

RESEARCH REPORT



NIEER



Quality in New Jersey's Abbott Preschool Program: A closer look across the years.

Milagros Nores & Erin Harmeyer

July 2023

Acknowledgments

Grateful acknowledgment is made to Robert Wood Johnson Foundation for their financial support of our research in New Jersey. The authors are solely responsible for the content of this report. The data for this report were collected by the Early Learning Improvement Consortium (and in later years, NIEER alone) with leadership from: Jessica Francis, Cynthia Esposito Lamy, and Ellen Frede, NIEER; Holly Seplocha and Janis Strasser, William Patterson State University; Saigeetha Jambunathan and Jo Anne Juncker, New Jersey City University; Heidi Ferrar, Montclair State University; Lorraine Wiley, Rowan University; and Ellen Wolock, New Jersey Department of Education

About NIEER

The National Institute for Early Education Research (NIEER) at the Graduate School of Education, Rutgers University, New Brunswick, NJ, conducts and disseminates independent research and analysis to inform early childhood education policy.

About the Author

Milagros Nores is the Co-Director for Research at the National Institute for Early Education Research (NIEER). Her expertise and interests are in early childhood development, data-driven policy development, evaluation design, equity, and dual language learners.

Erin Harmeyer, Ph.D. Dr. Harmeyer is an Assistant Research Professor at the National Institute for Early Education Research (NIEER). Dr. Harmeyer conducts research focused on early childhood programs and evaluations.

Correspondence regarding this report should be addressed to Milagros Nores at the National Institute for Early Education Research. Email: mnores@nieer.org.

Permission is granted to reprint this material if you acknowledge NIEER and the authors of the item. For more

information, call the Communications contact at (848) 932-4350, or visit NIEER at nieer.org.

Suggested Citation

Nores, M., & Harmeyer, E. (2023). *Quality in New Jersey's Abbott Preschool Program: A closer look across the years*. New Brunswick, NJ: National Institute for Early Education Research.

Key Findings

- √ Classroom quality in New Jersey's publicly funded "Abbott" district classrooms consistently improved from 2003 to 2015, as measured by the ECERS-R observation tool.
- √ In 2003, 22% of Abbott classrooms scored in the good to excellent range on the ECERS-R.
- √ By 2015, this percentage increased to 71% of Abbott classrooms scoring in the good to excellent range.
- √ In 2016, New Jersey transitioned to using the ECERS-3 to measure classroom quality and in 2016 and 2017, 40% and 32% of classrooms, respectively, scored in the good to excellent range on the ECERS-3.
- √ Overall and on all subscales, scores were lower in 2016 and 2017 with the introduction of the ECERS-3.
- √ In 2016 and 2017, 6% and 8% of classrooms were rated inadequate, respectively, compared to no inadequate ratings in 2015.
- √ Abbott classrooms consistently demonstrated high scores on the *Interactions* and *Program Structure* subscales of both the ECERS-R and ECERS-3.
- √ Scores on the *Personal Care Routines* subscale have been relatively low and improved less over time.
- √ *Learning Activities*, and *Language and Reasoning* subscale scores on the ECERS-3 are concerning.

Key Recommendations

Recommendations to improve quality in NJ pre-K programs based on this report are as follows:

- √ Implement continuous quality improvement systems that incorporate the use of classroom observation tools to monitor and enhance quality.
- √ Invest in ongoing quality monitoring efforts to address the wide variation in quality across classrooms, and its implications for equity.
- √ Prioritize interventions and resources for classrooms with inadequate levels of quality, ensuring urgent attention to quickly improving their quality.
- √ Continue to focus on maintaining and enhancing positive interactions and program structure, as reflected in the stronger scores on the *Interactions* and *Program Structure* subscales.
- √ Address the stagnant scores on the *Personal Care Routines* subscale by implementing targeted improvements in routines related to personal care, health, and safety practices.
- √ Target efforts to improve scores on *Learning Activities* and *Language and Reasoning* by providing resources, professional development, and support for teachers to enhance learning experiences.
- √ Take inspiration from the growth in quality reported, and continue to promote the use of reflective practice, targeted technical assistance, and coaching to drive continuous improvement in quality in preschool programs.

Introduction

In 1998, the New Jersey Supreme Court began a series of rulings in the *Abbott v. Burke* school finance case that led to a guarantee of high-quality preschool starting at age three in 31 high-poverty urban school districts, which subsequently became known as “Abbott districts” (Farrie, 2014). The New Jersey Supreme Court’s rulings mandated high standards for these preschool programs including a maximum class size of 15 with a teacher and assistant, teacher degrees (BA minimum) with an early childhood certification, and developmentally appropriate curricula, among others.

Under pressure from the Court to rapidly develop the preschool programs, the New Jersey Department of Education (NJDOE) worked with early childhood experts throughout the state to develop a system founded on high expectations and coherent standards. School districts contracted with pre-existing Head Start and private providers as well as delivering services within public school buildings under a single set of standards and with pay parity (Barnett et al., 2002). The NJDOE supported the existing workforce to complete college and earn early childhood certification and invested in state-of-the-art facilities (Farrie, 2014). The state also created a continuous improvement system in which both the state and districts were responsible for ensuring high quality, and higher education helped to collect data on children’s learning and development and on classroom quality, including the data used in this report (Barnett & Frede, 2017). Currently, over 53,000 students attend full-day, high-quality preschool in a mixed-delivery system of district-run programs, Head Start centers, and community providers, with an average per-child spending of about \$16,700 (Friedman-Krauss, et al., 2023). This ranks New Jersey second in the nation on per child spending.

This brief examines the evolution of quality as measured by the Early Childhood Environment Rating Scales (ECERS; Harms et al., 2015) between 2003 and 2017 in periodic assessments of pre-K quality in Abbott classrooms. From 2003 through 2008 average scores rose consistently. They subsequently plateaued, perhaps rising slightly again in 2015 the last year the Early Childhood Environment Rating Scale, Revised (ECERS-R) was used to assess quality. In 2016 we began to use the Early Childhood Environment Rating Scale, Third Edition (ECERS-3), a revised version of the ECERS-R. Scores on the ECERS-3 in 2016 and 2017 were lower than prior year scores on the ECERS-R in 2014 and 2015, which would be expected as research generally finds lower scores on the ECERS-3 than on the ECERS-R (Hestenes, et al., 2019; Neitzel et al., 2019). However, ECERS-3 scores decreased from 2016 to 2017 raising concerns that call for careful attention to quality state-wide, and to reinvigorating sustainable systems to support quality in both longstanding and new “expansion” programs.

Background

At the turn of the century, education advocates increasingly recognized that the target for school entry in the United States should be preschool, not kindergarten, and that providing high-quality preschool programs for three- and four-year-old children was a viable strategy for addressing achievement gaps in children from different cultural, economic and linguistic backgrounds (Whitebrook et al., 2008). It was in this context that the historic *Abbott v. Burke* decision was made in the New Jersey Supreme Court, with universal preschool education proposed as part of a comprehensive approach to fulfill the state’s constitutional obligation to meet the needs of children in high-poverty school districts. The *Abbott v. Burke* decision was ground-breaking in the inclusion of *high-quality* early education beginning at age 3 as an important element of educational adequacy for which state government would be held responsible.

The vision and standards for quality in the Abbott program were developed through a combination of follow-up court rulings and New Jersey Department of Education (NJDOE) actions. For example, the NJDOE developed new expectations (standards) for learning and teaching as well continuous improvement (Barnett et al., 2013). The Court expanded beyond its initial recommendations based on evidence regarding the characteristics of programs found effective in research, for example, a class size of 15 students or less with one lead teacher with an early childhood certificate and one assistant teacher per classroom. The court’s *Abbott VI* decision, in 2000, required the use of a developmentally appropriate curriculum, adequate facilities, and services for special education, bilingual education, transportation, and health (Whitebook et al., 2008). The *Abbott VII* decision, in 2002, ruled that teachers who had completed their bachelor’s degree and certification should have salaries raised to parity with K-12 teachers. Finally, along with providing high-quality preschool, Abbott districts were also required to enroll 90% of their preschool universe by the 2004-05 school year.

The timeline for providing the services to all children also was very short. The Court required Abbott districts to enroll 90% of their preschool universe by the 2004-05 school year. To meet that timeline, make maximum use of existing resources (personnel and facilities), and recognize the needs of families for child care, the Abbott Preschool Program developers committed to providing services in a mixed delivery system (e.g., in a mix of private childcare centers and public schools) with a system of wrap-around care beyond the school day and year for families that needed more hours of childcare.

In the first year (1999-2000) of the Abbott program, approximately 19,000 children were enrolled. Expansion of the program was rapid: five years later, more than 38,000 of the 54,000 three- and four-year-olds in these districts were in preschool (Frede et al., 2004). As the program expanded, so too, did efforts to measure program quality, and these efforts included various processes of collecting information on quality and mechanisms to help districts understand the results, as well as the development of state and local improvement plans (Frede, 2005). The NJDOE formed the Early Learning Improvement Consortium (ELIC) to measure the effects of Abbott preschool programs (Frede et al., 2004). Annual discussions between the NJDOE and ELIC personnel on district results, along with the data collected at the district level (districts were also required to collect similar information), informed annual professional development plans.

When the program began, few teachers (approximately 43%) were certified, and a representative sample of Abbott classrooms showed they were of moderate quality overall (3.86 out of 7 on the ECERS-R; Barnett et al., 2002). Additionally, about one in five classrooms scored on the range of inadequate quality (1.00 – 2.99). However, even amidst the rapid expansion efforts, ELIC demonstrated program improvements in the initial years of expansion: For example, in the 2003-04 school year, 24% of programs assessed scored in the good or excellent range, compared to 13% of programs scoring in that range in the year prior (Frede et al., 2004).

New Jersey's approach to preschool is important to study because rather than completely replacing systems that were already in place, the Abbott preschool program set out to provide training and education to enhance teaching quality and to maintain strong teaching by boosting pay and raising standards (Barnett et al., 2013). Measuring the quality of Abbott classrooms longitudinally thus provides important information for policymakers about the long-term impacts these types of programmatic decisions have had on quality. The purpose of this report is to look more closely at quality longitudinally in these Abbott programs.

This report includes data collected periodically between 2003 to 2017. The quality measures used in this report all stem from the Environment Rating Scales (ERS) system. The ECERS-R, published in the late 1990s, is one of the most commonly used measures of early childhood classroom quality, and was used in a number of national studies on early childhood, including the Early Childhood Longitudinal Study – Birth Cohort (ECLS-B; National Center for Education Statistics, 2016) and the Head Start Family and Child Experiences Survey (FACES; Moiduddin et al., 2012). However, some researchers have raised concerns about using the tool as a standalone measure of quality in high-stakes contexts (Perlman et al., 2004), citing concerns of the tool's weak associations with child outcomes. For instance, in a meta-analysis of 16 studies evaluating the association between the ECERS/ECERS-R and child outcomes, researchers found that while the tools appear to be associated with child outcomes including language and positive behavior, the associations are modest (Brunsek et al., 2017).

It is noteworthy that the continuous improvement system for the Abbott program employed multiple measures of quality including some specifically aligned with state standards and not on the ERS, alone (e.g., Frede et al., 2004). Patterns on these other measures over time are similar to those for the ERS. As they have been less widely used in studies outside New Jersey, and are less familiar to national audiences, this report focuses on ERS scores.

The ECERS-3 (Harms, Clifford & Cryer, 2015), introduced in the 2015-2016 school year, was revised to provide greater focus on the teacher's role in supporting cognitive development and social skills, and has a somewhat diminished focus on materials. Along with changes in scoring and conducting the observation, new items were added that emphasize research-based practices, including individualizing teaching and learning and the importance of high-quality interactions between teachers and children (Neitzel et al., 2019).¹ In a study that included 225 classroom observations across six states

¹ Additional differences include that the ECERS-3 only considers what is observed during the 3-hour time sample (additional time may be used only to review materials or the safety features of the playground), and does not include teacher interviews as the ECERS-R did; it is used in classrooms for 3-5-year-old children (the ECERS-R could be used for 2.5-5-year-old children); it has 35 items with 6 subscales (a reduction from 43 items with 7 subscales); requires less of a focus on accessible materials and more of a focus on how teachers interact with children using materials; major hazards for indoors and outdoors are separated into different items; and access to materials must be observed for 1 hour of the 3-hour assessment.

in which the ECERS-3 and ECERS-R were conducted simultaneously in classrooms, Neitzel and a team of researchers (2019) found significant differences between the mean total and subscale total scores of the two instruments. Total scores were significantly higher for the ECERS-R, as were all subscale scores, with the exception of Personal Care routines, in which the ECERS-3 score was higher. The largest score differences were between the total score (with an ECERS-3 mean total score 0.83 points lower than the ECERS-R mean total score); the Language subscale (with an ECERS-3 mean subscale score 1.2 points lower than the ECERS-R mean subscale score); and the Activities subscale (with an ECERS-3 mean subscale score 1.5 points lower than the ECERS-R mean subscale score). The authors of this study concluded that “the ECERS-R and ECERS-3 are two distinct quality measurement tools, rather than one instrument measuring the same features within learning environments” (p. 416). It is thus important to keep in mind these differences in the tools when looking longitudinally at the quality of classrooms hereby reported for the Abbott preschool program. This report covers data collected when the ECERS-R and the ECERS-3 tools were used; scores for the same tool will be contrasted, but comparisons of scores across the different tools were used should be interpreted with caution.

Study Methods

This report draws on data from multiple evaluations on the quality of the Abbott preschool program between 2003 and 2017. We address the following research questions:

1. What were the long-term trends in quality during the 2003 through 2017 period?
2. What areas of quality showed the strongest improvements in this period?
3. Were there areas of quality that experienced less positive long-term change?

Procedures

The datasets in the present analyses derived from annual statewide observations of classroom quality in the districts providing the program, which were conducted by New Jersey’s Early Learning Improvement Consortium and the National Institute for Early Education Research (NIEER) at Rutgers University. Samples in each year follow similar protocols for data collection, which include random selection of classrooms across districts, as well as training of the data collection team to reliability by NIEER personnel that have maintained reliability with the instrument developers. Since the fall of 2003, a random sample across the Abbott districts, stratified by public school or private provider, were observed periodically through 2017 (Frede, 2005).

After 2017, annual assessments of quality were no longer representative of the ‘former’ Abbott districts as the samples include preschool programs in additional districts due to New Jersey’s policies to expand the program across the entire state. Therefore, this report only reports progress through 2017. The number of classrooms included in this analysis varies year-to-year: in 2003, 94 classrooms are included, and in subsequent years the sample varied between 285 and 397 classrooms with the exception of 2013 in which 175 classrooms were included.

Measures

The Early Childhood Environment Rating Scale—Revised. ((ECERS-R; Harms, Clifford & Cryer, 1998) is an observation and rating instrument for preschool classrooms serving children ages two-and-a-half to five. The ECERS-R contains 43 items spread across seven subscales: Space and Furnishings, Personal Care Routines, Language-Reasoning, Learning Activities, Interaction, Program Structure, and Parents and Staff. The ECERS-R is administered over a 3-hour observation period, with observers noting the presence or absence of a large number of indicators. Following the observation, staff are asked questions about the types of activities that typically occur but could not have been observed; these responses are used to finish any unscored indicators. A rating scale between 1 and 7 is used. A rating of 1.00 – 2.99 indicates inadequate quality, 3.00 – 4.99 indicates moderate quality, and 5.00 – 7.00 indicates good to excellent quality.

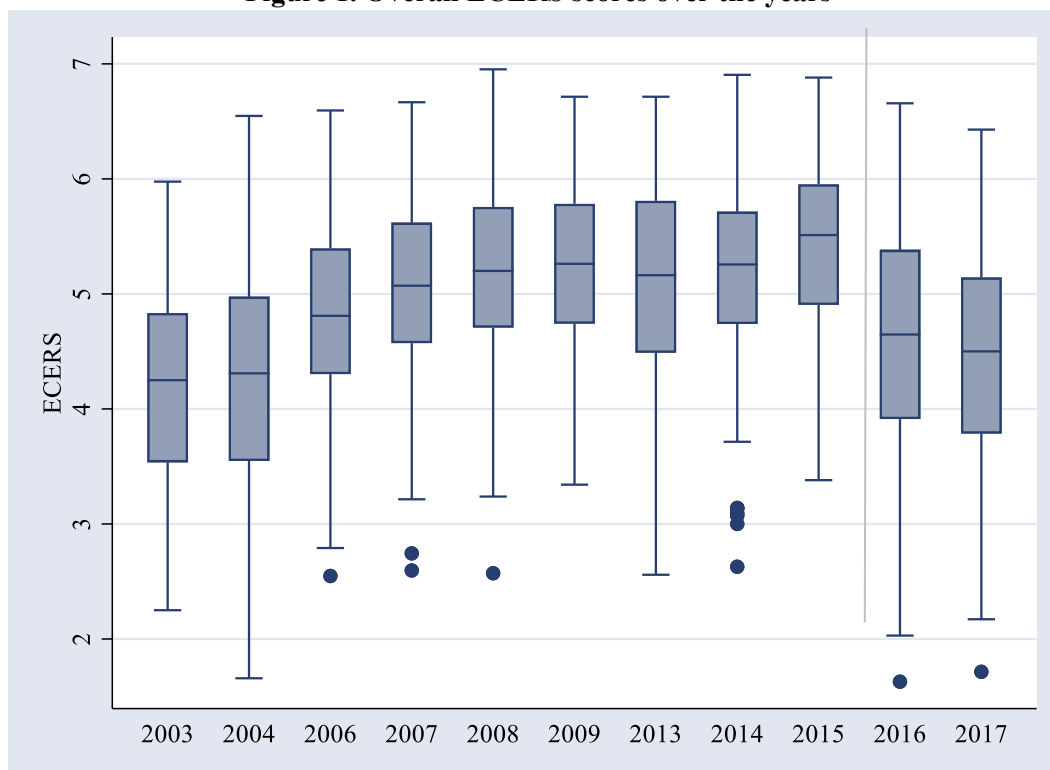
The Early Childhood Environment Rating Scale—Third Ed. (ECERS-3; Harms, Clifford & Cryer, 2014) was developed with the same structure and goals as the ECERS-R, however, there is a greater focus on teaching and interaction with materials, instead of just the materials and space itself (Neitzel, 2019). The total ECERS-3 score represents an average of the scores on 35 items spread across 6 subscales: Space and Furnishings, Personal Care Routines, Language and Literacy, Interaction, and Program Structure. A rating of 1.00 – 2.99 indicates inadequate quality, 3.00 – 4.99 indicates moderate

quality, and 5.00 – 7.00 indicates good to excellent quality. The most updated notes for clarification published by the Environment Rating Scales Institute (ERSI) were utilized each year when scoring all classrooms in this sample.

Results

To address the first research question about trends in quality over time, we looked at the overall ECERS scores for the classrooms in our sample. We find that ECERS-R scores consistently increased from 2003 – 2015 (Figure 1). The total mean score for classrooms observed in 2003 was 4.21 (demonstrating moderate quality). These mean scores increased year-to-year from 2003 to 5.21 in 2008 (indicating classrooms were, on average, of good to excellent quality) and to 5.11 and 5.41 in 2013 and 2015, respectively. In 2016, the first year that ECERS-3 was used, the mean total score across classrooms was 4.61 and in 2017, the mean score was 4.43.² Similarly, the complete distributions showed a shift to higher scores with a reduction on the number of classrooms scoring under 3 (at the inadequate level). While the ECERS-3 has not been utilized as broadly as the ECERS-R, scores exist for some preschool programs. As means of comparison, in 2018, the Seattle Preschool Program and the WV preschool programs were also assessed by NIEER and these showed average ECERS-3 scores of 4.94 and 4.04, respectively.

Figure 1. Overall ECERS scores over the years



Note: Box plots depict mean scores, confidence intervals and the overall distributions, as well as outliers if and when any were present. The vertical line separates the years for which ECERS-R was used, from those for which ECERS-3 was used. The 2003 and 2013 samples were reduced samples.

Although just 22% of classrooms scored in the good to excellent range in 2003 on the ECERS-R, by 2015 that number increased to 71% of classrooms in our sample. When the ECERS-3 was used in 2016 and 2017, there were 40% and 32% of classrooms scoring in the good to excellent range, respectively. While 12% of classrooms were rated as inadequate in 2003, there were no classrooms at this level in 2015. However, when the ECERS-3 was introduced in the spring of 2016 and 2017, 6% and 8% of classrooms were rated at the inadequate level, respectively. In summary, the proportion of classrooms rated in the good to excellent range increased by nearly 50 percentage points from 2003 to 2015

² We also estimated ECERS scores controlling for month of observation, type of ECERS (R or 3), number of children present, number of children with an IEP and age of the youngest and oldest child in the classroom. The trend stays positive with the difference between ECERS-R and ECERS-3 slightly attenuated.

(from 22% to 71% of classrooms). While fewer programs were rated in the good to excellent range with the onset of the use of the ECERS-3 in 2016, a majority of programs (54% in 2016 and 61% in 2017) landed in the moderate category in 2016 and 2017, with a small percentage falling in the inadequate range.

Figure 2. Percentage of Abbott Preschool Program classrooms observed scoring in the good to excellent range (mean ECERS score greater than 5.0)

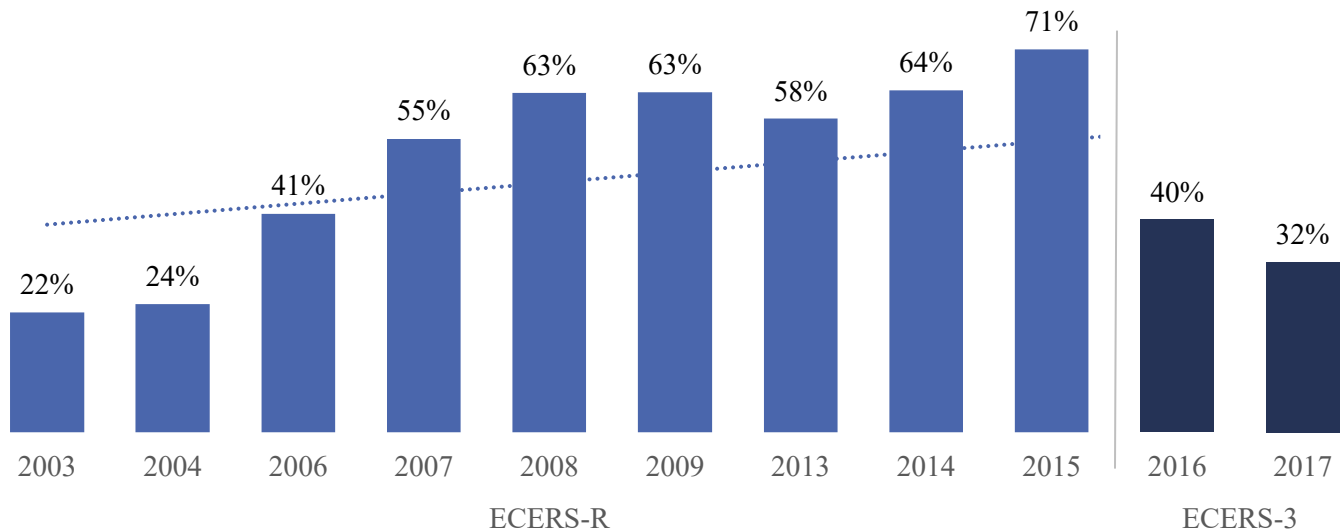
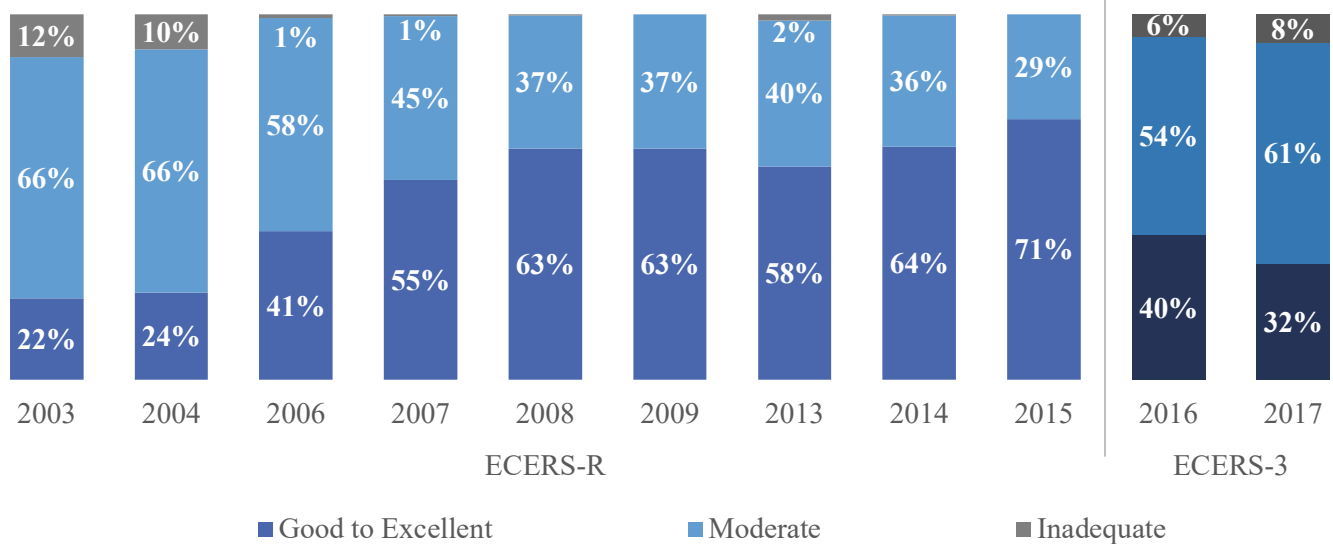


Figure 3. Percentage of Abbott Preschool Program classrooms observed scoring in all three quality categories



To answer our second research question about which areas of quality showed the starkest improvements over time, we looked at the individual subscale scores from 2003 through 2015, when the ECERS-R was used. During this timeframe, a similar trend to what was seen in the overall scores is noted, with classrooms generally showing an increase in quality over time as measured by the separate subscales. For example, the mean Space and Furnishings subscale score for classrooms in 2003 was 3.97; this score was 5.20 in 2015. Both the Learning Activities and Program Structure subscales showed similarly large increases over this timeframe, with the Learning Activities subscale scores increasing from 3.56 to 5.12 and the Program Structure subscale scores increasing from 4.51 to 6.07. The Personal Care Routines subscale also showed increased scores over time, although not consistently from year-to-year, and the overall change in scores was the smallest of the subscales: The mean score in 2003 was 4.21, and in 2015, it was 4.36.

When the ECERS-3 was used in 2016 and 2017, subscale scores were lower than those measured by the ECERS-R: For the Space and Furnishings, Personal Care Routines, Language and Literacy, and Learning Activities subscales,

scores were in the 4 range, or moderate quality. On both the Interaction and Program Structure subscales, scores were on average in the 5 range, demonstrating that programs were rated as having good to excellent quality in these domains. The subscales with the highest scores on the ECERS-R were Interaction and Program Structure, which was also the case with the ECERS-3.

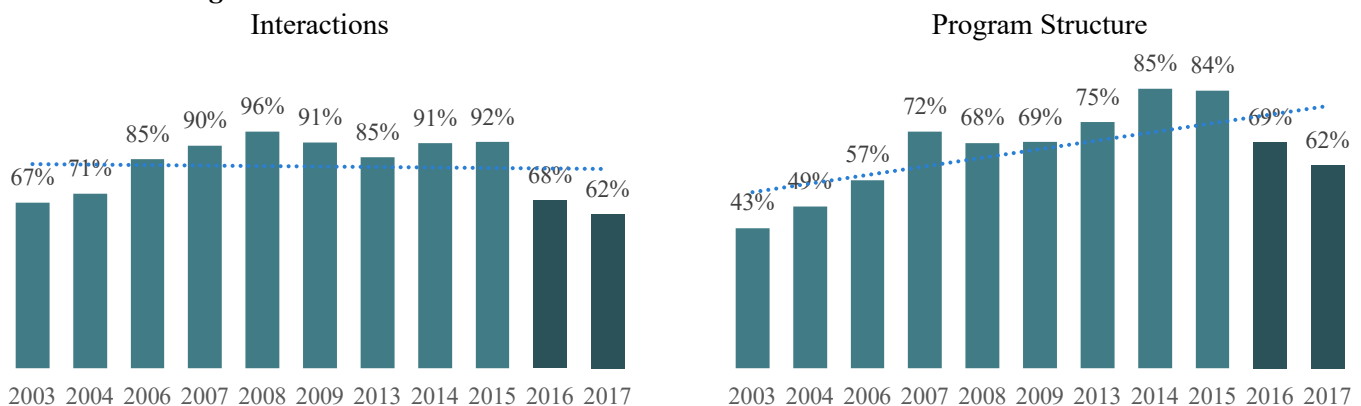
Table 1. Mean subscale scores for classrooms in the Abbott Preschool Program from 2003 – 2015, when the ECERS-R was used to measure quality and for 2016 & 2017 with the ECERS-3.

	2003 Mean (Range) N = 94	2004 Mean (Range) N = 310	2006 Mean (Range) N = 317	2007 Mean (Range) N = 397	2008 Mean (Range) N = 326	2009 Mean (Range) N = 316	2011 Mean (Range) N = 326 ^b	2013 Mean (Range) N = 175 ^c	2014 Mean (Range) N = 303	2015 Mean (Range) N = 286	2016 Mean (Range) N = 293 ^d	2017 Mean (Range) N = 300 ^d
Space & Furnishings	3.97 (2.12-6.63)	3.97 (1.5-6.88)	4.72 (2-6.88)	4.96 (2.25-6.88)	5.05 (2.38-7)	5.03 (3.13-7)	5.16 (2.5-7)	4.93 (2-7)	5.18 (2.43-7)	5.20 (2.63-7)	4.43 (2.14 – 7)	4.20 (2 – 7)
Personal Care Routines	4.21 (1.33-7)	4.01 (1-7)	4.17 (1.17-7)	4.25 (1.33-7)	4.32 (1.67-7)	4.35 (1-7)	4.49 (1.17-7)	4.32 (1.67-7)	4.12 (1.5-7)	4.36 (1.25-7)	4.36 (1.25 – 7)	4.26 (1.25 -6.75)
Language-Reasoning^a	4.42 (1-7)	4.56 (1-7)	5.03 (1.5-7)	5.17 (1.75-7)	5.48 (2-7)	5.56 (1.75-7)	5.84 (2.5-7)	4.41 (1-6.75)	4.86 (1.6-7)	5.17 (1.5-7)	4.86 (1.6-7)	4.56 (1 – 7)
Learning Activities	3.56 (2.1-5.2)	3.60 (1.2 – 6.3)	4.34 (1.8-6.8)	4.68 (2-6)	4.87 (1.6-7)	4.86 (2.5-7)	5.00 (2.3-7)	4.76 (1.3-7)	4.22 (1.18-6.8)	5.12 (2.2-7)	4.22 (1.18 – 6.8)	4.02 (1.45 – 6.72)
Interaction	5.11 (1-7)	5.40 (1-7)	5.93 (1-7)	6.22 (1.4-7)	6.45 (1.6-7)	6.33 (1.8-7)	5.98 (1.4-7)	6.02 (1-7)	5.26 (1-7)	6.30 (1.6-7)	5.26 (1 – 7)	5.17 (1-7)
Program Structure	4.51 (1-7)	4.63 (1-7)	5.02 (1-7)	5.48 (1-7)	5.4 (1-7)	5.45 (1.67-7)	5.41 (1.33-7)	5.64 (1-7)	5.20 (1-7)	6.07 (1-7)	5.20 (1-7)	5.00 (1-7)
TOTAL	4.21 (2.25-5.98)	4.27 (1.66-6.55)	4.82 (2.55-6.6)	5.08 (2.6-6.67)	5.22 (2.57-6.95)	5.24 (3.34-6.71)	5.30 (2.52-6.67)	5.11 (2.56-6.71)	5.22 (2.63-6.9)	5.41 (3.38-6.88)	4.61 (1.63-6.66)	4.43 (1.71-6.43)

^a2013 scores for only Camden, East Orange, Irvington, Jersey City, Long Branch, Newark, Paterson, and Phillipsburg. ^bFor 2011 we do not have the full sample to report in figures above. Averages reported here are those reported by NJDOE.³ ^cThis subscale was renamed Language & Literacy in the ECERS-3 version. ^dECERS-3 scores reported for 2016 and 2017.

Breaking down the percentage of classrooms scoring in the good to excellent range on each subscale, classrooms showed consistent improvement on the Interactions and Program Structure subscales, with a majority of classrooms scored as being in the good to excellent range in these areas, even with the switch to the ECERS-3 in 2016. In 2003, 67% of classrooms scored in the good to excellent range on Interactions; in 2015, 92% of programs met this bar. When the ECERS-3 was used in 2016 and 2017, 68% and 62% of classrooms scored in this range, respectively. Program Structure also showed increased scores over time, and most classrooms were rated in the good to excellent range. While 43% of programs met that standard in 2003, by 2015, 84% did so. With the switch to the ECERS-3 in 2016, 69% of programs scored in this range, as did 62% in 2017.

Figure 3. Percentage of Abbott Preschool Program classrooms observed scoring in the good to excellent range on the Interactions and Program Structure subscales.

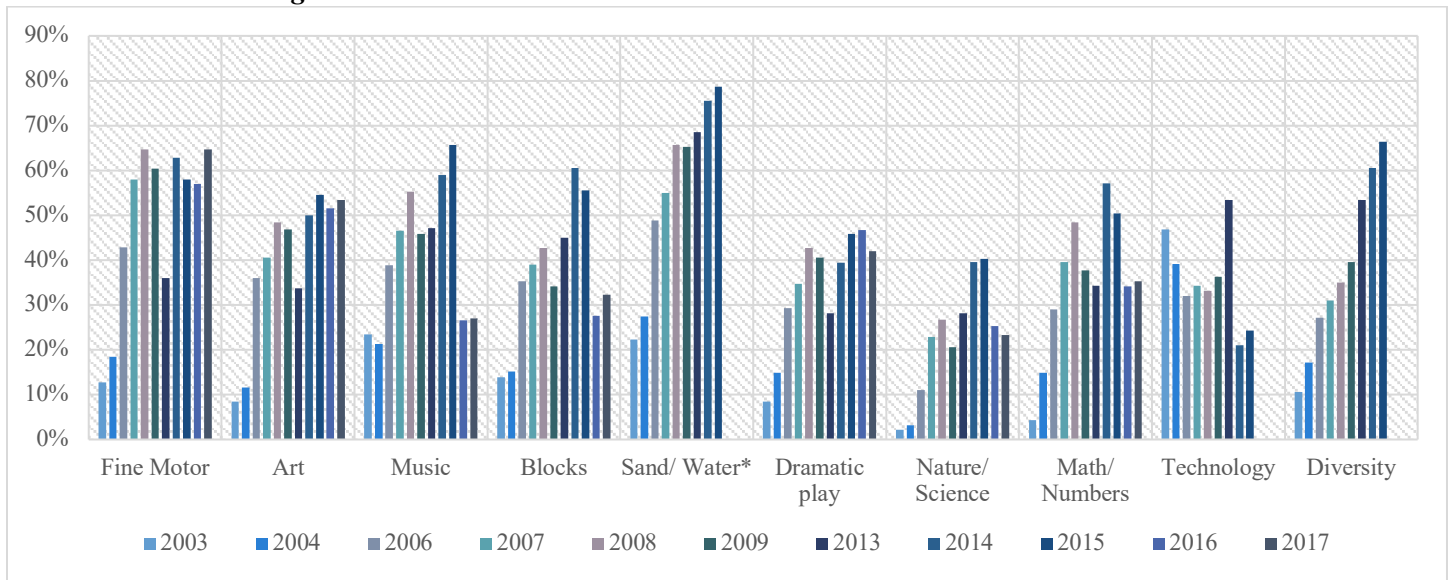


To further address this research question, we looked at the changes in item scores within the subscales. Items in the Learning Activities subscale, which address children’s access to and usage of materials including fine motor, art, and

³ <https://nieer.org/wp-content/uploads/2021/01/New-Jersey-Preschool-Quality-Evaluation-Study-Spring-2016-Summary-Report.pdf>

block play, showed the greatest improvement over time on the ECERS-R. For example, in 2003 just 22% of classrooms scored in the good to excellent range on the Sand/Water item of this subscale; that number increased to 79% of classrooms in 2015 (this item is not present as such in the ECERS-3). Similarly, while 13% of classrooms scored in the good to excellent range on the Fine Motor item in 2003, that number had increased to 57% of classrooms scoring in this range in 2015. The item Promoting Acceptance of Diversity also showed similar gains, with 11% scoring in the good to excellent range in 2003, and 66% of classrooms scoring in this range in 2015. Art, Music/Movement, Blocks, and Math/Number all showed similar increases, with the proportion of classrooms scoring in the good to excellent range in 2015 as compared to 2003 increasing by 40 percentage points or more for each of these items. With the change to the ECERS-3 in 2016, many classrooms still scored in the good to excellent range on these items; for example, 65% of classrooms scored in the good to excellent range on the Fine Motor item in 2017, and 53% of classrooms scored in the good to excellent range on the Art item in 2017. These percentages were lower for Nature/Science and Math/Numbers.

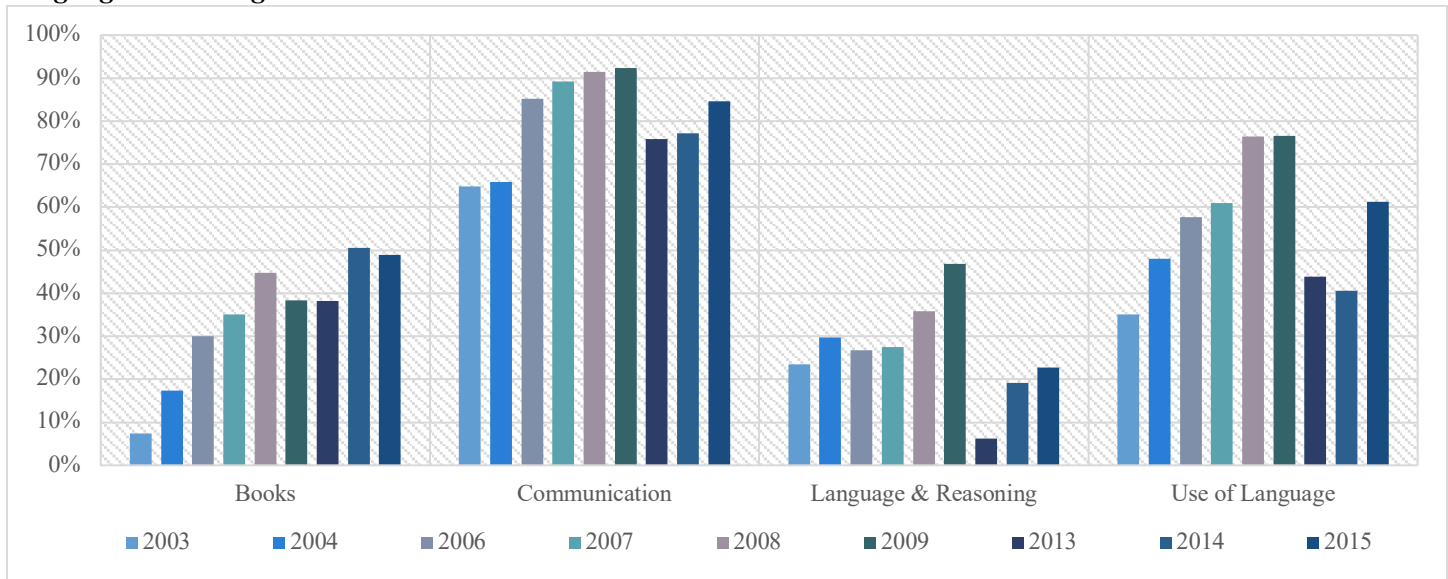
Figure 4. Percent of Abbott Preschool Program classrooms scoring on the good to excellent range across items included in the Learning Activities subscale.



Note: Three items from this subscale (sand/water, math/number, and use of technology) were revised on the ECERS-3. The graph excludes the item on use of technology. For the Math/Numbers item the 2016 and 2017 scores are an average of three items in ECERS-3 (Math materials and activities, Math in daily events and Understanding written numbers). Scores for 2016 and 2017 are not included for sand/water (*) as this item is not included as such anymore in the ECERS-3.

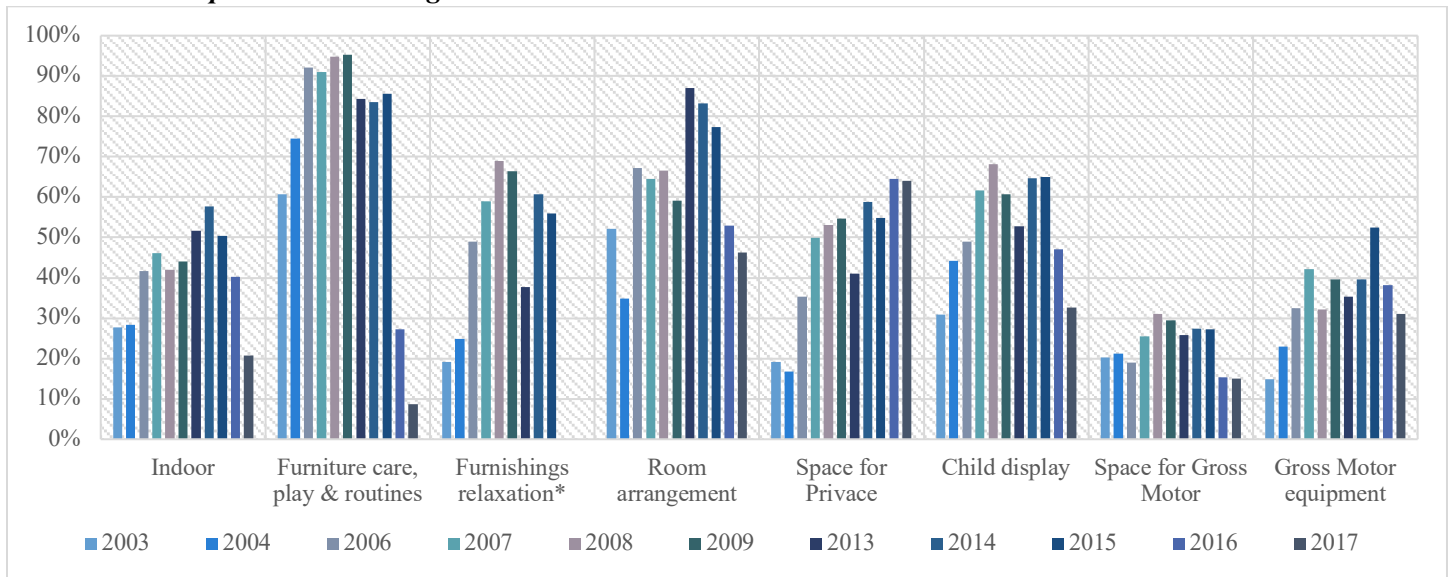
Although less pronounced, improved scores were noted across items on the subscales of Interaction, Language-Reasoning, Space and Furnishings, and some items of the Personal Care Routines subscale. For example, in the subscale Language-Reasoning, 7% of classrooms scored in the good to excellent range on the item Books & Pictures in 2003, and 35% of classrooms scored in that range on the Informal Use of Language item. In 2015, 49% of classrooms scored in the good to excellent range on Books & Pictures, and 61% scored in the good to excellent range on Informal Use of Language.

Figure 5. Percentage of Abbott Preschool Program classrooms scoring in the good to excellent range on the Language-Reasoning subscale.



Note: On the ECERS-3, the subscale was renamed Language and Literacy, and several items were revised and expanded making comparisons for 2016 and 2017 difficult and thus not included here.

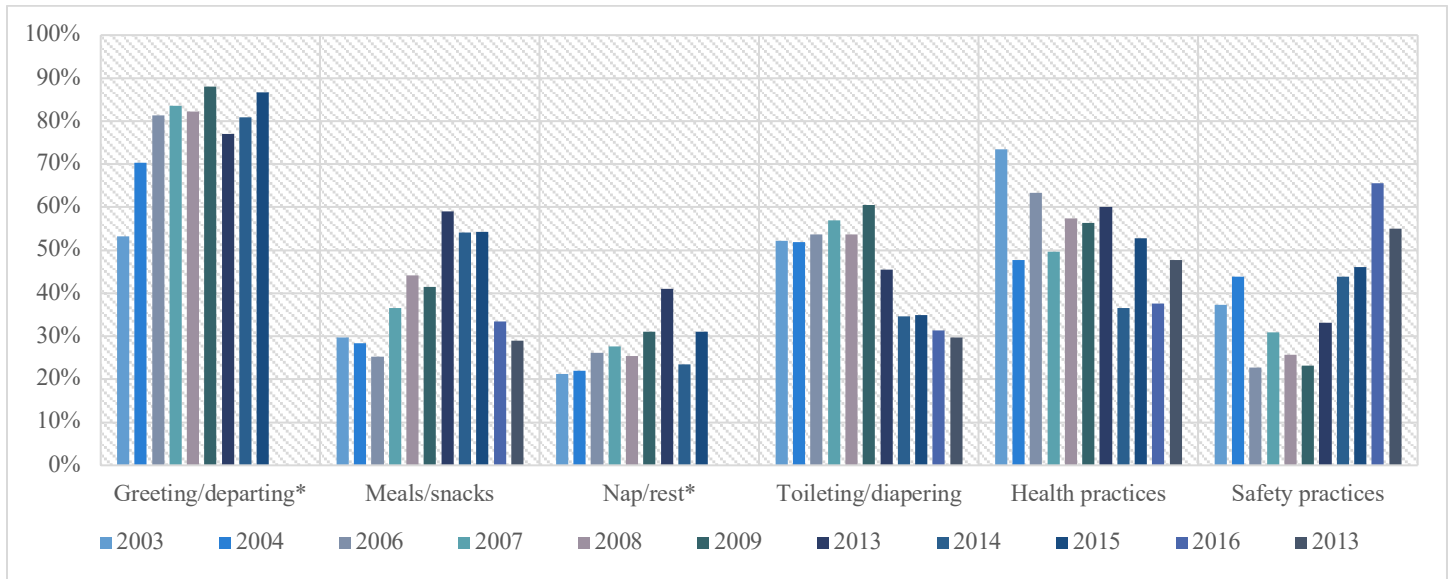
Figure 6. Percentage of Abbott Preschool Program classrooms scoring in the good to excellent range on items included in the Space & Furnishings subscale.



Note: The ECERS-3 measure does not include the “Furnishings for relaxation and comfort” item (*), so scores from 2016 and 2017 are not displayed here for that item.

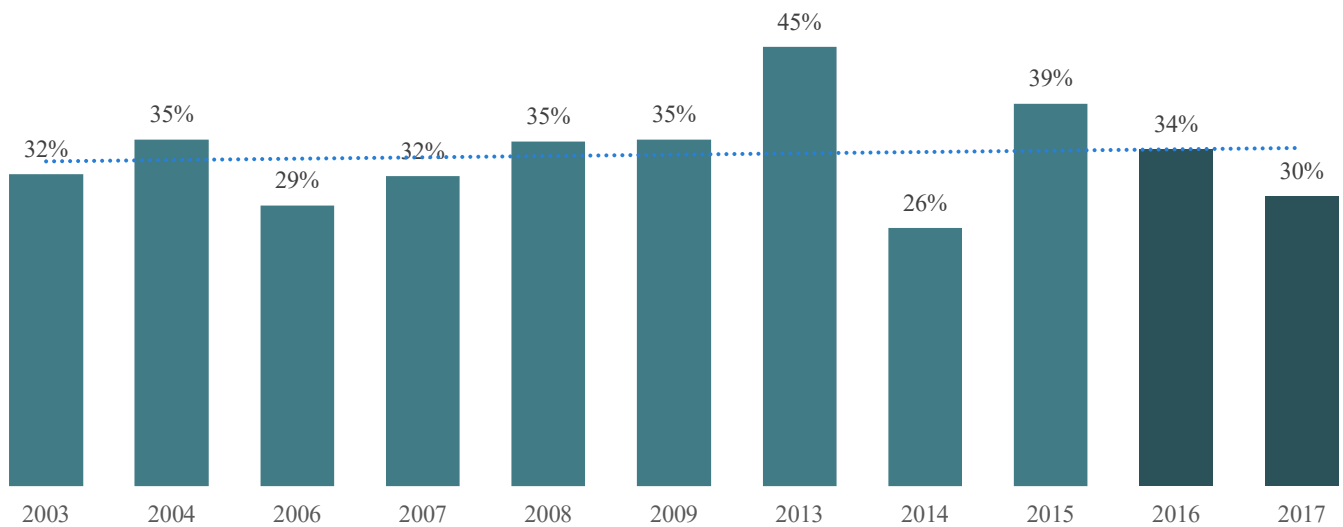
Finally, for our third research question, we were interested in understanding if there were areas of quality that were more resistant to long-term change within the Abbott Preschool Program. To do this, we looked at the subscale scores over time, particularly focusing on drops in scores in 2016 with the use of the revised version of the ECERS. In terms of Subscales not demonstrating much change over the timeframe of our study, this was most apparent with the Personal Care Routines subscale. While 32% of classrooms scored in the good to excellent range on this subscale in 2003, this number had increased to just 39% of classrooms in 2015. When the ECERS-3 was used in 2016 and 2017, these numbers stayed about the same, with 34% of classrooms in this range in 2016 and 30% of classrooms in this range in 2017. Breaking down the specific items within this subscale gives a clearer picture of this lack of change in scores: Both the Health Practices and Toileting/Diapering items on this subscale showed a downward trend from 2003 to 2017, with fewer programs scoring in the good to excellent range on both of these items over time.

Figure 6. Percentage of Abbott Preschool Program classrooms scoring in the good to excellent range across items on the Personal Care Routines subscale



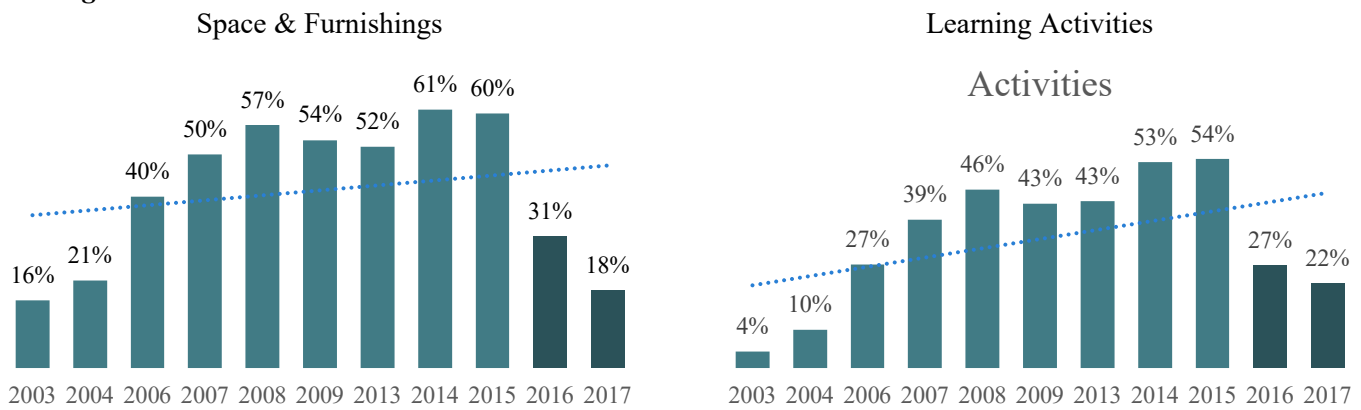
Note: The ECERS-3 measure does not include the “nap/rest” or “greeting/departing” items (*), so scores from 2016 and 2017 are not included here for those items.

Figure 7. Percentage of Abbott Preschool Program classrooms scoring in the good to excellent range on the Personal Care Routines subscale.



Finally, it is worth noting that while the Learning Activities and Space & Furnishings subscales showed consistent score increases over time with the use of the ECERS-R to measure quality, both subscales showed substantially diminished rates of classrooms scoring in the good to excellent range with the switch from the ECERS-R to the ECERS-3. On the Learning Activities subscale, just 4% of classrooms scored in the good to excellent range in 2003, and while this number climbed to 54% of classrooms in 2015, these numbers decreased to 27% and 22% of classrooms in 2016 and 2017, respectively. The Space & Furnishings subscales showed similar trajectories: While just 16% of classrooms were in the good to excellent range in 2003, this number reached 60% in 2015, yet fell to 31% in 2016 and 18% in 2017 with the switch to the ECERS-3.

Figure 8. Percentage of Abbott Preschool Program classrooms scoring in the good to excellent range on the Space & Furnishings and Activities subscales.



Summary

Through the use of longitudinal studies, researchers have demonstrated that the Abbott Preschool Program has a significant, positive impact on children’s later educational achievement (through grade 10) in language arts, mathematics and science (Barnett & Jung, 2021). In addition, the program has a positive impact on grade retention, with children who attended the preschool program being retained at a lower rate than children who did not. Along with these promising findings regarding child outcomes, the purpose of this study was to demonstrate changes in quality over time in Abbott preschool classrooms, as measured by two separate versions of one observational tool.

When looking historically at the goals for the Abbott preschool program, the focus on providing a *high-quality* preschool experience, through metrics including qualified teachers, small class sizes, and a developmentally appropriate curriculum, appears to be a worthwhile investment. By breaking down the average quality scores for classrooms in the Abbott preschool program over time, it is understandable why the program is having a lasting impact on children’s educational achievement. We found that from 2003 to 2015, when the ECERS-R was used to observe quality, the number of classrooms rated as good to excellent improved from 22% to 71% of those sampled; additionally, the number of classrooms rated as inadequate fell from 12% to 0%. When the observational tool used shifted from the ECERS-R to the ECERS-3 in 2016, the number of classrooms rated as moderate increased from 29% in 2015 to 61% in 2016; additionally, just 8% of classrooms were rated inadequate in 2017, while 32% were rated good to excellent.

The consistent upward trend of the proportion of classrooms scoring in the good to excellent range on the ECERS-R tool longitudinally seems to suggest that programs were able to utilize information from the scores obtained on the ECERS-R tool to make programmatic changes that improved quality, and that the high-quality goals of the NJ Abbott Preschool Program were being attained over time. Therefore, although a dip in percentage of classrooms scoring in the good to excellent range is evident with the switch to the ECERS-3 in 2016, the historical trend of increased quality over time as measured by the ECERS-R gives reason for optimism that classrooms will continue to progress in terms of quality ratings in the future as measured by a substantially different, more rigorous tool.

It is also important to recall that other researchers have noted that ECERS-3 scores are, on average, lower than ECERS-R scores. For example, Neitzel and a team of researchers (2019) demonstrated that there were significant differences in scores on all subscales of the ECERS-R and ECERS-3 when used simultaneously in the same classrooms, with the exception of the Personal Care Routines subscale. They additionally found that scores on the ECERS-R were significantly lower than scores on the ECERS-3. They found the tools to be only modestly correlated, because significant changes were made to the items on the ECERS-3, including different items and more difficult indicators that focus more heavily on teacher behavior rather than the provision of materials. Thus, although total scores and the proportion of classrooms scoring in the good to excellent range fell from 2015 to 2016, the adjustment to a revised tool to measure quality must be taken into account when interpreting the quality of Abbott Preschool Programs longitudinally. In addition, the updated version of the ECERS-3 provides a fresh perspective on areas that need improvement going forward.

Discussion & Recommendations

When looking at the individual subscales measured by the ECERS-R, in 2015, Interaction and Program Structure were the subscales with the highest scores, both with mean scores higher than 6. The Interaction subscale focuses on, among other things, general supervision of children, staff-child interactions, and interactions amongst children. Emotional support from teachers has consistently been found to be a predictor of positive outcomes for prekindergarten children, including greater social competence, fewer problem behaviors, and academic readiness skills in kindergarten (Brock & Curby, 2014; Palermo et al., 2007) Although the percentage of classrooms scoring in the good to excellent range on Interactions dropped from 92% of classrooms in 2015 to 62% of classrooms in 2017, it remains the subscale with the highest mean score (5.17 in 2017, or in the good to excellent range). Programs should continue to focus on promoting and supporting these high-quality interactions in the classrooms, as these relate to positive academic and socioemotional outcomes for children. Scores on the Program Structure subscale, which includes items such as schedule, free play, and transitions, were similarly strong through 2015. In addition, the mean score for Program Structure in 2017 was 5, that is, programs should continue to build on their strengths in setting up routines and structures.

There are other subscales of the ECERS-3 in which classrooms are, on average, of moderate quality; indicators of quality as measured by these subscales should be an area of focus for the future. For example, in 2015 the mean score on the Personal Care Routines subscale was 4.36; in 2017, the mean score was 4.26. Of all subscales, these scores showed the least change over time: most classrooms have consistently scored in the moderate or inadequate range on this subscale. Items on the Personal Care Routines subscale cover health and safety practices and routines including toileting/diapering. Of these, fewer than one-third of classrooms were rated in the good to excellent range on items covering meals and snacks and toileting and diapering, and less than half scored in this range on health practices. Other studies using the ECERS-3 and the ECERS-R across a large number of classrooms have also documented relatively low scores on the Personal Care Routines subscale (e.g., Cassidy et al., 2005; Setodji et al., 2017), and compliance with health and safety practices in child care settings is variable. For example, Clark and a team of researchers (2016) found that caregivers in a sample of early childhood centers in Arkansas washed their hands after just 30% of required handwashing events. Safety practices are an important metric of quality: a review conducted from 1996-2005 found that 2.1 million playground injuries occurred in this timeframe, and 27% of those injuries happened to children younger than 5 (Vollman et al., 2009). The National Association for the Education of Young Children (NAEYC) includes Health and Physical Environment as two of their 10 standards for early childhood programs because of the critical importance of ensuring early childhood environments are safe and high-quality health practices are used and modeled for children. A focus on health and safety practices such as hand washing and proper sanitation procedures, and the removal of indoor and outdoor hazards, should be receive attention.

While the average score on the Learning Activities subscale was in the good to excellent range the final year the ECERS-R was used in 2015 (5.12), with the switch to the ECERS-3, the mean score on this subscale was 4.22 in 2016 and 4.02 in 2017. Less than one-third of programs scored in the good to excellent range on this subscale on the items Music/Movement, Blocks, and Nature/Science. The ECERS-3 emphasizes much more the role of the teacher in supporting children's use of materials, and expanded observing staff support for logical concepts with more items on Math. All this signifies that programs should address the learning materials required by each of these items as well as the interactions teachers have with students while engaged with these materials in order to address these lower scores and enhance children's developmental experiences in the classroom.

In addition, the Language and Literacy subscale (formerly known as Language-Reasoning) showed some improvement over time when measured by the ECERS-R (the mean score was 4.42 in 2003 and 5.11 in 2015); however, mean scores in 2015 and 2016 were 4.86 and 4.70. The Language and Literacy subscale focuses on the use of print and spoken language with children, and the ways in which teachers encourage children to expand their vocabularies and use language in a meaningful way. This instructional skill is important: Researchers find that strategies such as shared book reading and targeted vocabulary instruction are associated with language skills in preschoolers (Leung, 2008; Neuman et al., 2011). Although scores are approaching the good to excellent range, classrooms are on average demonstrating moderate quality in this area which suggests the need to expand the use of instructional practices that promote language and literacy skills is an area to focus on.

It is important to note the wide range of scores recorded – both for the total ECERS-3 tool, and for each subscale. In the most recent year of data collection (2017), scores on five of the six subscales ranged from 1 to 7 on three of the six subscales, with the other subscales showing almost as much variation. This indicates that there is a wide range of quality observed in the system as a whole; and more importantly, it suggests that with adequate support, classrooms scoring in the

inadequate range can make significant improvements over time and that reducing variability may require targeted supports for programs with lower observed levels of quality.

Finally, it is recommended that programs continue to reflect on the information provided by the scores in the ECERS-3 or similar tools. Our findings demonstrate that an increase in quality of Abbott Preschool Programs over time (as measured by the ECERS-R) is feasible in a preschool program. Prior research indicates that initiatives using targeted professional development centered on teachers' ECERS scores impact future classroom quality (Helmerhorst et al., 2017; Hooks et al., 2006): The observed programs should continue to use the feedback provided by ongoing observations (using this or other tools) to continuously improve quality in preschool programs and effectively fund the necessary supports and workforce (i.e., coaches) for such continuous improvement to occur.

References

- Barnett, W. S., & Frede, E. C. (2017). Long-term effects of a system of high-quality universal preschool education in the United States. In H-P. Blossfeld, N. Kulic, J. Skopek, M. Triventi (Eds.), *Childcare, early education and social inequality: An international perspective* (pp. 152-172). Edward Elgar Publishing.
- 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.
- Barnett, W. S., Jung, K., Youn, M., Frede, E. C. (2013). Abbott preschool program longitudinal effects study: Fifth grade follow-up. New Brunswick, NJ: National Institute for Early Education Research. Retrieved from <http://nieer.org/wp-content/uploads/2014/09/APPLES205th20Grade.pdf>
- Brock, L. L., & Curby, T. W. (2014). Emotional support consistency and teacher-child relationships forecast social competence and problem behaviors in prekindergarten and kindergarten. *Early Education and Development, 25*(5), 661-680. <https://doi.org/10.1080/10409289.2014.866020>
- Brunsek, A., Perlman, M., Falenchuk, O., McMullen, E., Fletcher, B., & Shah, P. S. (2017). The relationship between the Early Childhood Environment Rating Scale and its revised form and child outcomes: A systematic review and meta analysis. *PLoS ONE, 12*(6), e0178512. <https://doi.org/10.1371/journal.pone.0178512>
- Cassidy, D., Hestenes, L., Hedge, A., Hestenes, S., & Mims, S. (2005). Measurement of quality in preschool child care classrooms: An exploratory and confirmatory factor analysis of the early childhood environment rating scale-revised. *Early Childhood Research Quarterly, 20*(3), 345- 360. doi: 10.1016/j.ecresq.2005.07.005
- Clark, J., Henk, J. K., Crandall, P. G., Crandall, M. A. & O'Bryan, C. A. (2016). An observational study of handwashing compliance in a child care facility. *American Journal of Infection Control, 44*(12), 1469-1474. doi: 10.1016/j.ajic.2016.08.006
- Early, D. M., Sideris, J., Neitzel, J., LaForett, D. R., & Nehler, C. G. (2018). Factor structure and validity of the Early Childhood Environment Rating Scale – Third Edition (ECERS-3). *Early Childhood Research Quarterly, 44*, 242-256 <https://doi.org/10.1016/j.ecresq.2018.04.009>
- Farrie, D. (2014). The Abbott Preschool Program: A 15-year progress report. Education Law Center. <https://edlawcenter.org/assets/files/pdfs/publications/AbbottPreschool15YearProgressReportMay2014.pdf>
- Friedman-Krauss, A. H., Barnett, W. S., Hodges, K. S., Garver, K. A., Weisenfeld, G., Gardiner, B. A., Jost, T. M. (2023). The State of Preschool 2022: State Preschool Yearbook. New Brunswick, NJ: National Institute for Early Education Research. <https://nieer.org/the-state-of-preschool-yearbook-2022#profiles>
- Frede, E. (2005). *Assessment in a continuous improvement cycle: New Jersey's Abbott Preschool Program*. Ewing, NJ: The College of New Jersey.
- Frede, E., Lamy, C. E., Seplocha, H., Strasser, J., Jambunathan, S., Juncker, J. A., & 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
- Harms, T., Clifford, R. M., & Cryer, D. (1998). *Early childhood environment rating scale*. Teachers College Press. New York, NY.
- Harms, T., Clifford, R. M., & Cryer, D. (2015). *Early childhood environment rating scale (3rd ed.)*, Teachers College Press. New York, NY.
- Hestenes, L. L., Rucker, L., Wang, Y. C., Mims, S. U., Hestenes, S. E., & Cassidy, D. J. (2019). A Comparison of the ECERS-R and ECERS-3: Different Aspects of Quality?. *Early Education and Development, 30*(4), 496-510.

- Helmerhorst, K. O. W, Fukkink, R. G., Riksen-Walraven, J. M. A., Gevers Deynoot-Schaub, M. J. J. M., & Tavecchio, L. W. C. (2017). Improving quality of the child care environment through a consultancy programme for center directors. *International Journal of Early Years Education, 25*(4), 361-378. doi:10.1080/09669760.2017.1321528.
- Hooks, L. M., Scott-Little, C., Marshall, B. J., & Brown, G. (2006). Accountability for quality: One state's experience in improving practice. *Early Childhood Education Journal, 33*, 399-403. DOI: 10.1007/s10643-006-0065-3
- Moiduddin, E., Aikens, N., Tarullo, L., West, J., and Xue, Y. (2012). *Child Outcomes and Classroom Quality in FACES 2009* (OPRE Report 2012-37a). Alexandria, VA: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.
- National Center for Education Statistics (2016) Early childhood longitudinal program (ECLS): data collection procedures. Available at: <http://nces.ed.gov/ecls/birthdataprocure.asp>
- Neitzel, J., Early, D., Sideris, J., LaForrett, D., Abel, M. B., Soli, M., Davidson, D. L., Haboush-Deloye, A., Hestenes, L. L., Jenson, D., Johnson, C., Kalas, J., Mamrak, A., Masterson, M. L., Mims, S. U., Oya, P., Philson, B., Showalter, M., Warner-Richter, M., & Wood, J. K. (2019). A comparative analysis of the Early Childhood Environment Rating Scale-Revised and Early Childhood Rating Scale, Third Edition. *Journal of Early Childhood Research, 17*(4), 408-422. <https://doi.org/10.1177/1476718X19873015>
- Neuman, S. B., Newman, E. H., & Dwyer, J. (2011). Educational effects of a vocabulary intervention on preschoolers' word knowledge and conceptual development: A cluster-randomized trial. *Reading Research Quarterly, 46*(3), 249-272. <http://www.jstor.org/stable/41228653>
- Palermo, F., Hanish, L. D., Martin, C. L., Fabes, R. A., & Reiser, M. (2007). Preschoolers' academic readiness: What role does the teacher-child relationship play? *Early Childhood Research Quarterly, 22*, 407-422. doi:10.1016/j.ecresq.2007.04.002
- Perlman, M., Zellman, G. L., & Le, V. (2004). Examining the psychometric properties of the Early Childhood Environment Rating Scale – Revised (ECERS-R). *Early Childhood Research Quarterly, 19*(3), 398-412. <https://doi.org/10.1016/j.ecresq.2004.07.006>
- Setodji, C. M., Schaack, D., & Le, V. (2017). Using the Early Childhood Environment Rating Scale-Revised in high stakes contexts: Does evidence warrant the practice? *Early Childhood Research Quarterly, 42*, 158-169. doi: 10.1016/j.ecresq.2017.10.001
- Whitebrook, M., Ryan, S., Kipnis, F., & Sakai, L. (2008). *Partnering for preschool: A study of center directors in New Jersey's mixed-delivery Abbott Program*. Center for the Study of Childcare Employment: Berkley, CA.