of Classrooms Scoring 1 - 7**
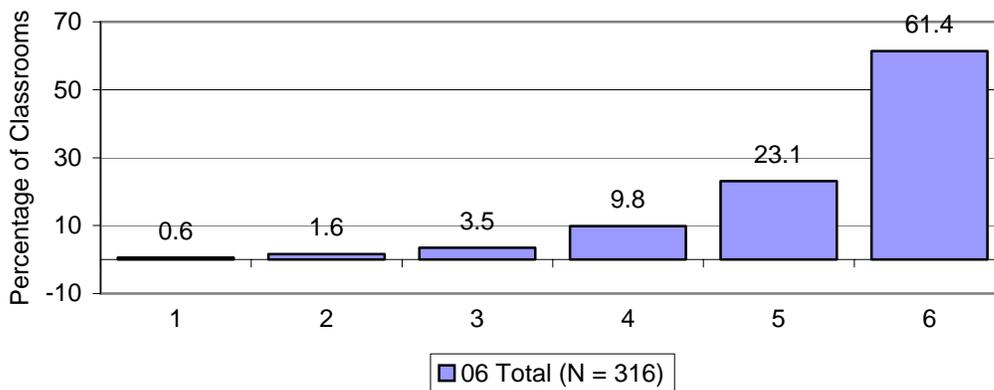
Interactions Subscale - Percentage of Classrooms Scoring 1 - 7
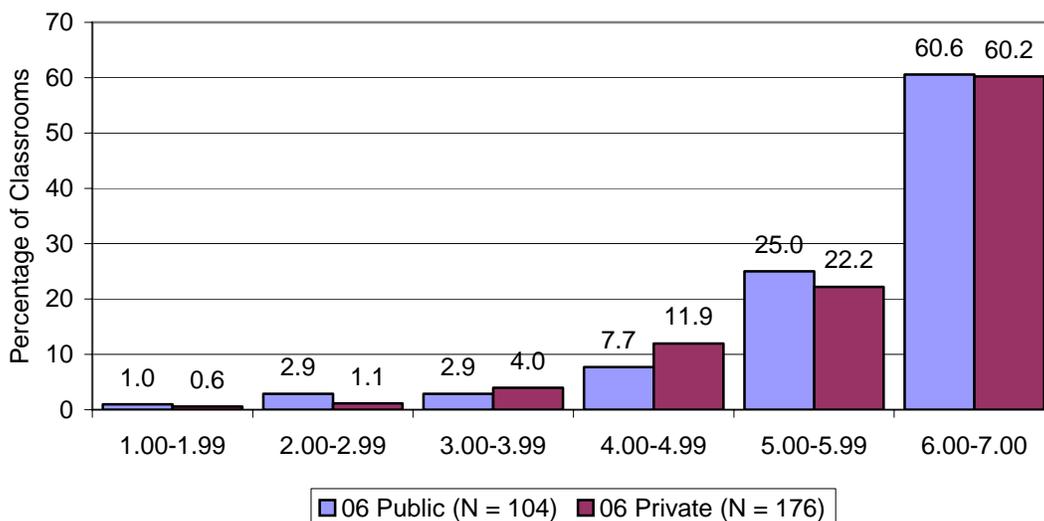
The Program Structure Subscale. The average Program Structure subscale score is slightly higher than the total average score, at 5.02 (*SD* = 1.40), placing the average classroom in good range. This subscale measures the extent to which children have an appropriate schedule of activities and groupings during the day, are given time to play with materials, and are provided for when they have special needs. About 57 percent score a 5 or better, placing them in the good to excellent range, while only 7 percent score below 3. The average scores for Program structure subscale for the public school and private program are 4.88 (*SD* = 1.50), and 5.15 (*SD* =1.37), respectively. About 12 percent of the public schools and about 5 percent of the private program classrooms score in the inadequate to minimal range (1 to 3), while about 54 percent of public school and about 60 percent of private program classrooms score in the good to excellent range (5 to 7).

**Program Structure Subscale - Classrooms Scoring 1 - 7**

**Program Structure Subscale -
Percentage of Classrooms Scoring 1 - 7**
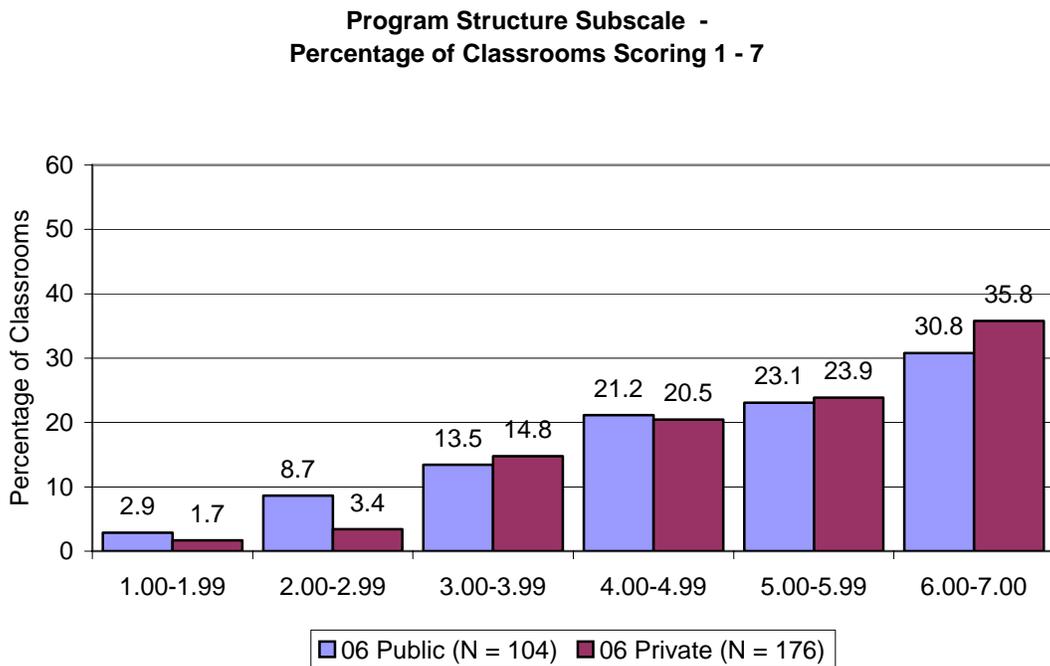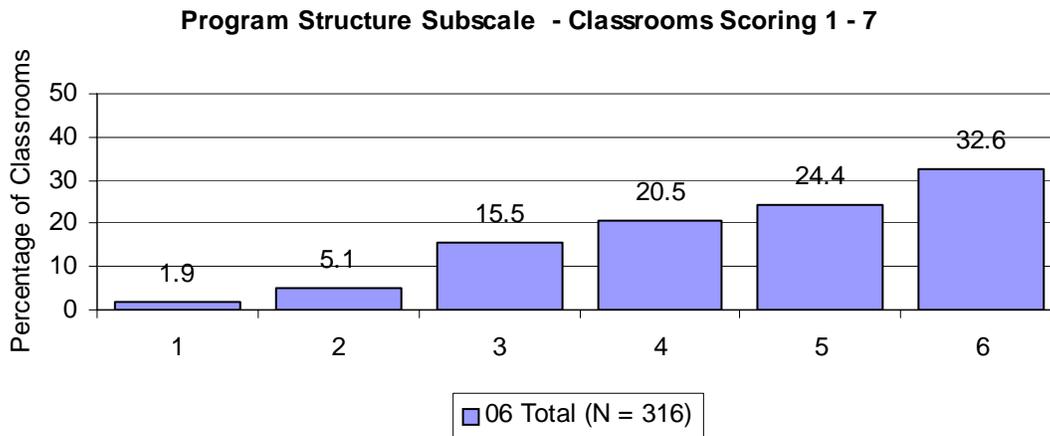
The Parents and Staff Subscale

The Parents and Staff subscale is the next highest scoring subscale across our sample, with an average score of 5.19 (SD = .89). This subscale measures provisions for the personal and professional needs of the program staff including professional development opportunities, evaluation, planning time, and space for private and work-related items. Provisions for the parents of the children who attend the program are also rated here, including the extent of information sharing between parents and program staff, and opportunities for parental involvement in program activities. Over 63 percent of classrooms score in the good to excellent range and about 36 percent score in the minimal to good range. The average scores for Parents and Staff subscale for the public school and private program are 5.58 (SD = .838) and 5.04 (SD = .87), respectively. Only about 1 percent of the public school and the private program classrooms score in the inadequate to minimal range (1 to 3), while about 80 percent of public school and about 57 percent of private program classrooms score in the good to excellent range (5 to 7).

Parents and Staff Subscale - Percentage of Classrooms Scoring 1 - 7

06 Total (N = 316)

Parents and Staff Subscale - Percentage of Classrooms Scoring 1 - 7

06 Public (N = 104)   06 Private (N = 176)