Longitudinal Study *of* Georgia's Pre-K Program

Final Report: Pre-K through 4th Grade

UNC
FRANK PORTER GRAHAM CHILD DEVELOPMENT INSTITUTE

Georgia Dept of Early Care and Learning
BRIGHT FROM THE START

Acknowledgments

We wish to acknowledge the members of our Georgia's Pre-K Program Evaluation Team who assisted with the 2018-2019 study, including the previous study PI, Ellen Peisner-Feinberg; the research staff: Karen Van Manen, Jada Walker, David Bogojevich, Shannon Casey, Elizabeth Giraud Love, Elizabeth Gunn, Cristina Luna Evans, Diana Knechtel, Tom Leggett, Jennifer Osborne, Judy Owen, Eric Savage, Stefanie Schwartz, Gina Walker; and data collectors Brita Alton, Jane Argo, Joan Boudousquie, Elizabeth Bragg, Brianna Castellano, Kathy Crumley, Akua Ducard, CL Dunn, Deborah Hallman, Stacey Harris, Khadija Highsmith, Melissa Johns, Kathy Crumley, Deborah Laing, Kim Lane, LaJuana Johnson-Mendez, Elizabeth Milling, Jacquelyn Parker, Katie Rutland, Carolina Sandoval, Steven Schell, Melissa Silva, Patricia Simpson, Moneesha Smith, Claudia Sorto, Karen Stewart, Alissa Sypsa, Peggy Tymes, Adriana Vivas-Sosa, Lori Warren; and the previous analysis team: Irina Mokrova and Margaret Burchinal.

In addition, we share our appreciation to all those who participated in and assisted with this study, including the children and families; the teachers, administrators, and other staff of Georgia's Pre-K Program; the kindergarten, 1st, 2nd, 3rd, and 4th grade classrooms and schools; and the staff of Bright from the Start: Georgia Department of Early Care and Learning (DECAL).

Report design by Hollis Oberlies, Purple Zante, Inc. Image on p. 8 by Melanie Busbee Photo.

This study was funded by Bright from the Start: Georgia Department of Early Care and Learning. The opinions expressed in this report do not necessarily reflect those of the funding agency.

Suggested citation: Soliday Hong, S., Zadrozny, S., Walker, J., Love, E.N.G., Osborne, J.D., Owen, J. L., & Peisner-Feinberg, E. (2023). *Longitudinal Study of Georgia's Pre-K Program Final Report: Pre-K through 4th Grade.* Chapel Hill, NC: The University of North Carolina, FPG Child Development Institute.

This report is available at http://fpg.unc.edu/projects/georgia-pre-kindergarten-evaluation or at http://www.decal.ga.gov/BftS/EvaluationGAPreKProgram.aspx.

Table of Contents

List of Tables

List of Figures

EXECUTIVE SUMMARY

Final results from the Georgia's Pre-K Longitudinal Study are summarized below. The study followed a statewide representative sample of children (n=1,169) from their pre-k experience (2013-2014) through the 4th grade (2018-2019). The study was conducted by researchers at the Frank Porter Graham (FPG) Child Development Institute at the University of North Carolina at Chapel Hill.

Study Components

- Standardized child assessments that measure skills across learning domains (by using standardized child assessments, children's scores can be compared to a nationally representative sample of children of the same age)
- Observations of classroom quality over time that measure the quality of teacher-child interactions
- Inclusion of a subsample of dual language learners (English-Spanish)
- Comparison sample of children who did not attend any pre-k added in the 3rd grade year and followed through 4th grade

Key Results from the Longitudinal Study

- The largest gains (higher than expected scores relative to the norming sample) were observed during children's Georgia's Pre-K year and 4th grade school years with small to moderate gains found in measures assessing literacy, math, and social skills.
- Initial gains persisted on most outcomes through kindergarten, with scores starting to level off in 1st grade and then decreasing or stabilizing through 3rd grade, followed by small increases in 4th grade in literacy, math, social skills, and behavior problems.
- A change in standard scores can reflect the influence of school experience above and beyond the expected typical growth for age over time. For 5 of the 7 assessments conducted from pre-k through 4th grade, children's scores were higher in 4th grade than at the pre-k or 1st grade baseline likely reflecting a cumulative benefit of pre-K and elementary grade instruction.
- Children's language scores, on average, were below the national norm throughout the study.
- Scores for children classified as Dual Language Learners (DLL) were slightly below the national norm in 4th grade for skills measured in English and well below the national norm for skills measured in Spanish. DLLs had the lowest levels of behavior problems throughout the study.
- The child characteristic most strongly associated with small differences in outcomes over time included English Language Proficiency at pre-k entry, Individualized Education Plan (IEP) status at the beginning of pre-k, and children's gender.
- Classroom quality, as measured by the Classroom Assessment Scoring System (CLASS) Pre-K, K-3, and Upper Elementary versions, was in the moderate range, on average; was highest in pre-k and 4th grade; and was associated with small gains in child skills in pre-k and 4th grade.

Key Results from the Comparison Sample

Children who attended Georgia's Pre-K Program had better language, literacy, and executive function skills in 3rd and 4th grades than children whose parents reported that the child did not attend any pre-k program (comparison group). These results are similar to the findings of the Longitudinal Study, where children who attended Georgia's Pre-K had higher scores in foundational literacy skills relative to the national norming sample. Together, these results suggest that foundational literacy skills, which are a focus of pre-k, were not obtained by children in the comparison sample.

Study Implications

- Overall, results from the study show that Georgia's Pre-K Program prepares students to enter kindergarten.
- While students in the study are at or above the national average at the end of 4th grade on most measures, the rate of growth in pre-k and kindergarten is not maintained in later grades (except for larger growth in social skills and math in 4th grade).
- A key strength of Georgia's Pre-K Program is building foundational literacy skills for reading.
- Vocabulary development is an area of growth for children ages birth to five, but also kindergarten to 4th grade, and opportunities to support language development are warranted.
- While classroom quality is highest in pre-k, quality is still in the moderate-to-high range throughout K-4th grades.
- This longitudinal study reflects a deep investment in ensuring that children fully benefit from enrollment in Georgia's Pre-K along with aligned and supportive K-4 experiences. The study also reflects a commitment to continuous improvement through an established state-university partnership.

LONGITUDINAL STUDY OF GEORGIA'S PRE-K PROGRAM THROUGH 4TH GRADE

Georgia's Pre-K Program is a state-funded universal pre-kindergarten program for four-year-olds that serves more than 80,000 children from all income levels each year.

The purpose of this evaluation study was to examine children's longitudinal academic and social outcomes associated with attendance in Georgia's Pre-K Program and to examine the quality of the classrooms attended. This report covers pre-k to 4th grade outcomes collected between the 2013–2014 and 2018–2019 school years as well as a sub-study comparing 3rd and 4th grade outcomes of children who attended Georgia's Pre-K Program with children who did not attend any pre-k program.

Primary Evaluation Questions

1. What are the longitudinal outcomes through 4th grade for children who attended Georgia's Pre-K Program?
2. What factors predict better longitudinal outcomes for children?
3. What is the quality of children's instructional experiences from pre-k through 4th grade?
4. Are there differential long-term outcomes for children with and without Georgia's Pre-K experience?

To address these questions, the evaluation study included a sample of 1,169 children (139 Spanish-speaking dual language learners/DLL) attending a random sample of 199 Georgia Pre-K classrooms in the first year of the study. These children were followed into 4th grade in the sixth year of the study. A comparison sample of 333 children (80 Spanish-speaking DLL) who did not attend Georgia's Pre-K Program were added to the study in 3rd grade (year 5), and 254 children (57 Spanish-speaking DLL) were followed into 4th grade in year 6.

Professional assessors conducted individual child assessments near the beginning and end of each school year. The assessment measures covered multiple areas of learning, including language, literacy, math, executive function, and teacher ratings of behavior skills. For the DLL subsample, parallel assessments were conducted in both English and Spanish. Researchers also conducted observations in a subset of children's classrooms each year from pre-k through 4th grade. In addition, parents and teachers were surveyed and some administrative pre-k data were used.

Overview of Georgia's Pre-K Program

Georgia's Pre-K Program is a state-funded universal pre-kindergarten program for four-year-olds from all income levels that serves more than 80,000 children each year. Georgia's Pre-K Program was established in 1992 and expanded in 1995, making Georgia one of the first states to offer a universal program. The following were key components of the program as of the 2013-14 school year when the study was initiated:

- Pre-K classrooms were located in a variety of settings: public school systems, private providers, and blended Head Start/pre-k classrooms.
- Georgia's Pre-K was based on a school-year model with instruction for 180 days/year and 6.5 hours/day.
- Class sizes were limited to 20–22 children with a lead and assistant teacher, resulting in adult: child ratios of 1:11.
- Lead teachers were required to have at least a bachelor's degree in early childhood education or a related field or a bachelor's degree in any field along with an approved early childhood education credential.

Prior to 2011-2012, Georgia's Pre-K Program provided 180 instruction days per year, but budget restrictions led to a reduction to 160 days in 2011-2012. In 2012–2013, the program year was increased to 170 days and in 2013–2014, it was returned to 180 days.

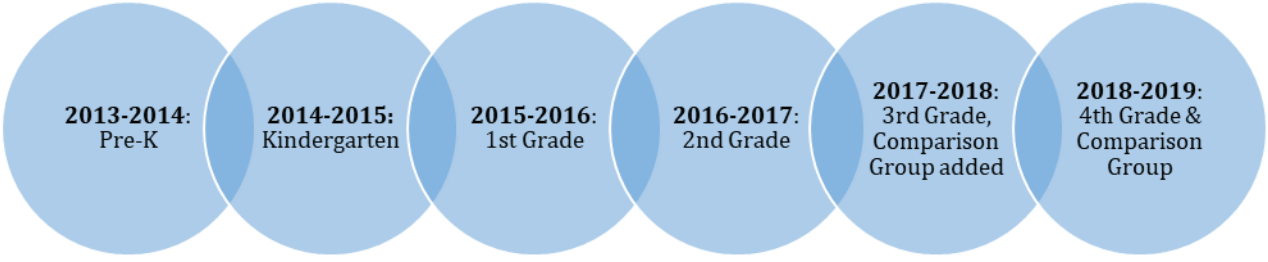
- Assistant teachers were required to have at least a Paraprofessional Certificate (issued by the Georgia Professional Standards Commission) or a Child Development Associate (CDA) credential.
- Program guidelines provided minimum salary requirements for lead teachers based on credentials, with funding provided by the state, as well as minimum salary requirements for assistant teachers meeting the credential requirements.

Previous Evaluation Studies of Georgia's Pre-K Program

Starting in 2011, the Georgia legislature funded a series of studies to evaluate the impact of Georgia's Pre-K Program on children's outcomes over time.

1. **Pre-K Outcomes Study,** conducted in 2011-2012, was designed to examine children's learning outcomes during pre-k, the factors that predicted better short-term outcomes, and the quality of children's experiences in Georgia's Pre-K classrooms. The Pre-K Outcomes study included 509 children recruited from a random sample of 100 pre-k classrooms. All children, including dual language learners, exhibited significant growth during their pre-k year across all domains of learning, including language and literacy skills, math skills, general knowledge, and behavioral skills. Classroom quality was in the moderate-to-high range.
2. **Kindergarten Comparison Study**, conducted in 2012-2013, was a quasi-experimental design (QED) study to investigate the effects of participation in Georgia's Pre-K on children's school readiness skills compared to eligible children who had not yet attended the program. This QED study utilized a regression discontinuity design (RDD) and included 1,181 children (611 children who attended pre-k and 570 children who had not yet attended pre-k). Children who attended Georgia's Pre-K had better language and literacy, math, and general knowledge skills during kindergarten than children who did not attend. No differences were observed between groups on social skills or behavior problems.
3. **Longitudinal Study**, occurring from 2013-2019, was designed to follow a sample of 1,169 children from pre-k through 4th grade, in order to examine the short- and long-term learning outcomes for children who attended Georgia's Pre-K Program, as well as the quality of their preschool and school experiences (see Figure 1). Previous reports showed that children had gains in language, literacy, math, general knowledge, and social skills during their pre-k, kindergarten, and 1st grade years. For a few measures (math problem-solving, vocabulary, reading comprehension), scores decreased slightly during the 1st grade year and again in the 2nd and 3rd grade years. For more advanced reading and math skills assessed only in 1st through 4th grades, standardized scores also showed decreases over time with children continuing to score in the expected range for their age.

FIGURE 1. Longitudinal Study Timeline



B

STUDY APPROACH

The study followed 1,169 children enrolled in 199 Georgia's Pre-K classrooms through the 4th grade, examining patterns of growth in language, literacy, math, and executive function skills, along with pre-k through 4th grade classroom quality.

This 4th grade/year 6 report includes data collected with a longitudinal cohort of children from Georgia's Pre-K through their 4th grade year (see Table 1). In year 1 (2013-14) data were gathered from a random sample of classrooms and children within classrooms to examine child outcomes and classroom quality in Georgia's Pre-K Program. This random sample of children was then followed into kindergarten during year 2 (2014-15), 1st grade during year 3 (2015-16), 2nd grade during year 4 (2016-17), 3rd grade during year 5 (2017-18), and 4th grade during year 6 (2018-19). During year 5, a sub-study was started and an additional n=333 children with no previous pre-k experience were recruited from the third grade classrooms of children enrolled in the original longitudinal cohort. At the beginning (fall) and end (spring) of each year, researchers conducted individual assessments of children's language and academic skills and gathered teacher ratings of behavior skills. Researchers also conducted classroom observations of teacher-child instructional interactions each year. Program and classroom characteristics, as well as teacher and child demographic data, were obtained from annual teacher and parent surveys.

Figure 2 provides an overview of the study activities in the children's 4th grade year. The following sections describe the technical details related to participant recruitment and data collection procedures.

FIGURE 2. Overview of Study Activities in 4th Grade

FALL 2018	SPRING 2019
• Re-recruitment of Longitudinal and Comparison Groups • Fall child assessments • Teacher survey of children's social skills	• Classroom observations • Spring child assessments • Teacher survey of children's social skills and teacher questionnaire

Recruitment

Children

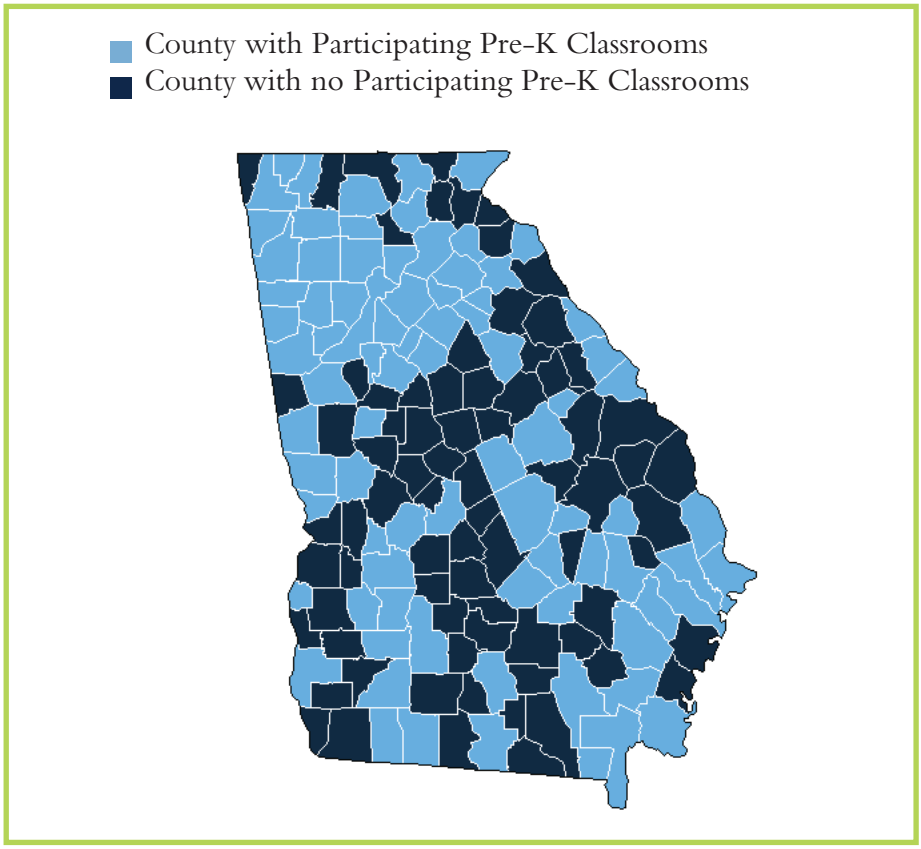
The original longitudinal sample of children who attended Georgia's Pre-K Program were recruited in the first year of the study (2013-14). Parent permission forms were distributed to all children in 199 randomly selected pre-k classrooms, with an overall permission rate of 73% (3,136 of 4,270 eligible children). From all returned permission forms, an average of six children per classroom were randomly selected for inclusion in the study.

Districts, Schools, and Teachers

Every fall, the research team began a process to recruit school districts, principals, and teachers where study participants attended school after pre-k. If the team was not able to get approval from the school district and the school principal, the participating student was not assessed. In kindergarten 104 districts, 428 principals, and 793 teachers agreed to participate in the study. In 1st grade, children were enrolled in 104 districts, with agreement to participate in the study and 421 principals, and 748 teachers agreed to participate. In 2nd grade 111 districts, 451 principals, and 745 teachers from 786 classrooms participated. In 3rd grade 104 districts, 432 principals, and 640 teachers agreed to participate. In 4th grade, the research team obtained study approval from

110 of 117 district superintendents for school districts where study participants attended 4th grade. From those districts, 394 school principals gave approval for schools and teachers to participate in the study in 4th grade. Of the 733 classrooms in which study children were enrolled, 693 teachers gave consent to participate. Participating schools were located throughout the state of Georgia regionally and included both rural and urban areas. See Figure 3 for a map of the districts that participated in pre-k and 4th grade. (Charter and private schools are not listed to protect anonymity of the study participants.)

FIGURE 3. Statewide Distribution of Participating Pre-K Classrooms



Participants

Teachers and Classrooms

Children in the original Georgia Pre-K Longitudinal Study sample initially attended 199 Georgia Pre-K classrooms in year 1 (2013–14), 822 kindergarten classrooms in year 2 (2014–15), 777 1st grade classrooms in year 3 (2015-16), 786 2nd grade classrooms in year 4 (2016-17), 718 3rd grade classrooms in year 5 (2017-18), and 754 4th grade classrooms in year 6 (2018-19) (see Table 1). About half of the Georgia Pre-K programs attended by children in the study sample were in public school settings (49%) and about half were in private sites (51%). As children were followed over time, the elementary school classrooms attended by children in the study were primarily located in public school settings (98% in 4th grade; 3rd=98%; 2nd=98%; 1st=98%; K=97%; PK=49%), with a few in charter schools and private schools (~1% in 4th grade; 3rd=1%; 2nd=2%; 1st=2%; K=2%; PK=51%).

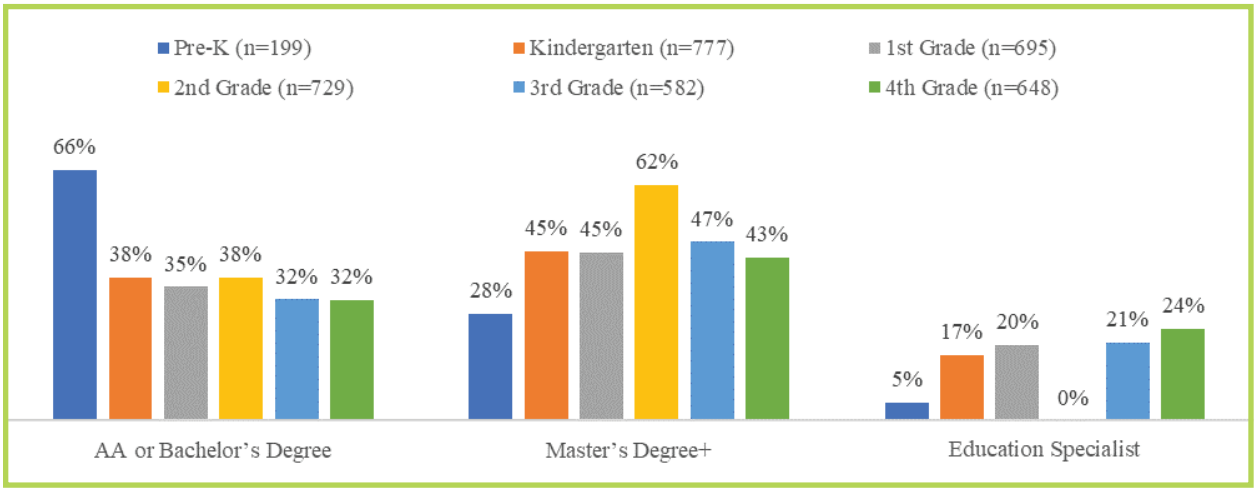
TABLE 1. Number of Classrooms and Children Participating in the Longitudinal Study

Study Year	Total Classrooms	Total Children in Longitudinal Study	DLL – Longitudinal Study
Pre-K (2013-14)	199	1,169	139
Kindergarten (2014-15)	822	1034 (88%)	118 (85%)
Grade 1 (2015-16)	777	969 (83%)	119 (86%)
Grade 2 (2016-17)	786	951 (81%)	116 (83%)
Grade 3 (2017-18)	718	857 (73%)	109 (78%)
Grade 4 (2018-19)	754	715 (61%)	72 (52%)

Note: Percentages noted in parentheses are the proportion of the original sample that was retained in the study by grade.

Classroom and teacher characteristics were collected each year through teacher surveys, as well as state administrative data from Bright from the Start: Georgia Department of Early Care and Learning (DECAL) in pre-k (see Appendix Table A1). The average class size was similar each year of the study (about 20-23 children, split equally among boys and girls). In pre-k, almost two-thirds (66%) of the teachers held an associate or bachelor's degree and about one-third (34%) held a master's degree, education specialist certification, or higher. In elementary grades, this pattern was reversed, with about one-third having an associate or bachelor's degree (K=38%, 1st=36%, 2nd=38%, 3rd= 32%, 4th=32%) and around two-thirds having a master's degree, education specialist certification, or higher (K=62%, 1st=64%, 2nd=62%, 3rd= 67%, 4th=68%) (see Figure 4). The majority (82%) of pre-k teachers were certified by the Georgia Professional Standards Commission. Teachers reported a substantial number of years of teaching experience on average (PK=11, K=15, 1st=14, 2nd=14, 3rd=13, 4th=14), including around half this time teaching at their current grade level (PK=6, K=9, 1st=7, 2nd=6, 3rd= 6, 4th=6). The majority of teachers were White (PK=67%, K=77%, 1st=70%, 2nd=72%, 3rd=73%, 4th=75%); Black teachers comprised almost one-fifth to one-quarter of teachers (PK=27%, K=18%, 1st=23%, 2nd=23%, 3rd= 19%, 4th=21%); a small proportion were of Hispanic/Latino ethnicity (PK=3%, K=2%, 1st=2%, 2nd=2%, 3rd= 3%, 4th=2%); and nearly all were female (PK=98%, K=98%, 1st=99%, 2nd=97%, 3rd= 95%, 4th=93%).

FIGURE 4. Teacher Education and Certification by Grade Level



Children

The final study sample consisted of 1,169 children enrolled in Georgia's Pre-K, with a DLL subsample of 139 children. Longitudinal analyses were conducted on the Georgia's Pre-K sample to estimate gains over time.

Georgia's Pre-K Sample

The longitudinal pre-k sample included 1,169 children in year 1 (pre-k), 1,034 children in kindergarten (88% of the original sample), 969 children in 1st grade (83% of the original sample), 951 children in 2nd Grade (81% of the original sample), 857 children in 3rd grade (73% of the original sample), and 715 children in 4th grade (61% of the original sample). At the beginning of the pre-k year children were 4.5 years old on average and 9.5 years old at the beginning of the last year of the study. Information about child and family characteristics for the study sample were obtained through parent surveys in pre-k (see Appendix Table A2). Across years 49-52% of the sample were girls and 15% Hispanic. The children were from varied racial/ethnic backgrounds which were fairly consistent across years of the study: majority White (42-53% across years) and Black/African-American (35-38%). Nine to ten percent of the children had limited English language proficiency across the study years, and 3% had an individualized education plan (IEP) at the beginning of pre-k. The education level for the majority (76%) of children's primary caregivers was between a high school diploma and less than a bachelor's degree, with 28% having a bachelor's degree or above, and 10% who had less than a high school diploma or GED by 4th grade.

Income categories were identified for pre-k participants and were provided to the study team from administrative data. Slightly more than half (53-54%) of the children were from low-income families as designated at pre-k entry as indicated by Pre-K Category One status (which represented participation in one or more programs including SNAP, TANF, SSI, CAPS, Medicaid, and free or reduced-price meals). Students with Category Two status were not income-eligible for the programs listed above.

The retention rate of participants was strong through 3rd grade when the study was initially scheduled to be completed. The study was extended to examine an additional year of children's

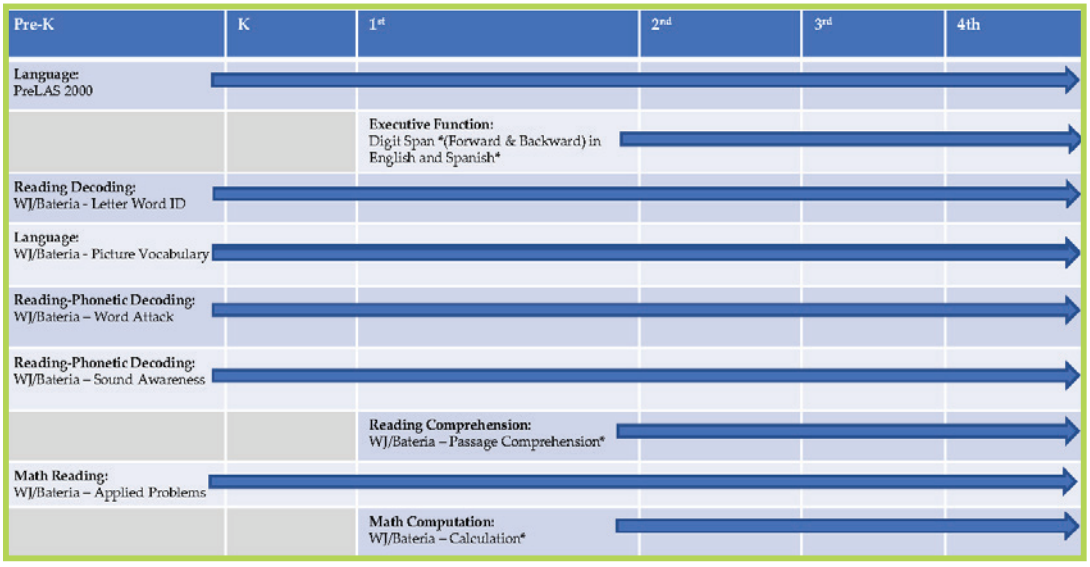
growth past 3rd grade. Children were re-recruited into the study in 4th grade which resulted in slightly lower retention rate in 4th grade. Previous longitudinal studies of other pre-k programs have had similar or lower retention rates (e.g., New Jersey Abbott Pre-K follow-up had an overall retention rate of 70% in 3rd grade; 65% overall retention rate in Tulsa Pre-K study at 3rd grade; and 36% retention rate in the Tennessee Pre-K Study at 3rd grade). These children included a subsample of Spanish-speaking dual-language learners (DLL subsample)—139 children in year 1 (pre-k), 118 children in kindergarten (85% of the original sample), 119 children in 1st grade (86% of the original sample), 116 children in 2nd grade (83% of the original sample), 109 children in 3rd grade (78% of the original sample), and 72 children in 4th grade (52% of the original sample).

Child Assessments

Individual assessments to measure children's growth in skills were conducted in classrooms each year. Children who attended Georgia's Pre-K were assessed in English at 12 time points, in the fall and spring (near the beginning and end of each year) from pre-k through 4th grade. In addition to parental permission, children's verbal assent was obtained before the assessments began. Children whose home language was Spanish also received a set of parallel assessments using Spanish language versions of these measures. The assessments in Spanish were conducted by a different (bilingual) data collector approximately two weeks after the English assessments.

Measures of language, literacy, and math skills were gathered from pre-k through 4th grade using the Woodcock-Johnson III: Tests of Achievement (Woodcock, McGrew, & Mather, 2001)[v] and the Batería III Woodcock-Muñoz: Pruebas de aprovechamiento (Batería III; Muñoz-Sandoval, Woodcock, McGrew, & Mather, 2005).[vi] The Forward and Backward Digit Span tasks were adapted from Gathercole & Pickering (2001).[vii] In 1st grade, more developmentally advanced measures of reading comprehension, number operations, and executive function were added to the battery (see Table 2 and Appendix Table B1 for listings of measures and years of administration). All child assessment measures were available in English and Spanish. Most of the measures used were norm-referenced, and therefore, standard scores could be used for these measures. These scores consider a child's age, so that the standardized mean score of 100 represents the expected performance for an average child at a given age.

TABLE 2. Constructs/Measures and Years Administered in the Longitudinal Study



Classroom Observations

Observations of classroom practices were conducted in classrooms attended by children in the sample each year. When the children were in pre-k, observations were conducted in all 199 randomly selected pre-k classrooms. In kindergarten, a random sample of classrooms was selected for observation. In subsequent years, the children enrolled in classrooms that were observed in kindergarten were followed, and their classrooms were observed in order to obtain a longitudinal sample associated with the classroom observations. In kindergarten, observations were conducted in 196 classrooms of the 822 attended by children in the sample (representing 434 children in the sample). In 1st grade, observations were conducted in 296 classrooms of the 777 attended by children in the sample (representing 447 children in the sample). In 2nd grade, observations were conducted in 280 classrooms of the 786 attended by children in the sample (representing 403 children in the sample). In 3rd grade, observations were conducted in 256 of the 706 classrooms attended by children in the sample (representing 496 children in the longitudinal sample). In 4th grade, observations were conducted in 228 of the 754 classrooms attended by children in the sample (representing 715 children in the longitudinal sample). These numbers are summarized below in Table 3. Classroom observations were conducted annually during the second half of the school year. Observations typically lasted 2-3 hours. Data collectors completed standard training procedures offered by the developers of the measures, had additional field practice, and met established reliability criteria before gathering data.

The classroom observations used the same measure each year to examine the quality of teacher-child instructional interactions, with appropriate versions for the age range of children. The Classroom Assessment Scoring System (CLASS) Pre-K Version (Pianta, La Paro, & Hamre, 2008)[ix] was used in pre-k, the CLASS K-3 (Pianta, La Paro, & Hamre, 2008)[x] was used in kindergarten through 3rd grade, and the CLASS Upper Elementary was used in 4th grade (Pianta, Hamre, & Mintz, 2012).[xi] Using these three grade-based versions of the CLASS, we measured teachers' interactions with children in the areas of social and emotional functioning, classroom organization and management, and instructional interactions, with children organized into three domains: Emotional Support, Classroom Organization, and Instructional Support. Total scores were also calculated.

TABLE 3. Total Number of Classrooms and Subset of Observed Classrooms Attended by Participating Children in the Longitudinal Study

Study Year	Total Classrooms	CLASS Observations
Pre-K (2013-14)	199	199
Kindergarten (2014-15)	822	196
Grade 1 (2015-16)	777	296
Grade 2 (2016-17)	786	280
Grade 3 (2017-18)	718	256
Grade 4 (2018-19)	754	228

Parent and Teacher Surveys

Parents completed demographic surveys each fall (pre-k to 3rd grade) about their family and household. The annual fall parent survey included information about parent education levels, income, languages spoken, and number of adults/children in the home. Parent surveys were received from 91% (1,067/1,169) of participating families in pre-k, 86% (888/1,034) of participating families in kindergarten, 85% (821/969) in 1st grade, 87% (832/951) in 2nd grade, and 71% (609/857) in 3rd grade. Fourth-grade parent surveys were distributed in the spring and included questions about parent engagement in their child's class and school activities/fundraising, the "school program fit," their child's behaviors, school safety, and where the child would attend school the following year. These parent surveys were sent directly to the parents' email or mailing address on file and were returned by email or postage-paid return envelopes. Spring surveys were sent to all participating children on record and were received from 68% (708/1040) of participating families in 4th grade.

Teachers completed surveys in the spring that included information about their classrooms and their background, including classroom composition (e.g., number of boys and girls in class and languages spoken by those children); years of teaching experience and degrees earned; available classroom resources; level of perceived control over what is taught and how; and levels of parent involvement and support. Teacher survey completion rates were 95% (189/199) in pre-k, 95% (777/822) in kindergarten, 91% (707/777) in 1st grade, 92% (727/786) in 2nd grade, 84% (590/699) in 3rd grade, and 90% (640/708) in 4th grade.

Comparison Group Sub-Study

The comparison group sub-study was planned to provide a relevant comparison group to answer Research Question 4: "Does academic performance in 3rd and 4th grade differ between children who enrolled in Georgia's Pre-K compared to children who had no pre-k experience prior to kindergarten entry?"

By addressing this question, we can build upon the comparisons made between the longitudinal cohort that attended Georgia's Pre-K and the national norming sample for the child assessment measures in the main study to better understand the impact of pre-k attendance. Ideally, this cohort would have been identified in the pre-k year of the study and would have been similar in every way with the pre-k group other than pre-k attendance, and thus both groups of children would have been followed over time. This design would have provided a better understanding of the degree to which attendance in pre-k is associated with better outcomes by eliminating any other source of influence on children's outcomes (and allowing the two groups to be exchangeable). Similar peers were identified later in the study (from children's 3rd grade classrooms whose parents reported that their children did not attend any pre-k program). We do not have demographic data on the comparison group children from their pre-k year, making the assumption of exchangeability difficult (i.e., similarity between the longitudinal pre-k group and the comparison group selected in 3rd grade). Despite these limitations, the comparison group sub-study provides another source for comparison to complement the longitudinal study results and comparisons with a nationally normed sample. Taken together, the results of these two studies contribute to our understanding of the impact of attending Georgia's Pre-K.

The comparison study utilized a nested cohort for the sub-study where the pre-k (exposed) group were those who were enrolled in Georgia's Pre-K and were willing to remain enrolled in the longitudinal study in 3rd grade. The comparison group for this cohort was composed of 333 3rd grade children who were in the same 3rd grade classes as the pre-k cohort children for the 2017-2018 school year, but whose parents reported that the child had no experience with any pre-k during the 2013-2014 school year. For the design of this sub-study, the original cohort was intended to be followed over time through Grade 4 and the comparison group recruited in Grade 3 would also be followed through Grade 4 longitudinally. We conceptualized assessments of children in the fall of Grade 3 as baseline measures for the beginning of this separate sub-study.

RESULTS FROM LONGITUDINAL STUDY

Larger than expected growth was observed during Georgia's Pre-K year in literacy, math, and social skills; gains were supported through kindergarten. Another period of higher than expected growth occurred in 4th grade.

Cumulative results from pre-k through 4th grade are organized in this section in five areas: (1) comparison of within-grade scores with national averages, (2) longitudinal patterns of growth in children's scores over time from pre-k through 4th grade in English, (3) longitudinal patterns of growth in children's scores over time from pre-k through 4th grade in Spanish, (4) factors associated with differential growth in children's scores over time, and (5) examination of classroom quality. The sections that follow include the analytic approach and results.

① Comparison of Within-Grade Scores with National Averages

Analytic Approach to Compare Within-Grade Change with National Averages

To examine change in within-grade scores, fall scores were subtracted from spring scores. Positive change suggests that spring scores were higher than fall scores in the same grade. Negative scores suggest that scores decreased from fall to spring within a given grade. Average scores across child outcome measures were within one standard deviation of the expected mean for children of the same age from the sample used to calculated age-standardized test scores. Within-grade gains in standard effect sizes are shown in Table 4 for measures assessed using standard scores, using HLM model adjusted means. Figure 5A and 5B shows the same differences in standard scores with the addition of executive function scores in their original scale. Adjusted means for the standardized outcome scores for the longitudinal sample are presented in Appendix Table B2.

Results — Comparison of Within-Grade Change with National Averages

The largest gains were observed during the pre-k and 4th grade years. From fall to spring in pre-k, children made small to moderate gains in two domains of literacy and social skills. In 4th grade, small fall-to-spring gains were observed for literacy (passage comprehension) and math (calculations), and large gains in social skills were observed. Small gains were also observed in kindergarten related to auditory processing and decoding skills, along with very small gains in letter-word identification. During 2nd grade, children's literacy and math skill growth was smaller relative to the national average (negative growth), but the rate of decline was in the very small range and resulted in children's scores moving closer to the national average. In 3rd grade, children's growth in their auditory processing and letter-word identification skills slowed from fall to spring, but the children showed a small amount of growth in their social skills relative to peers of the same age.

TABLE 4. Fall-Spring Changes – Effect Sizes

		Pre-K	Kindergarten	1st Grade	2nd Grade	3rd Grade	4th Grade
Language & Literacy	Picture Vocabulary	-0.02	0.04	-0.04	-0.03	0.01	0.09
	Sound Awareness	0.61	0.25	0.00	-0.15	-0.20	-0.10
	Letter-Word ID	0.40	0.12	-0.06	-0.14	-0.11	0.07
	Word Attack	0.07	-0.06	-0.13	-0.14	-0.07	-0.01
	Passage Comprehension	X	X	-0.10	-0.13	0.03	0.32
Math	Calculations	X	X	-0.07	-0.17	0.02	0.30
	Applied Problems	0.14	0.01	-0.06	-0.10	-0.08	-0.01
Social Skills	Social Skills	0.22	0.03	-0.04	-0.01	0.19	0.66
	Behavior Problems	-0.04	-0.01	0.01	0.03	0.10	0.08

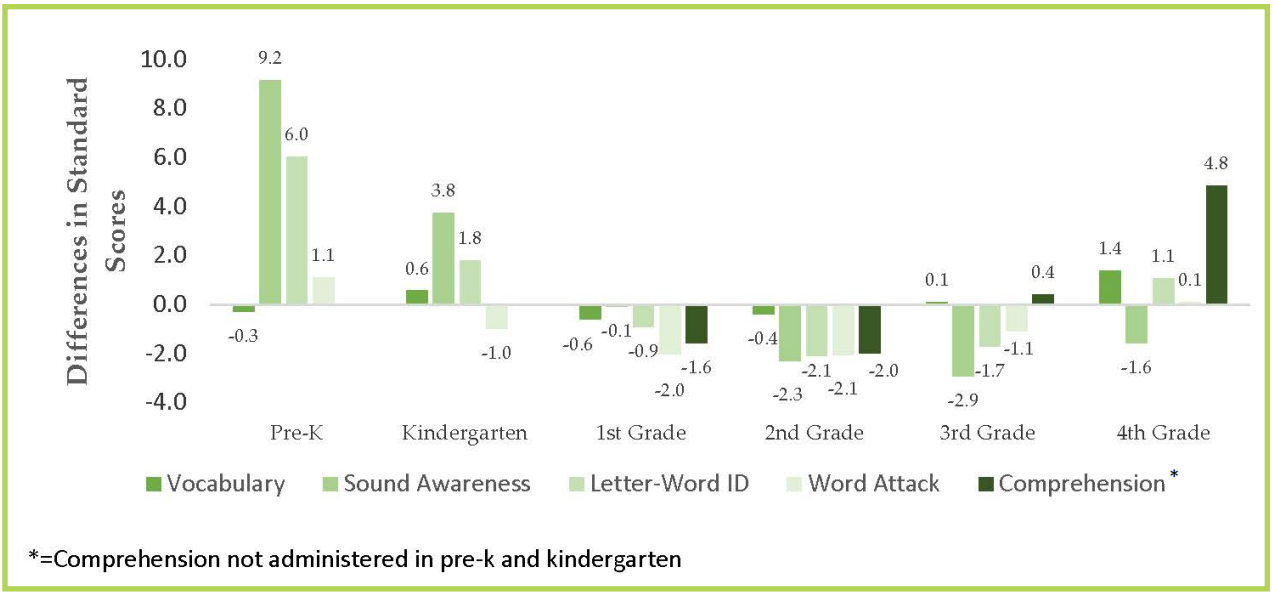
X
Not administered that study year.

Very small gains or losses
from fall to spring within each year (E.S. = 0.10-0.19).

Small gains or losses
from fall to spring within each year (E.S. = 0.20-0.49).

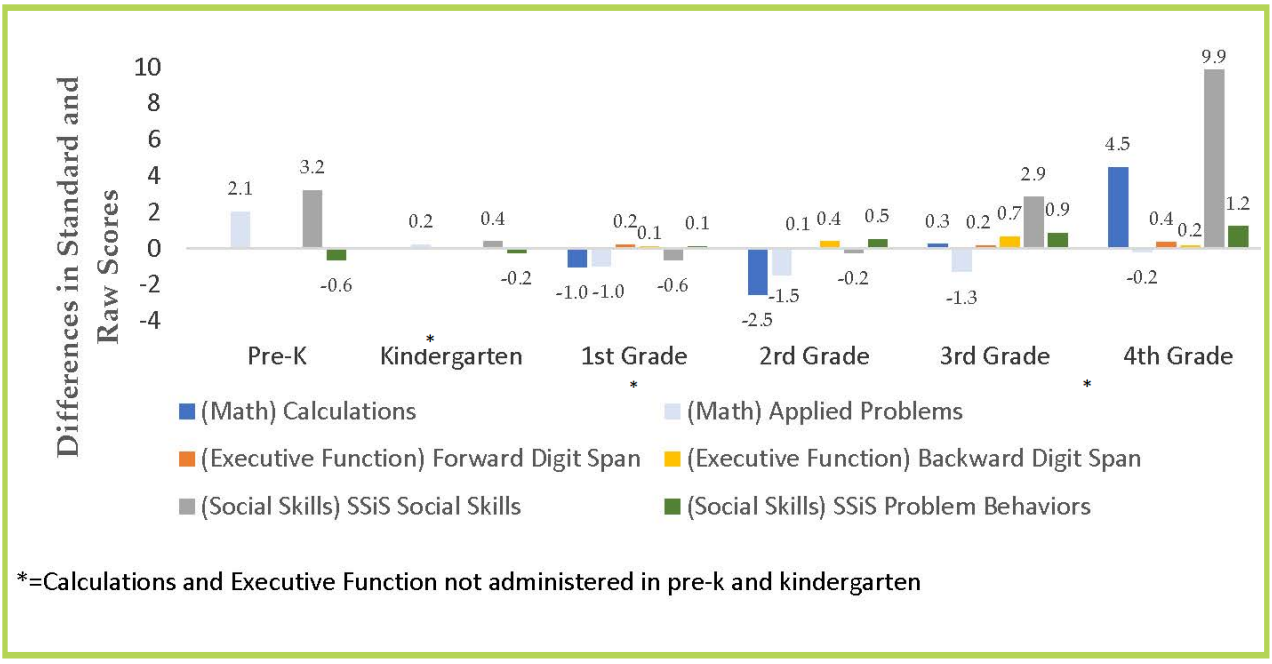
Moderate gains or losses
from fall to spring within each year (E.S. = 0.50-0.79).

FIGURE 5A. Fall-Spring Gains by Grade – Rate of Growth within School Years for Language and Literacy



*=Comprehension not administered in pre-k and kindergarten

FIGURE 5B. Fall-Spring Gains by Grade – Rate of Growth within School Years for Math, Executive Function, and Social Skills



*=Calculations and Executive Function not administered in pre-k and kindergarten

② Longitudinal Patterns of Growth in Children's Scores Over Time from Pre-K through 4th Grade in English

Analytic Approach to Children's Growth Over Time in English

To answer Research Question 1, "What are the longitudinal outcomes through 4th grade for children who attended Georgia's Pre-K Program?" we examined patterns of growth over time in child skills from pre-k through 4th grade for children who attended Georgia's Pre-K program. The related analytic question for Research Question 1 was as follows: "Did the rate of growth in children's outcomes change over time from pre-k through 4th grade for children who attended Georgia's Pre-K Program?" To answer this question, we estimated patterns of longitudinal growth in child outcomes from pre-k through 4th grade of children who attended Georgia's Pre-K using hierarchical linear models (HLM). These models test longitudinal growth of child outcomes over the six-year study duration. Separate sets of analyses were conducted for the full sample and for the DLL subsample for outcomes measured in English and in Spanish. Separate HLM analyses were also run for each outcome, and each included child, family, program, classroom characteristics, and quality as covariates. Specifically, the following covariates were included in all models: 1) Child/family characteristics – gender, race/ethnicity (Hispanic/Latino, White, Black, Multi-racial/Other), IEP status, English/Spanish preLAS language proficiency level at fall of pre-k, and family income; 2) Pre-k program/classroom characteristics – provider type (private setting vs. public school system) and class size; and 3) Classroom quality (low, mid, high). Separate HLM models were fit for each outcome.

These outcomes were assessed from pre-k through 4th grade: WJ-III Picture Vocabulary, WJ-III Sound Awareness, WJ-III Letter-Word Identification, WJ-III Word Attack, WJ-III Applied Problems, SSiS Social Skills, and SSiS Problem Behaviors standard scores. Additional outcomes were also assessed from 1st grade through 4th grade: WJ-III Passage Comprehension and WJ-III Calculation standard scores, and Forward Digit Span and Backward Digit Span adjusted scores.
In accordance with standard practice for model fitting, we tested the best pattern of growth or functional form for time for each model to examine changes in outcomes over time using a polynomial approach (e.g., using linear, quadratic, or cubic forms for time as appropriate). Non-linear HLM analyses followed a polynomial approach. The polynomial approach is best in illuminating the fluctuations in the rate of change of a given skill over time and is representing the most "natural" flow of skill growth without stringent functional restrictions (such as fitting the best line, or a U-shaped curve). Polynomial functional forms could vary by rates of change over time to be linear (Time), quadratic (Time2), or cubic (Time3) in shape depending on model fit for each outcome. The downside of this approach is that the interpretation of significant results is possible most of the time only in conjunction with an accompanying graph. The benefit is that the most accurate functional representation of the data can be presented. By graphing this functional representation, the reader can get a clear illustration of the fluctuations of skill growth over the entire time period from fall of pre-k through spring of 4th grade. Without allowing for fluctuations with squared or cubic forms of time, a straight line would be fit to represent all estimates over time. For this report, figures were created for the whole sample and for subgroups in cases of moderation testing.

All estimates and gains were calculated with covariates that were grand mean centered, and were adjusted for multiple comparisons using Benjamini-Hochberg[iv] adjustments to correct for the potential false discovery rate. Null hypothesis significance testing was done only for growth over time for each outcome model, so p-values were only included for specific effects that were tested and corrected.

Standard Scores: Standard scores were utilized to compare children's expected scores accounting for typical growth for age over time. Without intervention, it is expected that children's scores would remain fairly stable over time, a change of zero. A change greater than zero, or growth, indicates that children gained additional skills in that skill area relative to the amount of growth of other children of the same age in that skill area. Negative change over time indicates that children did not show as much growth in the skill area as same-age peers, often having average scores that trend downward toward the average score for children of the same age. Standard scores are available for all assessments except executive function measured by the Forward and Backward Digit Span tasks. Standard scores have a national norm mean of 100 and a standard deviation of 15.

Results — Analyses of Children's Growth Over Time

Child skills were examined relative to their peers primarily using standard scores on the Woodcock-Johnson III measure. If scores increased significantly over time, it suggested that children made gains in these skills relative to peers of the same age and that skills were growing at a rate greater than expected based on maturation alone. If significant growth was not observed, it meant that growth was taking place as expected based on the child's previous skills and age. Finally, decreasing scores indicate that children were continuing to learn and gain age-appropriate skills, but not at the expected rate relative to their same-age peers in the norming sample for the measure. HLM Results are shown in Appendix Table C1.

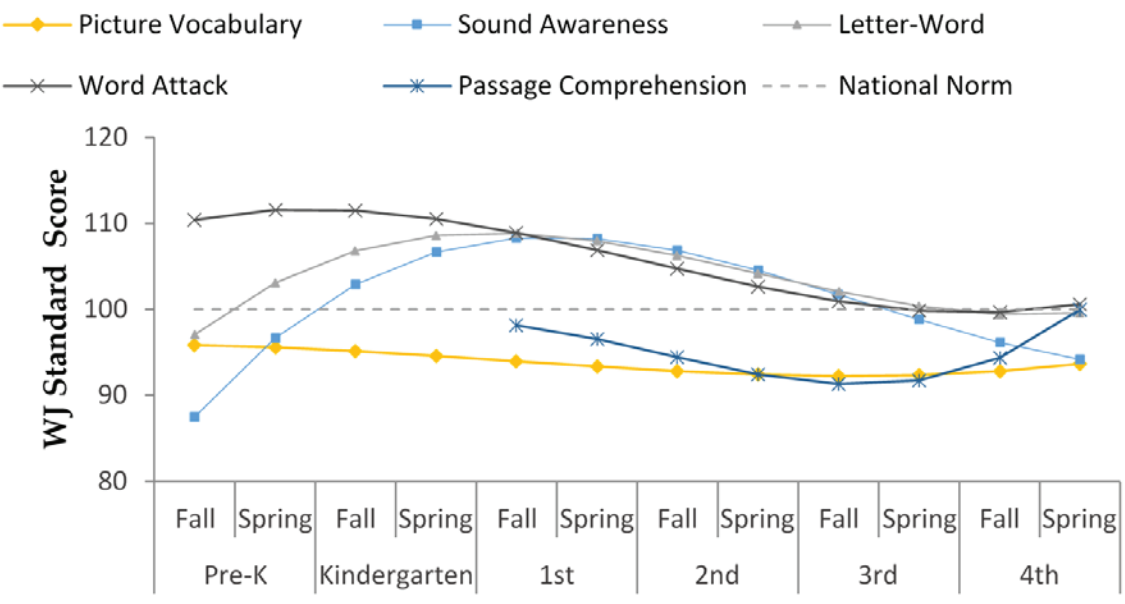
Results — Examining Patterns of Growth in Children's Scores Over Time by Outcome

Language and Literacy

For most language and literacy skills, children showed a pattern of negative gains (lower than expected growth) in 1st through 4th grade, with relatively steeper declines in scores in 2nd grade than in 1st grade (see Figure 6). For phonological awareness skills (WJ-III Sound Awareness) and letter word recognition (WJ-III Letter Word), children evidenced gains above peers each year from pre-k to 1st grade, then declining to approach the norming sample in 3rd and 4th grade (lower than average scores in 4th grade for phonological awareness). In contrast, for one measure of decoding skills (WJ-III Word Attack), scores peaked in pre-k at a mean that was higher than the norming sample with gradual decline toward the norming sample each year through 3rd grade. Picture vocabulary scores began slightly below the norming sample and demonstrated a very gradual but consistent decrease over time. For the literacy measure gathered in 1st through 4th grades only (WJ-III Passage Comprehension), the pattern showed similarly smaller decreases in scores in 1st grade through 3rd grade, but a steep increase in skills in 4th grade. Calculated estimates for growth in language and literacy outcomes over time are listed in Appendix Table C1.

FIGURE 6. Growth in Language and Literacy Skills (Pre-K–4th Grade)

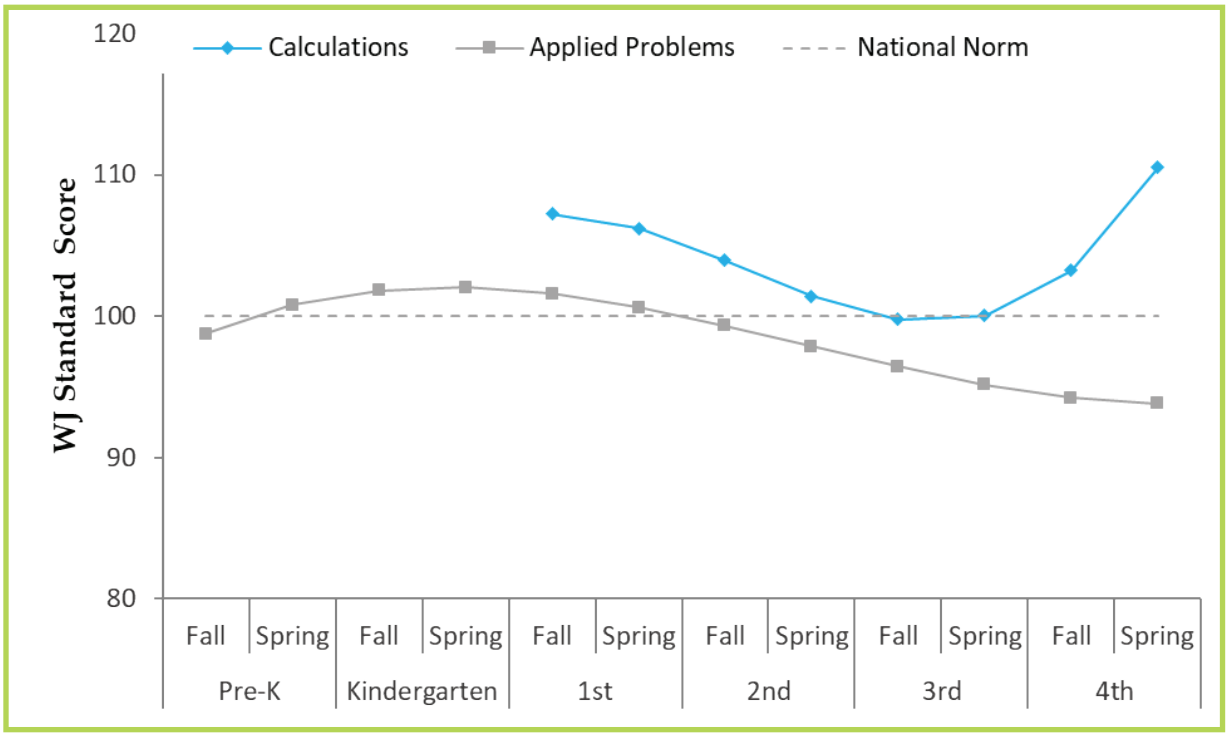
Gains in **literacy skills** were observed in pre-k through kindergarten, with declines or leveling off in skills relative to children of the same age in 1st through 4th grades. **Reading comprehension** scores increased in 4th grade. **Language** skills started and stayed below the national average through 4th grade.



Math Skills

Growth on math skills first assessed in pre-k declined in 1st and 2nd grades with high points in pre-k and kindergarten and a decline relative to national norms through 4th grade. This assessment focused on quantitative reasoning, math achievement, and math knowledge (assessed using the WJ-III Applied Problems subtest; see Figure 7). The ability to perform mathematical computations was assessed starting in 1st grade, and children performed higher than the national norm (using the WJ-III Calculation subtest) in 1st and 2nd grades, decreased to the sampling norm average in the 3rd grade, then rose significantly above the norm by end of 4th grade. Calculated estimates for growth in math outcomes are listed in Appendix Table C2.

FIGURE 7. Growth in Math Skills (Pre-K–4th Grade)

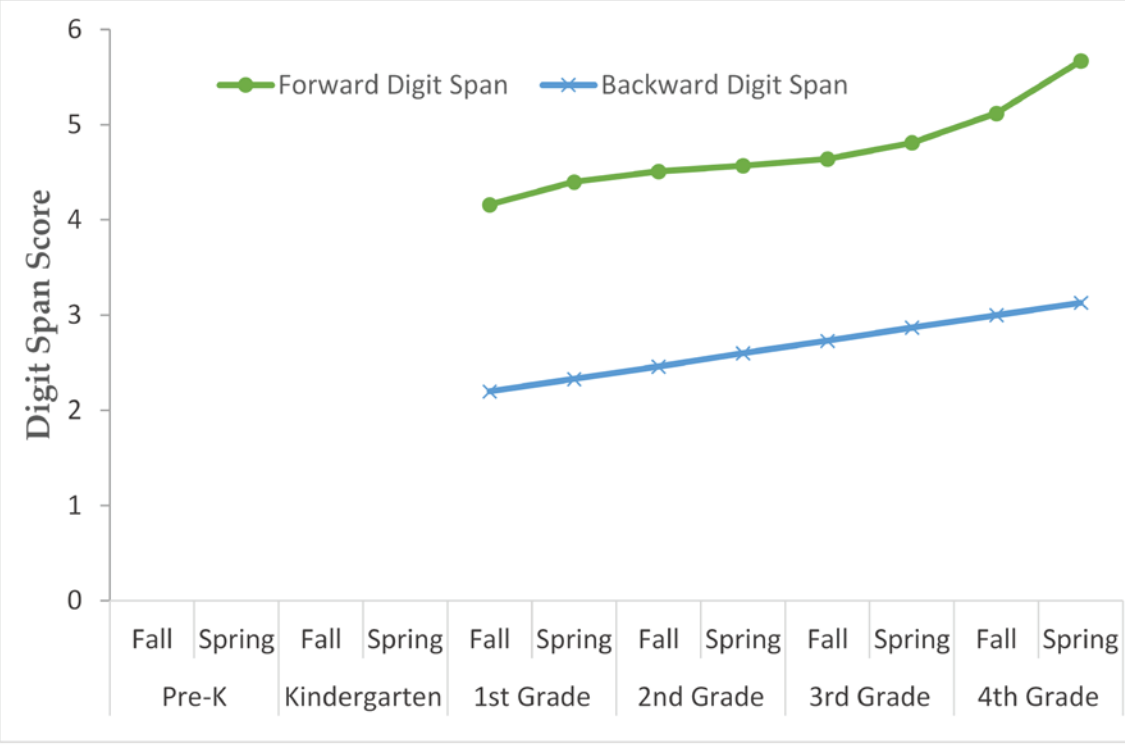


Executive Function

For both measures of executive function skills, specifically working memory (assessed using the Forward and Backward Digit Span tasks; see Figure 8), children showed a pattern of positive growth, with linear and cubic gains in 1st through 4th grades. These scores reflect increasing skill levels over time in both tasks, but these are not norm-referenced scores, so comparisons with expected scores at a given age cannot be made. The Backward Digit Span is the more taxing of the two tasks on working memory, and average scores on this task tend to be lower than Forward Digit Span scores. Calculated estimates for growth in executive function outcomes are listed in Appendix Table C2.

FIGURE 8. Growth in Executive Function Skills (1st–4th Grade)*

Gains in **executive function scores** were observed in 1st through 4th grades, with a slight increase in working memory gains in 4th grade.

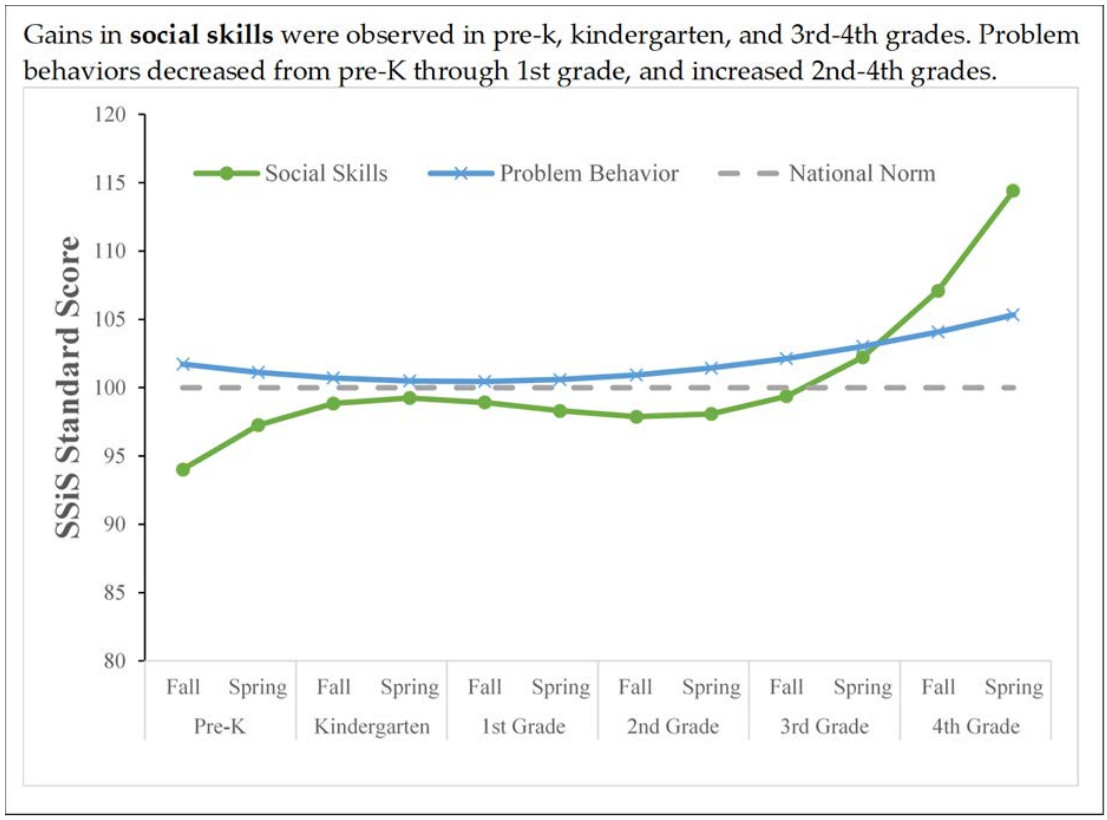


*Digit Span measures produce unstandardized scores; national averages are not available.

Social Skills

Both teacher-rated social skills and problem behaviors increased as children progressed from pre-k through 4th grade. Social skills improved through the spring of kindergarten, decreased slightly in 1st and 2nd grades, and then improved above the national norm in 3rd and 4th grades (see Figure 9). Overall scores were in the average range. Problem behaviors stayed relatively stable over time with increases above the national norm in the spring of 2nd grade through 4th grade. There were slightly greater decreases in problem behavior ratings (improvements in behavior) in pre-k and slightly greater increases in ratings (declines in behavior) from the end of 2nd through 4th grades. Calculated estimates for growth in social skills and problem behavior outcomes are listed in Appendix Table C2.

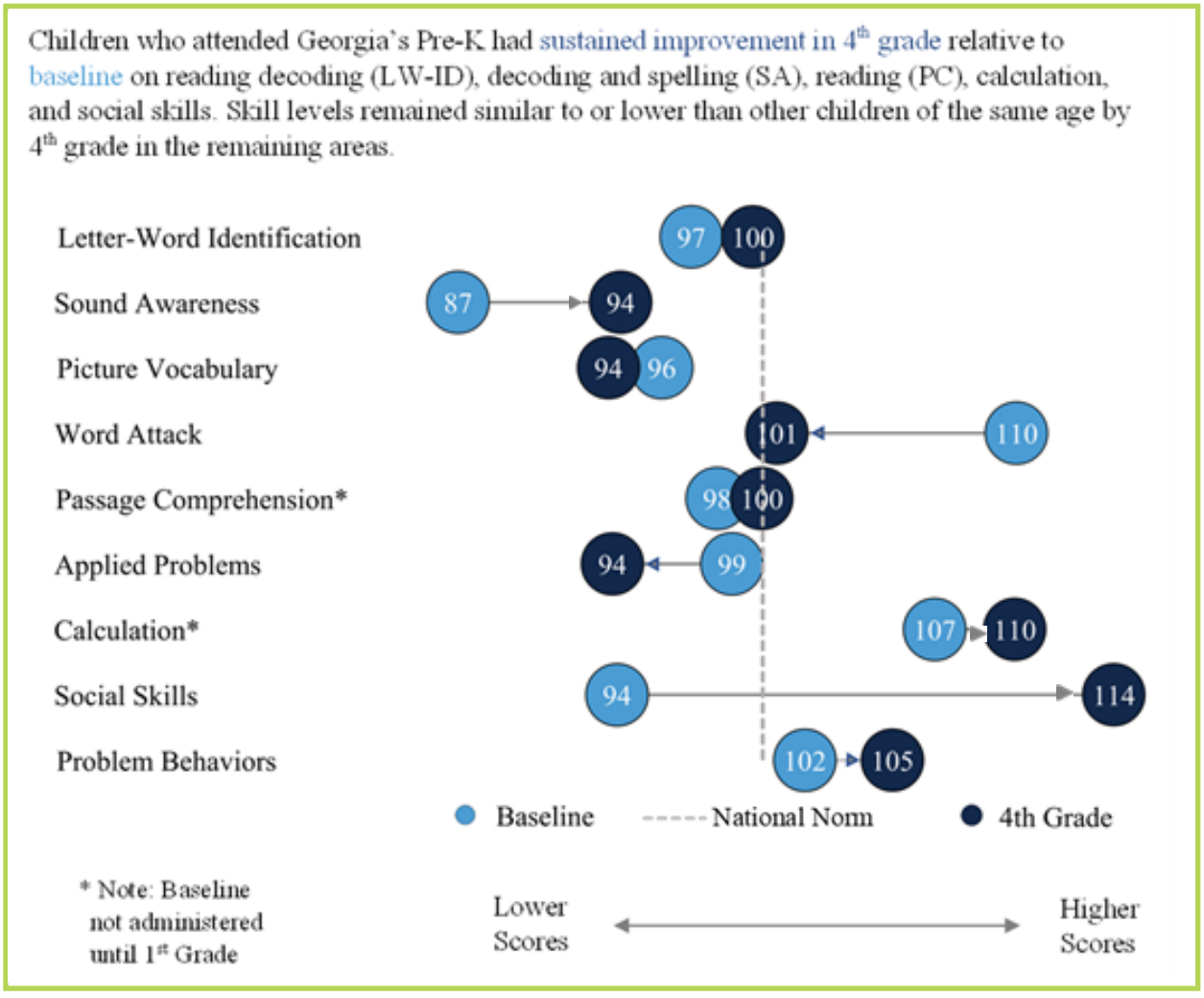
FIGURE 9. Growth in Social Skills (Pre-k–4th Grade)



Comparisons of Baseline to Final Scores

Comparisons of baseline pre-k scores (or 1st grade) for assessments are found in Figure 10.

FIGURE 10. Total Adjusted Gains Over Time between Baseline (Pre-K and 1st Grade) and 4th Grade Adjusted Scores



③ Longitudinal Patterns of Growth In Children's Scores Over Time From Pre-K through 4th Grade in Spanish

Analysis Approach to Children's Growth Over Time in Spanish

A parallel set of analyses was conducted specifically for dual language learners (DLL). Like the full sample analyses, patterns of growth were examined over time in child skills from pre-k through 4th grade for children who attended Georgia's Pre-K Program.

Results — Children's Growth Over Time in Spanish

Like the results for the full sample, children showed larger gains during pre-k and somewhat lesser gains during kindergarten, with smaller-than-expected gains during 1st grade and leveling out or greater decreases during 2nd and 3rd grades, and an increase in some skills in 4th grade. However, compared to the full sample, the average scores were somewhat lower in English (see Figures 11, 13, and 15; Appendix Table B3) and much lower in Spanish for DLL (see Figures 12 and 14; Appendix Table B4).

Average standard and adjusted scores for dual language learners (DLL) for the full longitudinal sample are presented in Appendix Table B3 for outcomes assessed in English and Appendix Table B4 for outcomes assessed in Spanish. Average standardized scores across child outcome measures were in the average and expected range relative to children of the same age in the norming sample. English language skills (WJ-III Picture Vocabulary) started out and persisted in the lower than average range despite gains in the literacy domain that yielded scores in the average range. Similar to English language skills, vocabulary skills in Spanish were in the low (70-79) or very low (60-69) range, which persisted through 3rd grade. Literacy and math scores were mostly in the low average range (80-89).

Language and Literacy Skills: Dual Language Learners

Similar to patterns observed with the full sample, patterns of growth in language and literacy skills for DLL were observed in pre-k and kindergarten with lower than expected growth in 1st and 2nd grades with some growth in 3rd and 4th grades (see Figures 11 and 12). DLL language skills in English (WJ Picture Vocabulary) began well below the norming sample and demon-strated increases in pre-k-kindergarten and 3rd-4th grades. For the literacy measure gathered in 1st through 4th grades only (WJ-III Passage Comprehension), the pattern showed similarly smaller decreases in scores in 1st through 3rd grades and increases in skills in 4th grade. Phonological awareness skills (WJ-III Sound Awareness) grew relative to children of the same age through 2nd grade, with steep declines through 4th grade. Spanish skills generally started and stayed lower than English language skills with similar patterns of decline and growth over time. Calculated estimates for growth in language and literacy outcomes are listed in Appendix Tables B3 and B4.

FIGURE 11. Rate of DLL Growth in English Language and Literacy Skills (Pre-K–4th Grade)

Gains in **English literacy skills** were observed in pre-k through kindergarten, with declines or leveling off in skills relative to children of the same age in 1st through 4th grades. **Reading comprehension** scores increased in 4th grade. **Language** skills started and stayed below the national average through 4th grade.

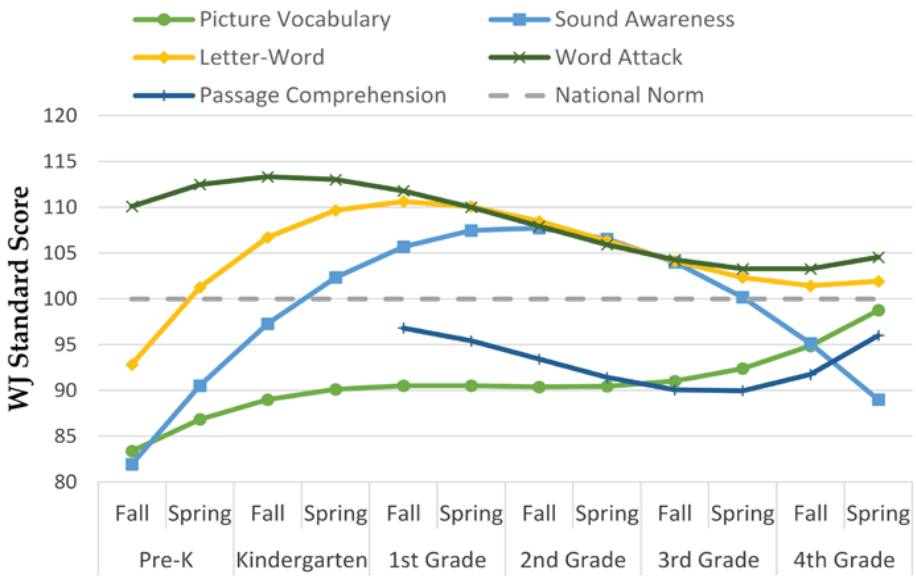
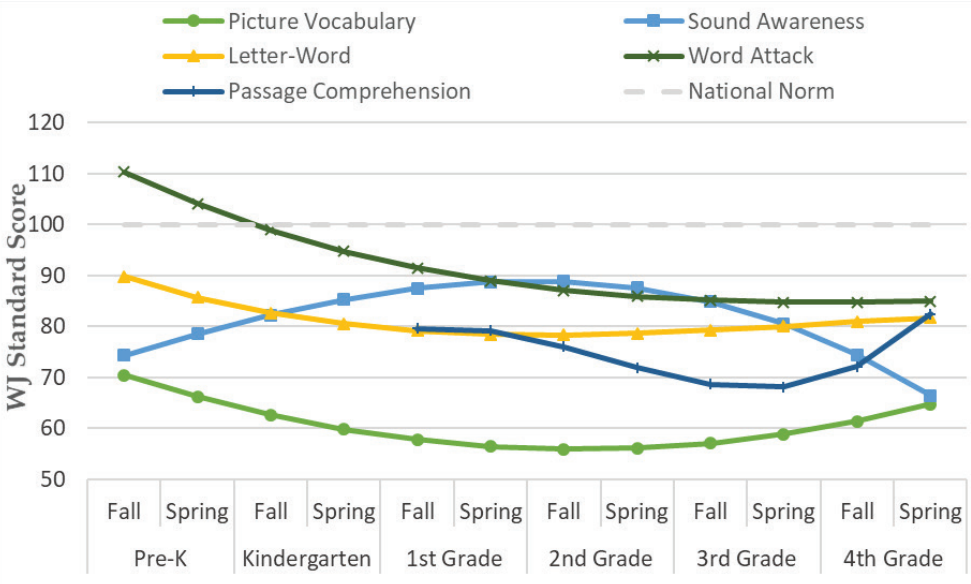


FIGURE 12. Rate of DLL Growth in Spanish Language and Literacy Skills (Pre-K–4th Grade)

All but one measure of **Spanish language and literacy skills** started and stayed below the national average through 4th grade. Gains in **Spanish phonological awareness skills** were observed in pre-K through 2nd grade, with declines or leveling off in skills relative to children of the same age in 2nd through 4th grades. **Reading comprehension** scores increased in 4th grade and **vocabulary** scores increased in the 3rd and 4th grades.



Math Skills: Dual Language Learners

Growth on math skills for DLL first assessed in pre-k showed greater than expected growth for age in pre-k and kindergarten. Scores on both math assessments declined relative to national norms in 1st-2nd grade and increased in 3rd-4th grade (see Figures 13 and 14). The WJ-III Applied Problems assessment focused on quantitative reasoning, math achievement, and math knowledge. The ability to perform mathematical computations was assessed starting in 1st grade, and children mostly performed higher than the national norm (using the WJ-III Calculation subtest). Calculated estimates for growth in math outcomes are listed in Appendix Table C4 and C5.

FIGURE 13. DLL Growth in English Math Skills (Pre-K–4th Grade)

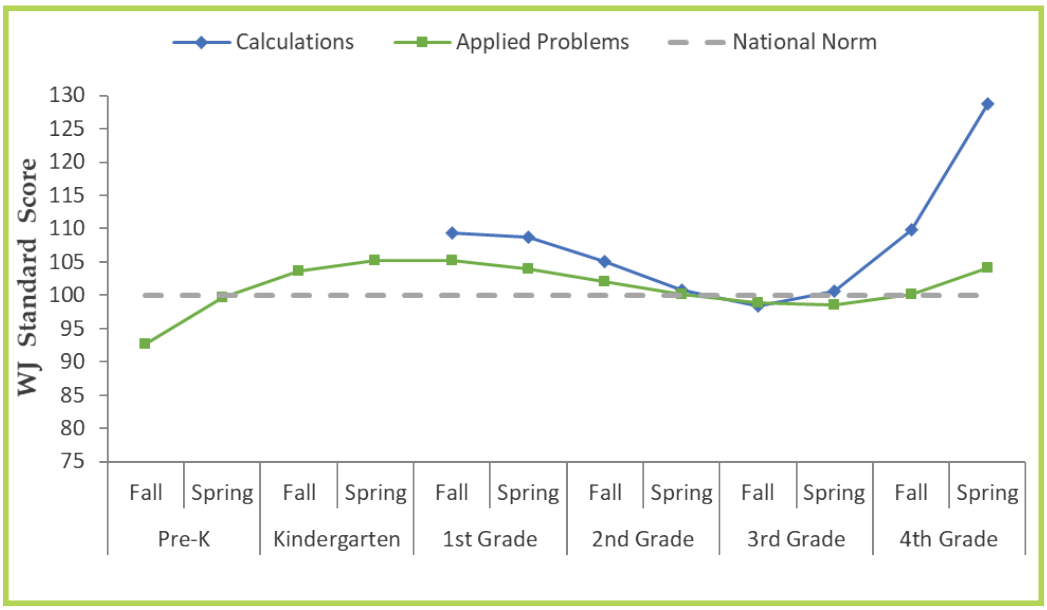
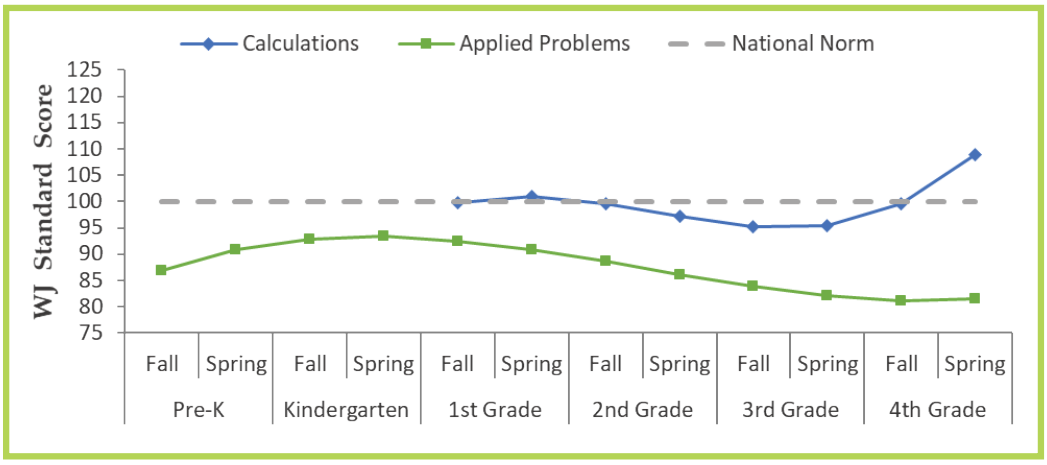


FIGURE 14. DLL Growth in Spanish Math Skills (Pre-K–4th Grade)

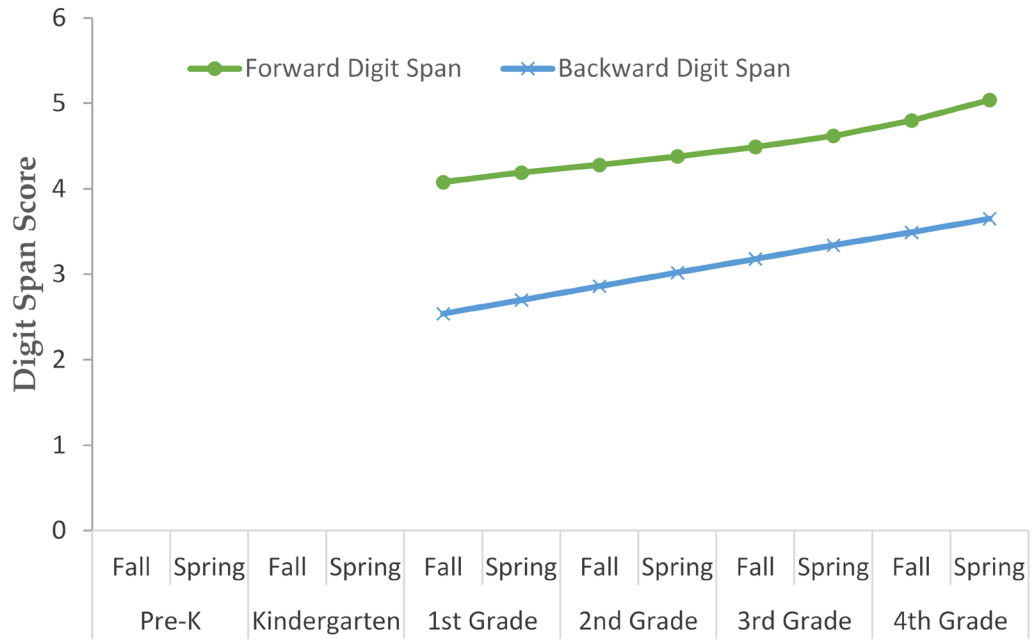


Executive Function Skills: Dual Language Learners

For both measures of executive function skills used with DLLs, children showed a pattern of positive growth in 1st through 4th grades (see Figure 15). Overall scores were lower than those observed with the full sample. These scores reflect increasing skill levels over time in both tasks, but these are not norm-referenced scores, so comparisons with expected scores at a given age cannot be made. The Backward Digit Span is the more taxing of the two tasks on working memory, and average scores on this task tend to be lower than Forward Digit Span scores. Calculated estimates for growth in executive function outcomes are listed in Appendix Table C5.

FIGURE 15. DLL Growth in Executive Function Skills (1st–4th Grade)*

Gains for DLLs in **executive function scores** were observed in 1st through 4th grades, with a slight increase in working memory gains in 4th grade. Scores were slightly lower overall than the full sample averages.



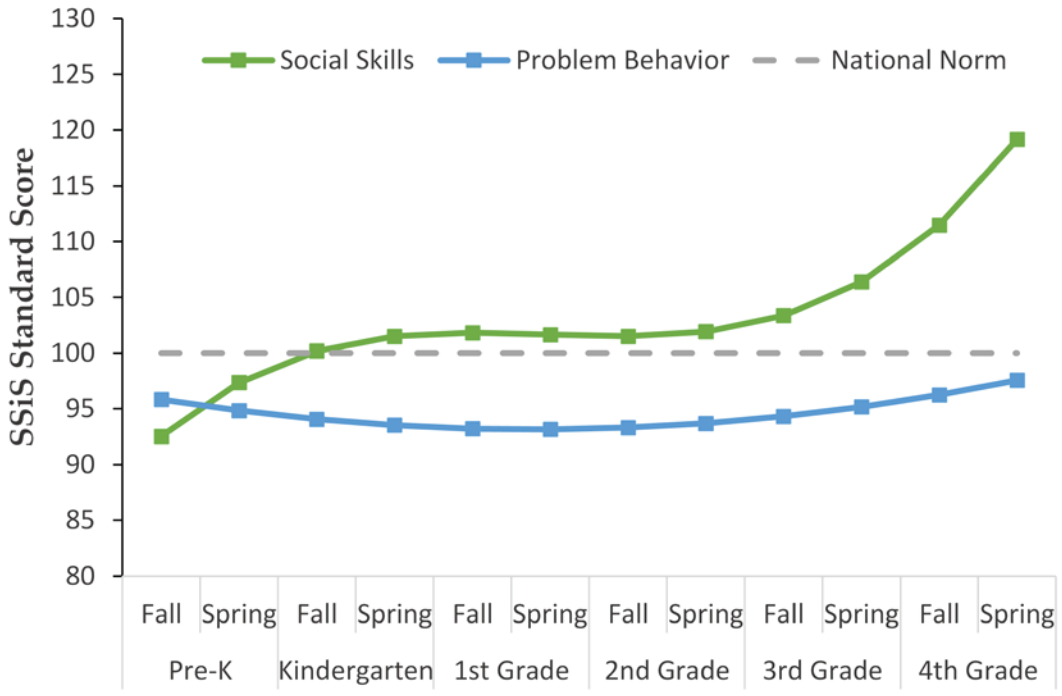
*Digit Span measures produce unstandardized scores; national averages are not available.

Social Skills: Dual Language Learners

Both teacher-rated social skills and problem behaviors increased for DLL as children progressed from pre-k through 4th grade. Social skills improved through the spring of kindergarten, decreased slightly in 1st and 2nd grades, and then improved above the national norm in 3rd and 4th grades (see Figure 16). Overall scores were in the average range. Problem behaviors stayed relatively stable over time with increases above the national norm in the spring of 2nd grade through 4th grade. There were slightly greater decreases in problem behavior ratings (improvements in behavior) in pre-k and slightly greater increases in ratings (declines in behavior) from the end of 2nd through 4th grades. Problem behaviors were rated much lower for DLL than for the full sample.DLL were also rated as having lower social skills in pre-k and kindergarten. Social skills were assessed by teachers, so separate assessments in English and Spanish are not applicable. Calculated estimates for growth in social skills and problem behavior outcomes are listed in Appendix Table C4.

FIGURE 16. DLL Growth in Social Skills (Pre-K–4th Grade)

Gains in **social skills** for DLLs were reported by their teachers in pre-k, kindergarten, and 3rd-4th grades. Problem behaviors decreased slightly from pre-k through 1st grade, and increased slightly from 2nd-4th grades. Social skills were rated much lower than the full sample before 1st grade, and overall problem behaviors were well below the national average for DLLs.



④ Characteristics Associated with Differential Patterns of Growth in Children's Scores Over Time – Full Sample

Analysis Approach to Examining Predictors of Growth in Children's Scores Over Time

To answer Research Question 2: "What factors are associated with better longitudinal outcomes for children?" HLM analyses were conducted with additional interaction variables for time and the covariate of interest. These additional HLM analyses were conducted as described in the section for Research Question 1 with additional interaction variables for time and the covariate of interest to test for effect heterogeneity via moderation (i.e., does growth for children in one group differ over time compared to growth in another group). These moderation models examined whether patterns of growth were similar for different groups of children by comparing growth trajectories associated with different child/family characteristics (i.e., child gender, race/ethnicity, IEP status, oral language proficiency, and family income) and pre-k classroom characteristics (i.e., provider type and class size).

Results — Examining Predictors of Growth in Children's Scores Over Time

For children who attended Georgia's Pre-K Program, average scores were within the expected and average range (standard score of 90-110), with some skills in the "high average" range (standard score of 111-120). Patterns of longitudinal growth were significant over time for most language, literacy, math, and social skills measured from pre-k through 3rd grade assessed in English (see Figures 6, 7, 8, 9, and Appendix Table B2). On most of these norm-referenced measures, children showed a pattern of initial gains in scores during pre-k and kindergarten (i.e., larger gains than expected relative to the norming sample), with scores starting to level off in 1st grade and then growth decreasing, or scores stabilizing, through 4th grade.

These moderation models examined whether children's growth in academic and behavior skills are similar for children with different English proficiency levels and racial and ethnic backgrounds. Adjusted mean scores over time estimated from the HLM models where moderation results were statistically significant are provided for the full longitudinal sample in Figures 17, 18, 19, 20 and 21 and the HLM coefficients are provided in Appendix Tables C1, C2, C3, C4, and C5. To adjust for multiple comparisons that increase the likelihood of falsely identifying a statistically significant result, adjustments were applied to p-values. The following child characteristics predicted differences in growth on some longitudinal outcomes for children:

- Children's racial and ethnic background were not associated with statistically different patterns of growth.
- Different patterns of growth over time were observed for some outcomes including child English proficiency level at pre-K, IEP status at the beginning of pre-K, and the gender of the of child.

The figures listed in the next section are for moderation differences that were statistically significant only.

Sensitivity Analysis for Child Race/Ethnicity: Previous analyses of outcomes by children's race and ethnicity modeled race/ethnicity as a combined variable; in these analyses, Hispanic ethnicity is modeled as a separate variable from racial categories. A sensitivity check was conducted with the race and ethnicity variables using both approaches, and there were no difference in results (adjusted means available by request).

Moderation by Child English Proficiency Level at Pre-K

Children's English proficiency level at pre-k entry was associated with differences in longitudinal growth in receptive vocabulary skills (Picture Vocabulary subtest; see Figure 17), quantitative reasoning (Applied Problems subtest; see Figure 18), and social skills (Social Skills Rating; see Figure 19). There were five levels of English proficiency, with level 1 being the least proficient and level 5 being proficient. There were too few children at levels 2 and 3 to analyze separately, so these two levels were combined.

FIGURE 17. Growth in Picture Vocabulary Skills Over Time by Pre-K English Language Proficiency Level

Children who started out with lower levels of English language proficiency in pre-k (**level 1** and **levels 2/3**) had the same or higher scores by 4th grade than at pre-k, whereas pre-k attendees who entered with higher language skills ended up with lower scores over time. Children who entered with very low language skills finished with skills well below children of the same age.

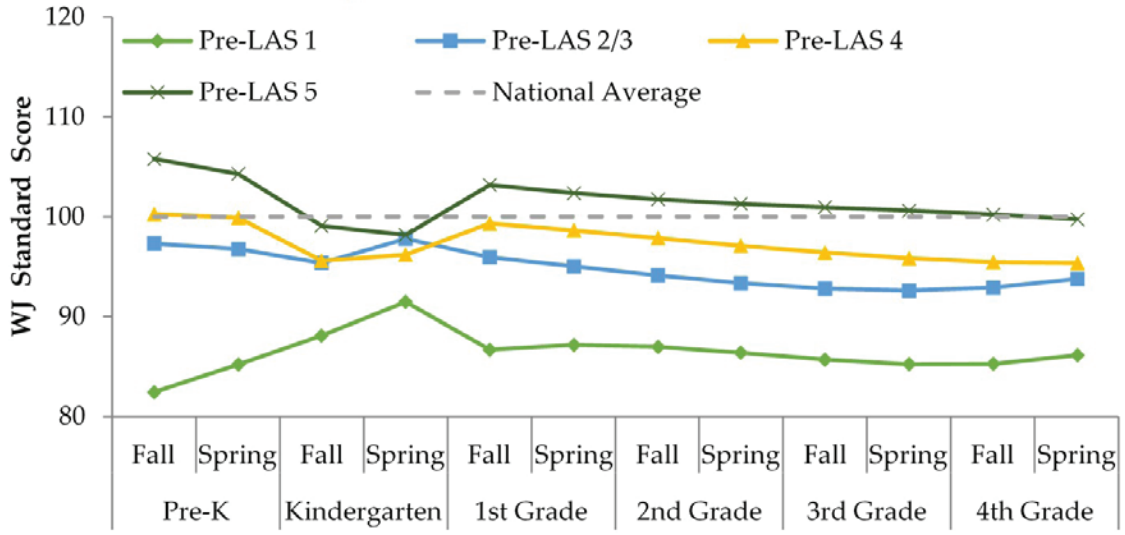


FIGURE 18. Growth in Applied Problems Skills Over Time by Pre-K English Language Proficiency Level

Children who started out with lower levels of English language proficiency in pre-k (**level 1**) showed patterns of **increased growth in quantitative reasoning skills** over time, but still ended up with scores below the national average. Children in the **remaining four categories ended up above the national mean by 4th grade.**

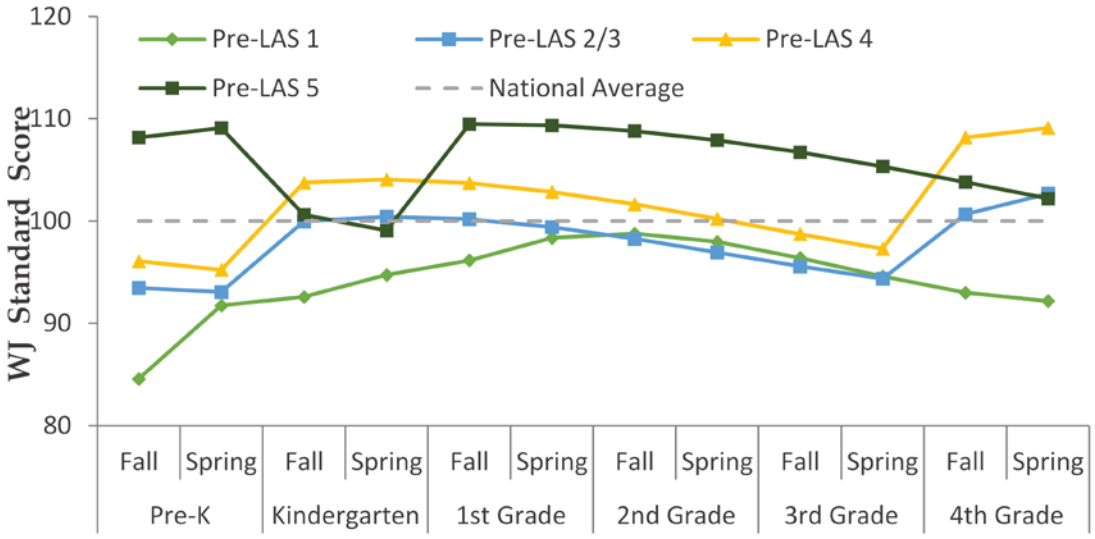
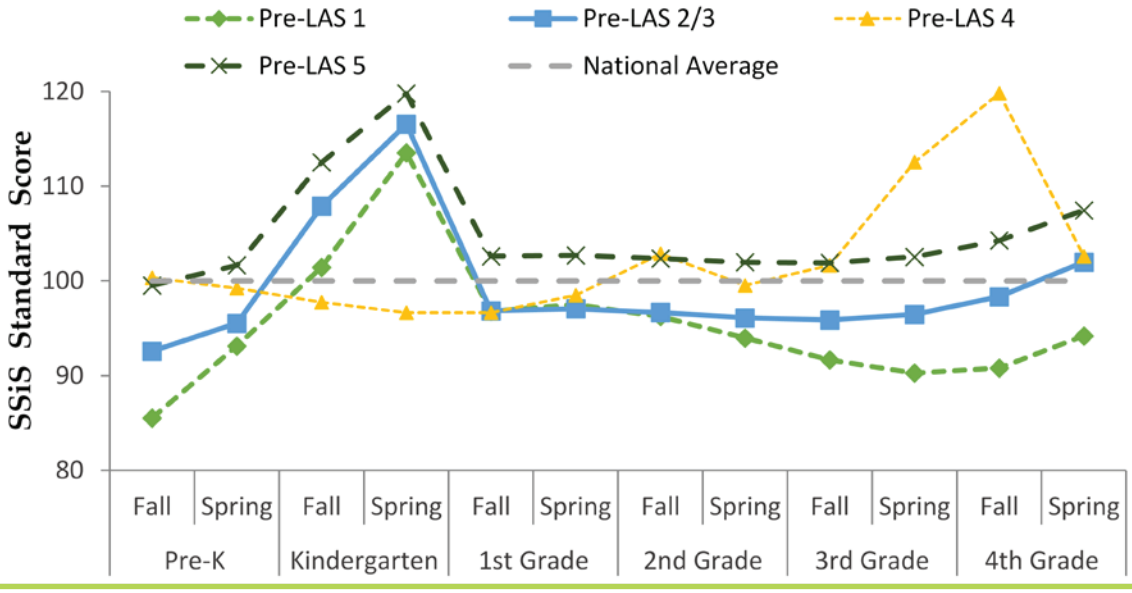


FIGURE 19. Growth in Social Skills Over Time by Pre-K English Language Proficiency Level

Children who had **mid-to-high levels of English language proficiency ended up with social skills above the national norm,** and children with the lowest level of proficiency ended up lower than the national average.

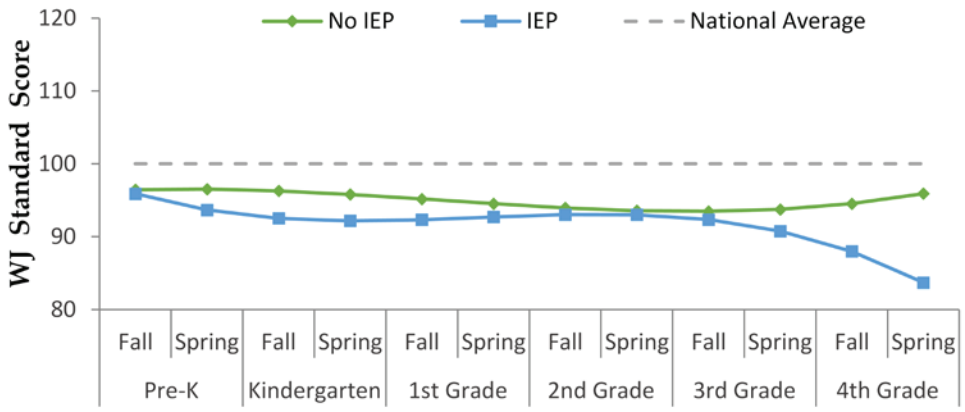


Differences in Outcomes for Children with Different IEP Statuses at Pre-K

Vocabulary skills did not improve as rapidly for children with IEPs at the beginning of pre-k (see Figure 20). Children with an IEP at the beginning of pre-k did catch up in 2nd grade but fell below peers again by the spring of 4th grade.

FIGURE 20. Growth in Vocabulary Skills Over Time by IEP Status at Pre-K

Children with IEPs in pre-k showed slower patterns of growth in vocabulary skills than children without an IEP in pre-k through 1st grade, with some catch-up in 2nd grade and ending well below the national average by the spring of 4th grade.

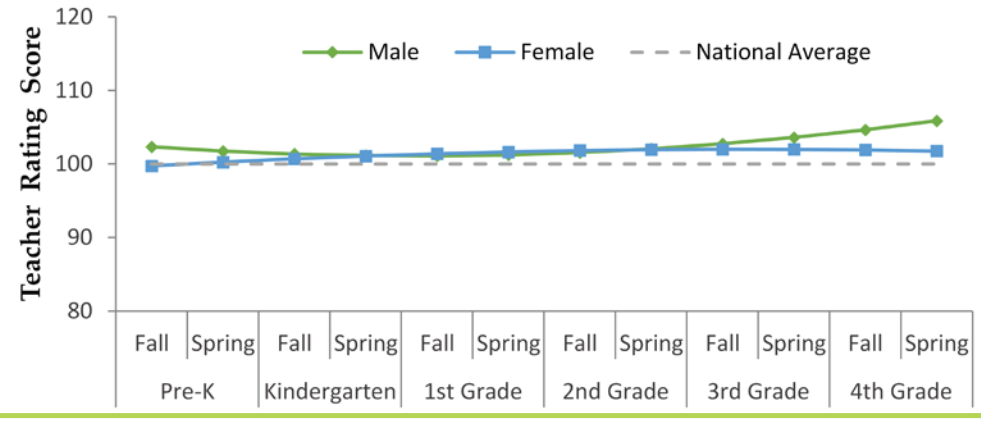


Differences in Outcomes for Children by Biological Sex (Gender)

The child's gender (assigned sex at birth) was also significantly associated with patterns of change in problem behaviors over time. Girls had lower levels of problem behaviors in pre-k and 4th grade than boys with similar levels of problem behaviors in the spring of kindergarten through 2nd grade (see Figure 21).

FIGURE 21. Growth in Problem Behaviors Over Time by Biological Sex of Child

Girls had lower levels of problem behaviors in pre-k, 3rd, and 4th grades than boys. Levels were similar in the spring of kindergarten through 2nd grade.



⑤ Pre-K Program/Classroom Characteristics

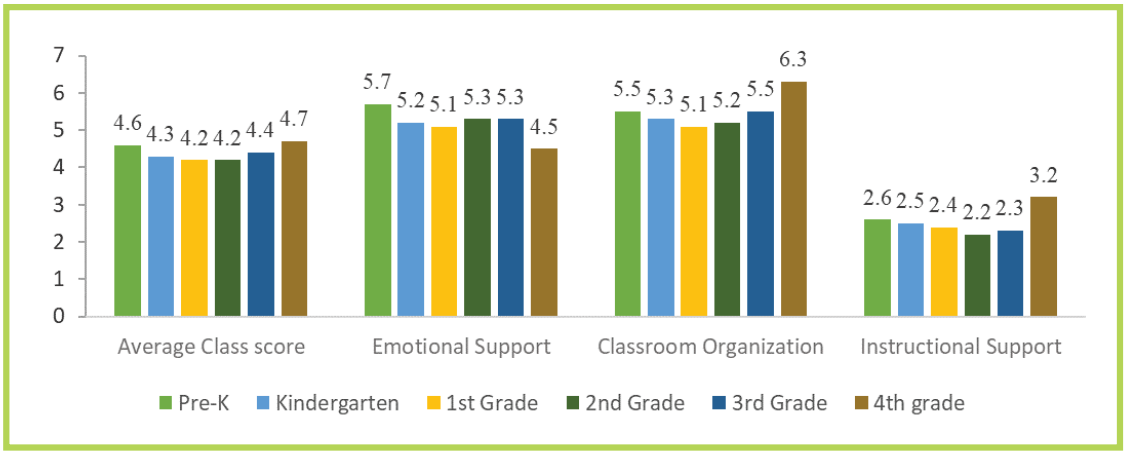
Classroom Quality

To answer the third Research Question 3, "What is the quality of children's instructional experiences from pre-k through 4th grade?" both descriptive and inferential analyses were conducted to estimate the quality of children's classroom experiences over time and associations between the quality of classrooms and patterns of child outcomes. These analyses were conducted for the subsample of children whose classroom quality was observed pre-k through 4th grade. Classroom quality was measured over time using the appropriate grade-level total score from the Classroom Assessment Scoring System: pre-k classroom quality (Pre-K CLASS), kindergarten through 3rd grade classroom quality (K-3 CLASS), and for 4th grade classroom quality (Upper Elementary CLASS). CLASS Domain scores were considered high quality according to the research on quality thresholds (Emotional Support score of 5 or higher, Classroom Organization score of 5 or higher, and Instructional Support score of 2.5 or higher) and were compared with classrooms in the low-to-mid-quality range (scores below the thresholds listed previously). The quality coefficient can be interpreted as the average effect of coming in and out of high-quality classrooms over time on child outcomes.

The quality of instructional practices was examined by the study team each year from pre-k through 4th grade for the observed samples of classrooms attended by children in the study. We conducted descriptive analyses for CLASS Pre-K, CLASS K-3 in kindergarten, 1st, 2nd, and 3rd grade; and CLASS Upper Elementary scores in 4th grade, including means and frequency distributions on the total, domain, and dimension scores.

Using the scoring for the Pre-K, K-3, and Upper Elementary versions of the CLASS recommended by the publishers, average classroom quality total scores were in the mid-range across all years (3-5.9). See Figure 22 and Appendix Table D1. Classroom quality was higher in the pre-k year, slightly lower and consistent through 2nd grade, then rose through 3rd grade to its highest level in 4th grade. The Emotional Support domain scores in 4th grade were in the moderate range of quality, below those in the high range from pre-k through 3rd grade. Classroom Organization scores were slightly lower in 1st grade and 2nd grade and increased to the same levels as pre-k by 3rd grade. In 4th grade, the majority of classrooms (96%) were in the high-quality range for Classroom Organization. Finally, Instructional Support scores were in the low end of the range (1-2.9) from pre-k to 3rd grade, but significantly increased to the moderate range (3-5.9) in 4th grade.

FIGURE 22. Average CLASS Total Scores by Grade



A higher proportion of programs were in the high-quality range in the 4th grade compared to pre-k through 3rd grade scores for the overall CLASS score, or average total score (29%; see Figure 23 and Appendix Table D2). Compared to the pre-k through 3rd grade total scores, the Emotional Support domain score in 4th grade was proportionally higher in the mid-range (66%; see Figure 24). A higher proportion of classrooms were also in the high-quality range for the Classroom Organization domain (96%; see Figure 25), and a higher proportion of classrooms were in the mid-quality range for the Instructional Support domain in 4th grade (61%) compared with pre-k to 3rd grade scores (see Figure 26).

FIGURE 23. Frequency of CLASS Average Total Scores by Grade

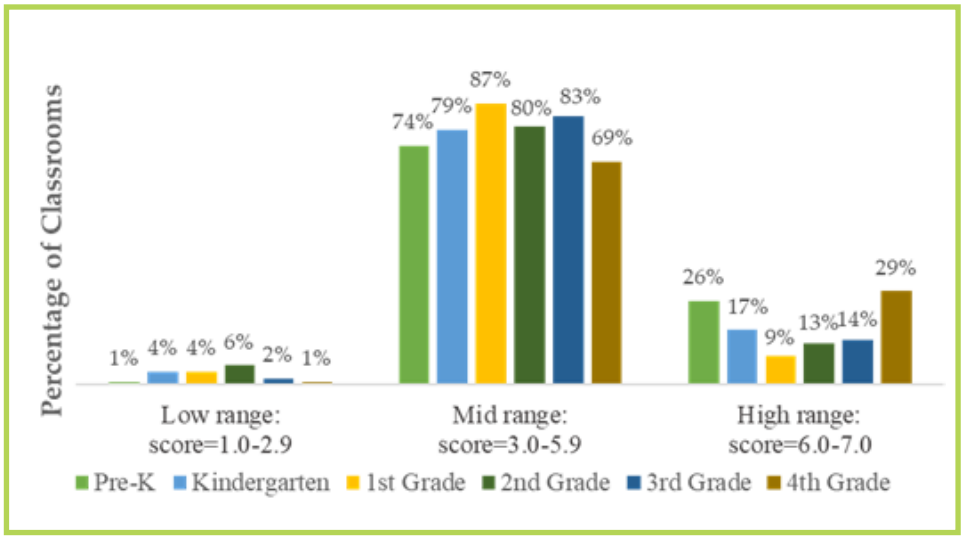


FIGURE 24. Frequency of CLASS Average Total Scores by Domain: Emotional Support

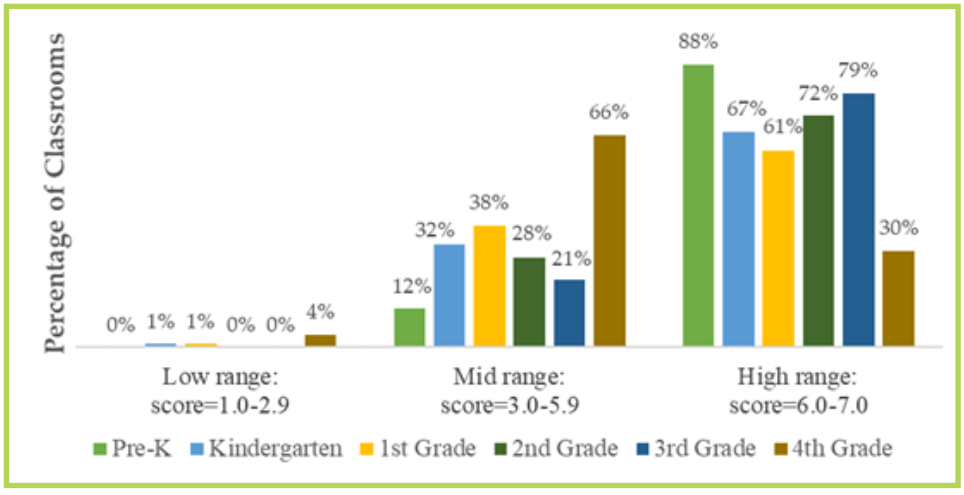


FIGURE 25. Frequency of CLASS Average Total Scores by Domain: Classroom Organization

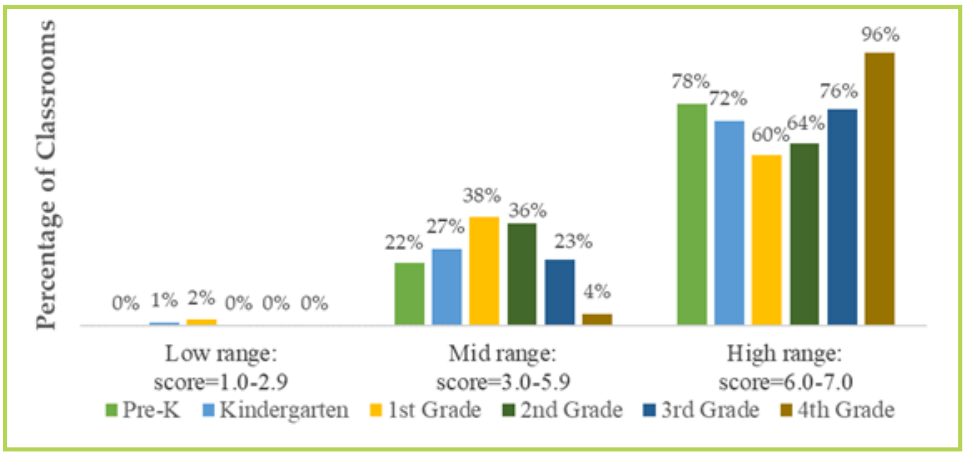
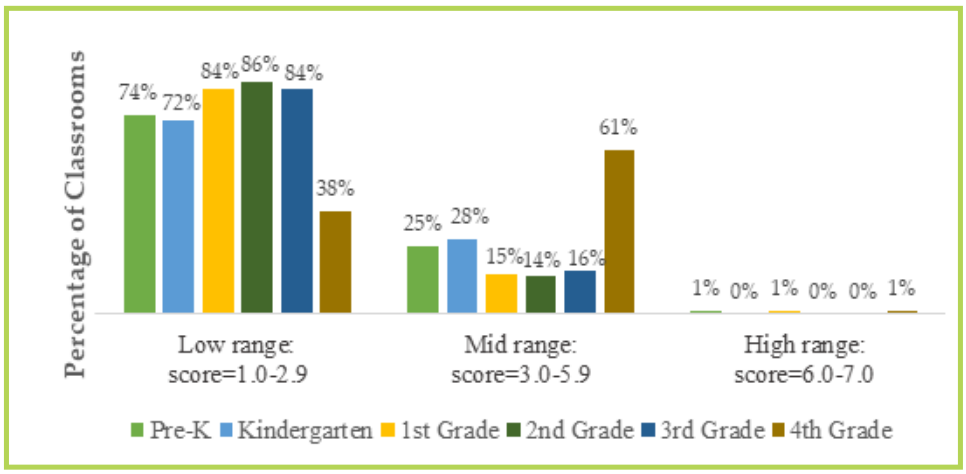


FIGURE 26. Frequency of CLASS Average Total Scores by Domain: Instructional Support



*Note: Scores changed in 4th grade due to a change in CLASS measures.

Classroom Quality and Child Outcomes

The average association between enrollment in high-quality classrooms versus enrollment in classrooms of moderate-to-low quality was small increases in longitudinal growth in phonemic awareness and letter identification (Letter-Word Identification subtest; see Figure 27). See Appendix Table D3 for HLM coefficients. Children could be enrolled in a high- or low-to-moderate-quality classroom each year. We examined the average effect of enrollment in high-quality classrooms across the years. Children who were in a higher proportion of high-quality classroom experiences initially had better skills, but children enrolled in a higher proportion of moderate-to-low-quality classrooms in kindergarten through 4th grade caught up to children in high-quality classrooms. Similar patterns were observed for the Word Attack subtest (see Figure 27). Enrollment in high-quality classrooms was associated with small math gains in pre-k (see Figure 28). See Appendix Table D4 for HLM coefficients. However, an opposite pattern was observed in 4th grade, with enrollment in low-to-moderate-quality classrooms associated with better scores. Patterns of growth related to problem behaviors and classroom quality changed over time. From the spring of kindergarten through the spring of 1st grade, enrollment in high-quality classrooms was associated with fewer behavior problems in 4th grade as compared to children in low- or moderate-quality classrooms (see Figure 29).

FIGURE 27. Growth in Letter-Word Identification and Word Attack Scores Over Time by Pre-K–4th Grade Classroom Quality

Enrollment in **high-quality classrooms** was associated with initial gains in pre-k, but growth was similar for children enrolled in high- and low- quality classrooms in kindergarten through 3rd grade. The average effect of enrollment in low-mid quality was associated with slightly better letter-word identification skills in 4th grade.

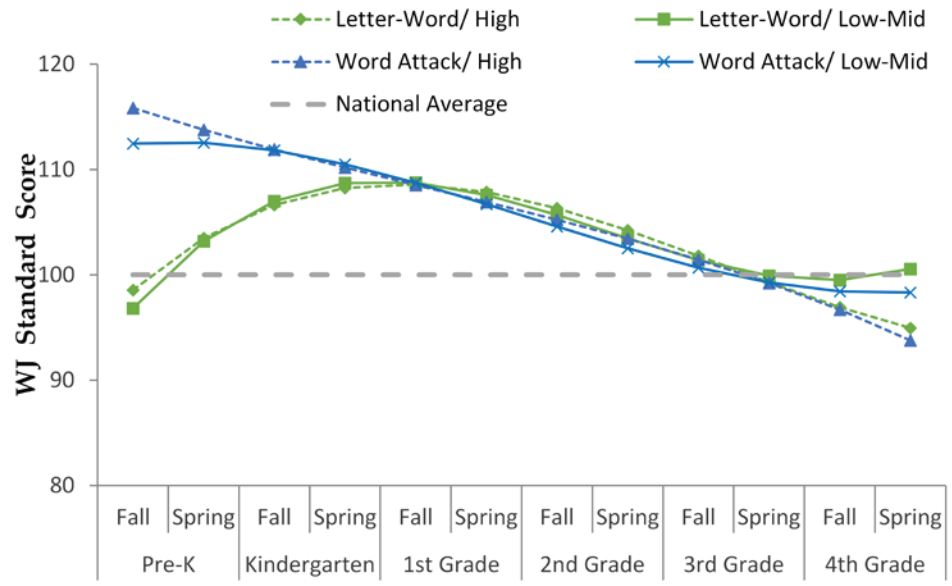


FIGURE 28. Growth in Applied Problems and Calculation Scores Over Time by Kindergarten–4th Grade Classroom Quality

Enrollment in **high-quality classrooms** was associated with small gains in pre-k, but enrollment in **low to moderate quality** classrooms was associated with gains in 4th grade on the Calculations Subtest. A similar pattern was observed for Applied Problems in 4th grade.

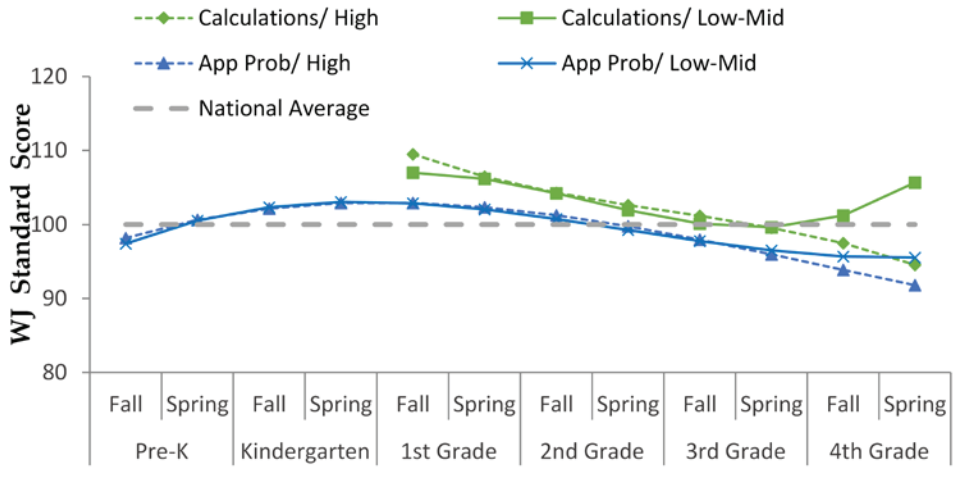
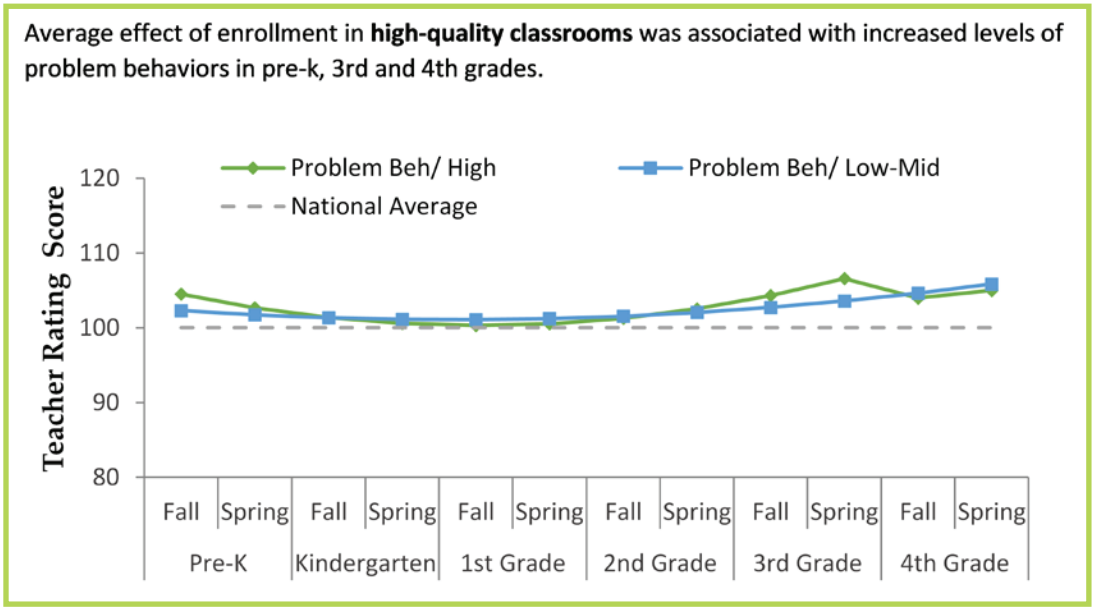


FIGURE 29. Growth in Problem Behaviors Over Time by Kindergarten-4th Grade Classroom Quality



Classroom Quality: Fourth Grade Sensitivity Analysis

Due to a change in the classroom quality measurement between 3rd and 4th grades that resulted in a higher proportion of classrooms falling into the high-quality range, sensitivity analyses were conducted with the original publisher recommended scoring, and a total score was estimated using a scoring structure as similar as possible to the Pre-K and K-3 versions. A detailed description of the measures is provided below. The Pre-K and K-3 versions of the Classroom Assessment Scoring System are composed of the same dimension and domain scores: (1) Emotional Support: Positive Climate, Negative Climate, Teacher Sensitivity, and Regard for Student Perspectives; (2) Classroom Organization: Behavior Management, Productivity, and Instructional Learning Formats; and (3) Instructional Support: Concept Development, Quality of Feedback, and Language Modeling. The Upper Elementary version used in 4th grade had a slightly different scoring structure: (1) Emotional Support: Positive Climate, Teacher Sensitivity, and Regard for Student Perspectives (no Negative Climate); (2) Classroom Organization: Behavior Management, Productivity, and Negative Climate (added Negative Climate and removed Instructional Learning Formats); (3) Instructional Support: Instructional Learning Formats, Content Understanding, Analysis & Inquiry, Quality of Feedback, and Instructional Dialogue (removed Concept Development and Language Modeling, added Content Understanding, Analysis & Inquiry, and Instructional Dialogue); and (4) additional dimension: Student Engagement. To estimate a total score using a scoring structure as similar as possible to the Pre-K and K-3 versions, the following adjustments were made to the 4th grade quality scores as part of the sensitivity analysis: (1) Positive Climate was included in the Emotional Support Domain score, (2) the Negative Climate was included in the Classroom Organization Domain score, (3) the Instructional Domain score was composed only of the Quality of Feedback dimension, and (4) Total Score was estimated using the following dimensions: Positive Climate, Teacher Sensitivity, Regard for Student Perspectives, Behavior Management, Productivity, and Negative Climate, Instructional Learning Formats, and Quality of Feedback (removed new dimensions added in the Upper Elementary version: Concept Development and Language Modeling, Content Understanding, Analysis & Inquiry, and Instructional Dialogue).

LONGITUDINAL STUDY RESULTS SUMMARY

Comparing pre-k baseline scores to 4th grade scores, some pre-k gains were sustained in reading decoding, decoding and spelling, and social skills.

The following section highlights key findings from the longitudinal portion of the current study, where children were assessed while enrolled in Georgia's Pre-K through 4th grade:

Literacy: Children who attended Georgia's Pre-K scored above the national average for reading skills related to decoding and phonics skills in kindergarten through 2nd grade, with declines to the national average in 3rd and 4th grades. Within-grade gains were biggest in pre-k for phonological awareness and letter-word identification. Phonological awareness skills declined in 3rd and 4th grades. Passage comprehension gains were below the national average, with small gains to the national average in 4th grade.

Language: Children's vocabulary skills started out within the average range relative to peers of the same age in pre-k and declined into the lower end of that range by 4th grade. Very little within-grade change was observed.

Math: Children's math skills were within the average range and were relatively stable over time for quantitative reasoning, math achievement, and math knowledge skills; they were at the higher end of the average range in 1st and 4th grades, with declines to scores closer to the average score for children of the same age in 2nd and 3rd grades. Within-grade gains were small for calculations in Grade 4.

Executive function: Children's executive function skills grew steadily during the study period.

Social Skills: Scores were close to the national average. Small gains in social skills were observed in pre-k, and moderate gains were observed in 4th grade. Behavior problems were in the low average range through 2nd grade and increased to the high average end of the range in 3rd and 4th grades.

Dual language learners: These children showed a similar pattern of growth in English and Spanish; however, vocabulary scores were in the lower end of the average range in English and improved over time, and Spanish vocabulary skills were in the below-average range. Problem behaviors were overall much lower than in the full sample.

Comparing pre-k baseline scores to 4th grade scores, some pre-k gains were sustained, with **children having higher-than-expected scores after controlling for child, family, and classroom characteristics in reading decoding (L-W ID), decoding and spelling (SA), and social skills.** Very small to small decreases in the other norm-referenced outcomes were observed related to the remaining language and literacy and math skills. No norming sample was available for the executive function outcomes. These results are consistent with other longitudinal studies of pre-k effects.[i, ii, iii]

Predictors of Longitudinal Outcomes

Most of the child and classroom-level moderators tested in this analysis were not associated with differential growth in child outcomes, including racial and ethnic backgrounds and pre-k characteristics—pre-k located in a public/private setting and total CLASS Pre-K score. However, children's English proficiency level at pre-k was associated with growth in language and quantitative reasoning. Children who started out with the lowest levels of English language skills showed the greatest growth over time in these domains. Children with IEPs had slower patterns of growth in vocabulary skills than children without an IEP at the beginning of pre-k, and girls had slightly fewer problem behaviors at pre-k, 3rd grade, and 4th grade.

Classroom Quality for Pre-K through 4th Grade

Overall classroom quality was largely in the middle range from pre-k through 4th grade, with shifts in domain scores in 4th grade. The proportion of classrooms that fell in the high-quality range was highest in pre-k and 4th grades. Children in classrooms with high overall quality had slightly higher literary scores in pre-k compared to children enrolled in low- and middle-quality classrooms; they also had higher math scores but more problem behaviors in 4th grade.

COMPARISON SUB-STUDY

Children who attended Georgia's Pre-K had significantly better scores than their peers who did not attend any pre-k program in literacy, oral language, and executive function outcomes.

The cohort of children included in the longitudinal study of Georgia's Pre-K was followed from pre-k through 4th grade. Tracking growth over time using standardized scores provides some understanding of how well children's skills grow over time relative to a norming sample of same-age peers. However, the longitudinal analyses did not provide insight into how children would have performed if they had no exposure to any childcare program prior to kindergarten enrollment. To compare 4th grade outcomes between children who enrolled in Georgia's Pre-K Program with children who had no exposure to pre-k prior to school-entry, a nested cohort sub-study was conducted to examine differences in academic and social outcomes between the two groups in the 3rd and 4th grades.

The comparison group was composed of 333 children who were enrolled in the same 3rd grade classrooms as the children in the longitudinal pre-k cohort in the 2017-2018 school year, but who had no experience with any pre-k during the 2013-2014 school year as reported by their parent or primary caregiver. Baseline assessments for the comparison group were administered in the fall of 3rd grade for the beginning of this separate sub-study, and assessments continued in the spring of 3rd grade and fall and spring of 4th grade. Differences in academic and social skills for children were estimated by comparing children who enrolled in Georgia's Pre-K with children who did not have any pre-k exposure.

Comparison Group Recruitment

To identify a similar comparison group to the longitudinal sample, the recruitment process for this sample group included contacting parents from the same 3rd grade classrooms attended by the longitudinal pre-k sample. For each classroom that had at least one participant retained in the longitudinal sample (n=667), 25 recruitment packets were sent home to parents. Parents who responded to the recruitment request completed a consent form and answered a short survey. The 3rd grade fall parent surveys were used both to screen participants for the comparison group and, if eligible, to provide baseline demographic information about students in the comparison group and their families. The surveys, only sent to comparison group families, included information about parent education levels, income, languages spoken in the home, and number of adults/children in the home. In order to assure that the children in the comparison group had not attended pre-k, parents of children in the comparison group were also asked if their child had attended pre-k/preschool as a four-year-old, and if so, the number of hours/day and months attended. Parent surveys were distributed to families through the classrooms and returned in sealed envelopes to teachers for retrieval by the research team. All children in the comparison group received the same assessments as the students in the longitudinal sample described above.

From these 3,682 recruitment survey responses, most were excluded since they attended some type of pre-k, another type of child care, or educational program (91%). All children with permission to participate who had not previously attended a pre-k program were included in the sub-study. Therefore, permission for the 333 (9%) children was gathered from third graders who had no pre-k experience (birthdates ranging from 9/2/2008-9/1/2009).

Comparison Group Participants

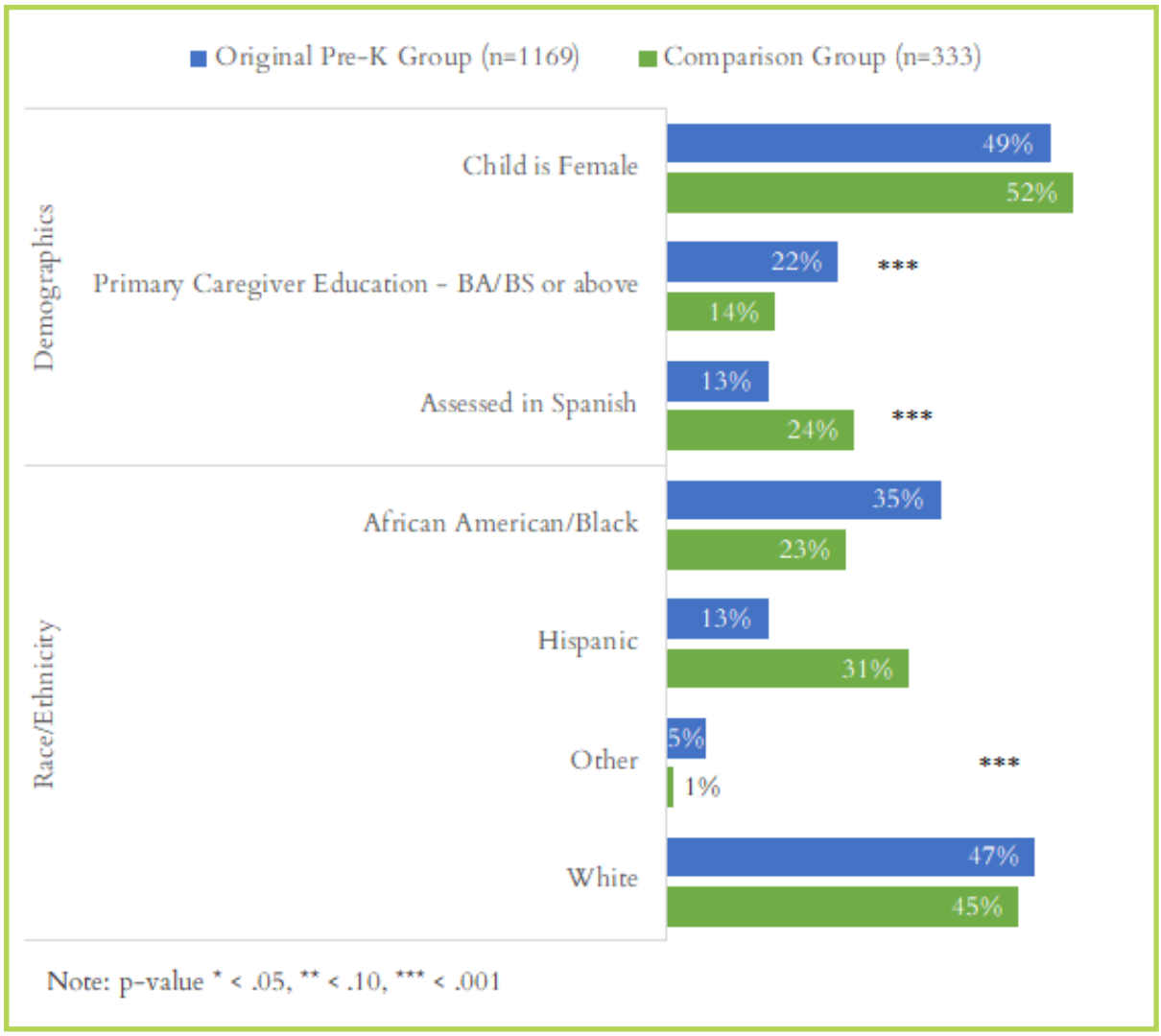
The comparison group included 333 children in 3rd grade (80 of whom were DLLs) and 254 in 4th grade (57 of whom were DLLs). See Table 5. Demographic and family information about child and family characteristics for the comparison group in 4th grade was obtained from parent survey data in 3rd grade (see Figure 30). To examine the degree to which the longitudinal study sample and the comparison sample differed in meaningful ways that may be related to outcomes, we compared demographic and family characteristics for the two groups. Results from t-test comparisons between the longitudinal pre-k sample and the comparison group characteristics at 4th grade showed that the comparison group was more likely to be Hispanic (31% versus 13%), African American/Black (23% versus 35%), less likely to be white (45% versus 47%), had primary caregivers with less education (14% with a bachelor's degree versus 22%), lower income (134% of the federal poverty level versus 163%), and slightly larger household size (5 people versus 4).

TABLE 5. Number of Classrooms and Children Participating in the Comparison Group Sub-Study

Study Year	Total Children in Comparison Group	DLL – Comparison Group
Grade 3 (2017-18)	333	80
Grade 4 (2018-19)	254 (76%)	57 (71%)

Note: Percentages noted in parentheses are the proportion of the original sample that was retained in the study by grade.

FIGURE 30. 4th Grade Comparison Group versus Original Pre-K Group

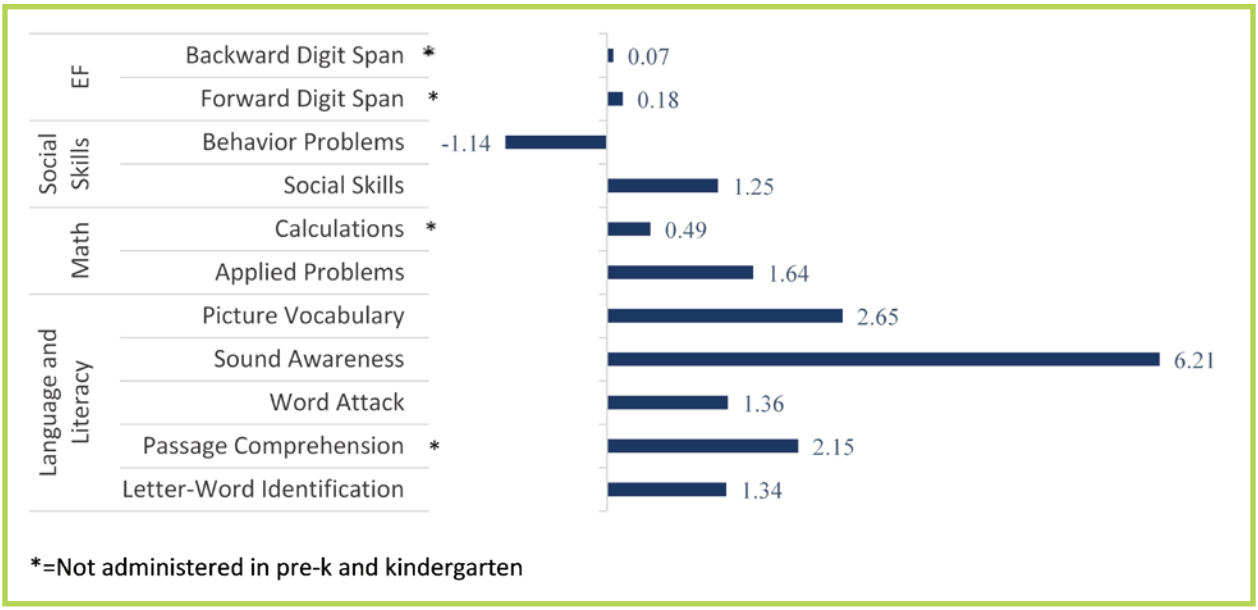


Comparison Group Analyses

To answer Research Question 4:"Does academic performance in 3rd and 4th grade differ between children who enrolled in Georgia's Pre-K compared to children who had no pre-k experience prior to kindergarten entry?" we examined whether there are differences in the learning outcomes in 3rd and 4th grade for children who attended Georgia's Pre-K compared with outcomes for children with no pre-k experience. To estimate differences in academic and social outcomes at the fall of 4th grade for children who had been enrolled in Georgia's Pre-K Program compared to children who were not enrolled in any pre-k program, separate generalized linear models were fit for each outcome. These models controlled for the demographic characteristics on which the comparison group differed (see Figure 31 and Appendix Table E1).

Children who attended Georgia's Pre-K had significantly better scores on three of the five language and literacy outcomes in 3rd and 4th grade compared to children who did not attend any pre-k program, including Picture Vocabulary, Sound Awareness, and Passage Comprehension (see Figure 31 and Appendix Table E1). Differences were also statistically significant for executive function skills as measured by the Forward Digit Span task. Differences in language and literacy outcomes were in the small to moderate range (Effect Size = 0.14-0.41). The largest effect size was associated with Sound Awareness, where children who attended Georgia's Pre-K had moderately higher scores than their peers who did not attend pre-k (Effect Size = 0.41).

FIGURE 31. Differences in Standard and Adjusted Mean Scores for Pre-K Group versus Comparison Group in 4th Grade



Summary and Conclusions on the Comparison Sub-Study

Children who attended Georgia's Pre-K had significantly better scores than their peers who did not attend any pre-k program on the following outcomes:

Literacy

- Reading comprehension and lexical knowledge (Passage Comprehension subtest)

Oral Language

- Rhyming and manipulating phonemes (Sound Awareness subtest)
- Identifying and naming pictures (Picture Vocabulary subtest)

Executive Function

- Attention efficiency and capacity (Forward Digit Span task)

The largest difference between the comparison and pre-k groups was related to the Sound Awareness subtest. This subtest assesses auditory processing skills that underlie decoding and spelling as basic skills for acquiring literacy. In the sequence of skills necessary for literacy acquisition, these skills are some of the least complex and are most likely to be targeted with young children in a pre-k program. The difference between groups was small to moderate in size in 4th grade (Effect Size = 0.41). This is similar to the fall-to-spring gain in pre-k for the longitudinal study group (Effect Size = 0.61). If these skills are not acquired early in literacy acquisition, they are unlikely to be taught at a later point in time. Given the differences in the two samples on key demographic characteristics and reasons for selecting into attending Georgia's Pre-K, these results should be cautiously interpreted as some of the results may be related to those differences and not exclusively to differences in outcomes related to pre-k enrollment.

STUDY IMPLICATIONS

Overall, children who attended Georgia's Pre-K were prepared for kindergarten, and a key strength of the pre-k program was building foundational literacy skills for later reading.

Georgia has one of the few fully implemented state-funded universal pre-kindergarten programs in the United States, available to all 4-year-olds in the state[xii]. Georgia's Pre-K quality standards and program requirements are founded on knowledge about high-quality practices that support young children's growth across multiple domains of development. The Georgia General Assembly funds Bright from the Start: Georgia Department of Early Care and Learning (DECAL) to contribute to knowledge about pre-k best practices and continuous quality improvement by investing and engaging in ongoing research related to its pre-k program. As a part of these ongoing research efforts, DECAL has established a model state-university partnership with the FPG Child Development Institute at the University of North Carolina at Chapel Hill. DECAL and FPG have worked collaboratively on evaluation and research projects since 2007, providing an example for other states for ongoing efforts to establish and improve programs for young children.

Overall, children who participated in Georgia's Pre-K in this study were prepared for kindergarten, with a key strength of the pre-k program found in building foundational literacy skills for later reading. While students in the study were at or above the national average at the end of 4th grade on most measures, the rate of growth in pre-k and kindergarten is not maintained in later grades (except for larger growth in social skills and math in 4th grade). Results from this six-year longitudinal study of children who attended Georgia's Pre-K contribute knowledge about the long-term outcomes for children enrolled in universal statewide pre-k programs. These results can also be used for understanding program effectiveness, continuous quality improvement, and decision-making for future activities and funding. In addition to understanding growth in academic and social skills during the pre-k year, this study provides a picture of what happens in the years after students attend pre-k and provides a glimpse into the skills that children bring with them when they enter pre-k that may inform the development of programming to support children prior to pre-k entry as well.

Language Skills and Supporting Individual Children's Needs

Language skills, vocabulary knowledge, and semantics facilitate the development of other academic and social skills[xiii]. On average, children in this study started pre-k with language skills below the national norm and maintained age-related growth in pre-k, but their skills decreased slightly relative to peers of the same age from kindergarten through 3rd grade, with a slight increase in 4th grade. Additionally, children who started pre-k with lower vocabulary, semantic-related language use, and/or an IEP continued to have lower language, math, and social skills scores through 4th grade. A closer look into how individual children's language skills can be supported in pre-k and continued in an aligned manner afterward is warranted.

Dual Language Advantage and Need for Home Language Support

Consistent with previous research, children whose home language was Spanish excelled in their ability to apply phonological awareness and decoding skills to nonsense words relative to their monolingual English speaking peers[xiv]. Growth in English early literacy and language skills was slightly higher than their peers, and they ended up with higher scores when assessed in English throughout the study. DLL reading comprehension skills in English also mirrored their peers. Math skills assessed in English and teacher-rated social skills were also much higher than their peers. Despite these strengths, DLL language and literacy skills in their home language declined relative to same-age peers over time. Finally, DLL had higher scores on the Backward Digit Span,

but not Forward Digit Span executive function measures by 4th grade. Although DLL children benefited from their knowledge of at least two languages, their growth in Spanish was not supported in the same manner. Given the relative advantages of bilingualism on children's skills in this study, support for further home language acquisition could be considered.

Classroom Quality

While classroom quality is highest in pre-k, most classrooms were in the moderate (74-83%) to high range (9-26%) of quality throughout the study and slightly higher with a change in measures in 4th grade (69% of classrooms in the high quality range). Patterns of quality were fairly consistent with these ranges for the Emotional Support and Classroom Organization Domains of the observational measure, with Instructional Support scores in the expected lower end of the range, as is consistent with other observational studies of pre-k and early elementary quality. As with other longitudinal research on the effects of classroom quality, associations with child outcomes were small and nuanced and are likely only fully maximized with continuous exposure to high-quality classrooms[xv]. Enrollment in higher-quality classrooms was associated with better literacy, math, and social skills outcomes, but slightly more behavior problems in pre-k. Enrollment in low or moderate quality classrooms in 4th grade was associated with small benefits in math, and enrollment in high quality classrooms was associated with more behavior problems. In this study we examined the average effect of having been enrolled in high-quality classrooms across grades, so it's possible that the grade-specific patterns that we observed with this approach were influenced by the number of years that children spent in high-quality classrooms. A further examination of the association between dosage (number of years) of enrollment in high-quality classrooms and outcomes is a potential next step. Further, most of the classrooms observed in this study were not at the lowest range of quality, where we would expect the biggest differences in children's outcomes.

Changes in Academic and Social Skills Over Time and Curricular Alignment

This study also allows us to examine the degree to which children's growth was consistent over time, highlighting the potential influence of curricular alignment across grades. The use of learning standards may lead to emphasis on particular skills and benchmarks in specific grades, but by looking at skill domains over time in this study we may also see where there may be a need for more continuous support. For example, children's phonological awareness skills, knowledge of letter-sound correspondence and blending skills, and reading comprehension showed a rapid rate of growth from entry into pre-k through the spring of kindergarten to levels well above the national average. However, those literacy skills did not continue to increase at the same rate after kindergarten, and by third grade average skill-levels among the same students were falling below the national norm. Although foundational literacy skills are largely acquired by 1st grade, these results suggest that children may benefit from continued support later in their elementary years through increased curricular alignment. A similar pattern was observed for math skills, with the exception being a steep increase in rate of growth for the calculations subtest in 4th grade when student ability to perform mathematical computations increased to levels similar to those observed in 1st grade, suggesting a possible curricular focus in this skill set that may not have been emphasized as much in the 2nd and 3rd grades. The development of social skills is another area where we see most growth taking place during the pre-k year, when a specific focus on developing friendships and the ability to engage successfully in a group environment are emphasized. Social skills then stayed at similar levels from the end of pre-k through the fall of 3rd grade when children had

a steep increase in these skills through 4th grade relative to same-age peers. These results suggest that some combination of classroom support and possible developmental change in 3rd grade support the growth of social skills in a similar way to what children experienced in their pre-k year. Additionally, the continued rate of growth from pre-k to kindergarten suggests potential curricular alignment those years with increasing mismatch with the skill sets in 1st and 2nd grades.

In summary, this longitudinal study reflects a deep investment in ensuring that children fully benefit from the alignment between pre-k through 4th grade. Children experienced moderate-to-high-quality instruction from pre-k through 4th grade and had higher scores than would be expected due to age on most outcomes, particularly related to literacy, although increased curricular alignment after kindergarten may be beneficial. Language skills are foundational and there was little growth on average over time. Children with low language skills at pre-k entry also continued to have lower language, math, and social skills over time. Further investigation into supports for language skill growth is warranted, with particular attention to supporting home language acquisition in addition to acquisition of English language skills.

Appendix A. Classroom and Child Characteristics

TABLE A1. Characteristics of Classrooms and Teachers (Pre-K–4th Grade)

	Pre-K		Kindergarten		1st Grade		2nd Grade		3rd Grade		4th Grade	
	n	Mean/%	n	Mean/%	n	Mean/%	n	Mean/%	n	Mean/%	n	Mean/%
Classroom Characteristics												
Class size	189	21.1	772	20.1	694	20.7	728	21.0	581	21.5	648	23
Percentage of boys	189	50.1%	768	51.3%	695	51.0%	725	51.0%	581	51.4%	648	51.0%
% of children with home language other than English	199	11.2%	770	14.7%	687	12.0%	700	11.0%	580	14.5%	637	14.3%
Teacher Characteristics												
Years teaching current grade level	184	5.8	777	8.5	695	6.8	728	5.9	582	5.9	645	5.5
Years teaching at current school	184	5.1	774	8.8	693	8.5	729	8.5	582	7.8	646	7.4
Total years teaching	179	11.1	777	14.5	688	14.2	728	13.7	582	12.7	647	13.8
Female	186	98.4%	762	98.1%	688	99.1%	708	97.1%	582	95.%	605	93.0%
Race/Ethnicity												
White	128	67.4%	596	76.7%	492	69.6%	523	71.9%	427	73.4%	484	74.80%
Black	52	27.4%	143	18.4%	164	23.2%	168	23.1%	113	19.4%	134	20.70%
Hispanic/Latino	5	2.7%	13	1.7%	16	2.3%	16	2.2%	20	3.4%	13	2.00%
Other	5	2.6%	25	3.2%	35	5.0%	20	2.8%	22	3.8%	16	2.50%
Georgia PSC Certification	162	81.8%	-	-	-	-	-	-			-	-
Highest Degree Earned												
Associate's Degree	4	2.2%	1	0.1%	2	0.3%	0	0.0%	1	0.2%	0	0.0%
Bachelor's Degree	118	64.1%	293	37.8%	245	35.3%	275	37.7%	186	32.0%	207	31.9%
Master's Degree	52	28.3%	338	43.5%	300	43.2%	446	61.2%	267	46.0%	279	43.0%
Education Specialist	9	4.9%	135	17.4%	138	19.9%	0	0.0%	120	20.6%	157	24.2%
PhD/EdD/PsyD	1	0.5%	10	1.2%	9	1.3%	8	1.1%	7	1.2%	6	0.9%

TABLE A2. Characteristics of Children in the 4th Grade Longitudinal Cohort Over Time

	Pre-K		Kindergarten		1st Grade		2nd Grade		3rd Grade		4th Grade	
	n	Mean/%	n	Mean/%	n	Mean/%	n	Mean/%	n	Mean/%	n	Mean/%
Child Characteristics												
Child's age on Sep 1	1169	(4.5)	1034	(5.5)	969	(6.5)	951	(7.5)	857	(8.5)	715	(9.5)
Female	601	(51%)	538	(52%)	491	(51%)	483	(51%)	434	(51%)	353	(49%)
Hispanic	170	(15%)	152	(15%)	150	(16%)	143	(15%)	132	(15%)	92	(13%)
Race/Ethnicity												
White	616	(53%)	552	(53%)	514	(53%)	397	(42%)	366	(43%)	334	(47%)
Black/African American	448	(38%)	393	(38%)	369	(38%)	352	(37%)	307	(36%)	252	(35%)
Latino									132	(15%)	92	(13%)
Multi-racial/Other	105	(9%)	89	(7%)	86	(9%)	59	(6%)	52	(6%)	37	(5%)
Limited English language proficiency (designated by DECAL)	116	(10%)	108	(10%)	101	(10%)	95	(10%)	86	(10%)	63	(9%)
Individualized Education Program (IEP)	34	(3%)	32	(3%)	30	(3%)	27	(3%)	22	(3%)	19	(3%)
Family Characteristics												
Income												
Category One[a]	636	(54%)	571	(55%)	522	(54%)	512	(54%)	454	(53%)	377	(53%)
Category Two[b]	533	(46%)	463	(45%)	447	(46%)	439	(46%)	403	(47%)	338	(47%)
Percent Federal Poverty Level at pre-k	803	(163%)	787	(176%)	730	(186%)	794	(186%)	618	(199%)	551	(191%)
Household Size	851	(4.6) (2,14)	842	(4.6) (2,12)	788	(4.5) (2,12)	830	(4.6) (2,12)	628	(4.5) (2,12)	564	4.5 (2,12)
Primary caregiver education												
< High School	120	(10%)	95	(9%)	78	(8%)	78	(8%)	46	(5%)	46	(6%)
High School-GED	263	(23%)	184	(18%)	175	(18%)	163	(17%)	127	(15%)	124	(17%)
Some College	749	(64%)	584	(56%)	546	(57%)	587	(62%)	447	(52%)	391	(55%)

[a] Income Category One: Participation in one or more programs that were income-dependent, including SNAP, TANF, SSI, CAPS, Medicaid, and free or reduced-price meals.
[b] Income Category Two: Child was not eligible for income-dependent programs.

Appendix B. Measures and Adjusted Mean Scores on Child Outcome Measures

TABLE B1. Child Assessment and Classroom Observation Measure Overview

Measure	Scoring
Language and Literacy Skills	
Vocabulary	
WJ-III[v] Picture Vocabulary (Subtest 14) / Bat-III[vi] Vocabulario sobre Dibujos (Prueba 14)	Standard score Mean=100, SD=15
Phonological Awareness	
WJ-III Sound Awareness (Subtest 21) / Bat-III Discernimiento de Sonidos (Prueba 21)	Standard score Mean=100, SD=15
Letter-and Word Recognition	
WJ-III Letter-Word Identification (Subtest 1) / Bat-III Identificación de Letras y Palabras (Prueba 1)	Standard score Mean=100, SD=15
Written Comprehension	
WJ-III Passage Comprehension (Subtest 9) / Bat-III Comprensión de Textos (Prueba 9)	Standard score Mean=100, SD=15
Phonemic Awareness and Decoding	
WJ-III Word Attack (Subtest 13) / Bat-III Análisis de Palabras (Prueba 13)	Standard score Mean=100, SD=15
Math Skills	
Written Calculation	Standard score
WJ-III Calculation (Subtest 5) / Bat-III Cálculo (Prueba 5)	Mean=100, SD=15
Math Problem-Solving	Standard score
WJ-III Applied Problems (Subtest 10) / Bat-III Problemas Aplicados (Prueba 10)	Mean=100, SD=15
Executive Function	
Working Memory	Unstandardized score
Forward Digit Span and Backward Digit Span (English/Spanish)[vii]	Range=1-8
Behavior Skills	
Social Skills	Standard score
SSiS[viii] Social Skills subscale	Mean=100, SD=15
Problem Behaviors	Standard score
SSiS Problem Behaviors subscale	Mean=100, SD=15
Classroom Quality	
Teacher-Child Instructional Interactions	
CLASS[ix] / CLASS K-3[x]/CLASS Upper Elementary[xi] Emotional Support, Classroom Observation, Instructional Support Domains & Total	Total and Domain scores range=1.0–7.0

TABLE B2. Child Adjusted Mean Outcomes for Full Longitudinal Sample (Pre-K–1st Grade)[a]

	Pre-K				Kindergarten				1st Grade			
	Fall		Spring		Fall		Spring		Fall		Spring	
Measure	Mean	(SE)	Mean	(SE)	Mean	(SE)	Mean	(SE)	Mean	(SE)	Mean	(SE)
Language & Literacy												
WJ-III Picture Vocabulary[a,b]	95.86	(.70)	95.59	(.69)	95.14	(.69)	94.57	(.70)	93.94	(.70)	93.34	(.71)
WJ-III Sound Awareness[b,c]	87.49	(1.38)	96.66	(1.35)	102.91	(1.36)	106.66	(1.37)	108.30	(1.38)	108.23	(1.39)
WJ-III Letter-Word Identification[b,c]	97.07	(1.12)	103.10	(1.11)	106.81	(1.11)	108.60	(1.12)	108.85	(1.13)	107.94	(1.14)
WJ-III Word Attack[b,c]	110.42	(1.01)	111.54	(.97)	111.48	(.97)	110.52	(.97)	108.90	(.98)	106.88	(.99)
WJ-III Passage Comprehension[b,c]	-	-	-	-	-	-	-	-	98.13	(1.05)	96.56	(1.04)
Math												
WJ-III Calculation[b,c]	-	-	-	-	-	-	-	-	107.06	(1.20)	106.06	(1.18)
WJ-III Applied Problems[a,b]	98.74	(.89)	100.80	(.87)	101.84	(.88)	102.05	(.88)	101.59	(.89)	100.63	(.89)
Executive Function												
Forward Digit Span[d]	-	-	-	-	-	-	-	-	4.16	(.08)	4.40	(.07)
Backward Digit Span[d]	-	-	-	-	-	-	-	-	2.20	(.05)	2.33	(.05)
Behavior Skills												
SSiS Social Skills[b]	94.02	(1.30)	97.26	(1.26)	98.85	(1.28)	99.25	(1.29)	98.92	(1.30)	98.31	(1.31)
SSiS Problem Behaviors[b]	101.73	(1.31)	101.13	(1.31)	100.72	(1.29)	100.50	(1.31)	100.46	(1.32)	100.60	(1.34)

TABLE B2 (continued). Child Adjusted Mean Outcomes for Full Longitudinal Sample (2nd–4th Grade)[a]

	2nd Grade				3rd Grade				4th Grade			
	Fall		Spring		Fall		Spring		Fall		Spring	
Measure	Mean	(SE)	Mean	(SE)	Mean	(SE)	Mean	(SE)	Mean	(SE)	Mean	(SE)
Language & Literacy												
WJ-III Picture Vocabulary[a,b]	92.81	(.71)	92.42	(.72)	92.25	(.73)	92.36	(.76)	92.81	(.85)	93.67	(1.06)
WJ-III Sound Awareness[b,c]	106.84	(1.40)	104.55	(1.42)	101.73	(1.44)	98.80	(1.49)	96.14	(1.65)	94.17	(2.02)
WJ-III Letter-Word Identification[b,c]	106.25	(1.15)	104.17	(1.16)	102.08	(1.18)	100.37	(1.22)	99.41	(1.30)	99.60	(1.49)
WJ-III Word Attack[b,c]	104.71	(1.00)	102.65	(1.01)	100.95	(1.03)	99.87	(1.07)	99.66	(1.18)	100.58	(1.43)
WJ-III Passage Comprehension[b,c]	94.43	(1.04)	92.44	(1.05)	91.30	(1.06)	91.71	(1.09)	94.36	(1.30)	99.96	(2.05)
Math												
WJ-III Calculation[b,c]	103.78	(1.18)	101.26	(1.18)	99.60	(1.20)	99.84	(1.24)	103.07	(1.70)	110.35	(3.18)
WJ-III Applied Problems[a,b]	99.36	(.90)	97.88	(.91)	96.43	(.92)	95.16	(.95)	94.24	(1.08)	93.83	(1.37)
Executive Function												
Forward Digit Span[d]	4.51	(.07)	4.57	(.07)	4.64	(.08)	4.81	(.08)	5.12	(.11)	5.67	(.21)
Backward Digit Span[d]	2.46	(.05)	2.60	(.05)	2.73	(.05)	2.87	(.06)	3.00	(.06)	3.13	(.06)
Behavior Skills												
SSiS Social Skills[b]	97.88	(1.32)	98.08	(1.34)	99.38	(1.36)	102.23	(1.44)	107.09	(1.68)	114.42	(2.22)
SSiS Problem Behaviors[b]	100.93	(1.35)	101.44	(1.37)	102.14	(1.39)	103.02	(1.44)	104.08	(1.51)	105.33	(1.62)

[a] Growth models were adjusted for child sex, IEP, ethnicity, race, English proficiency, poverty status, public preschool site, class size, and classroom quality.
[b] Indicates standard score on norm-referenced measure with mean=100, SD=15.
[c] Scores reflect use of updated normative tables (2007).
[d] Possible range=1-8.

TABLE B3. Child Adjusted Mean Outcomes for DLL Sub-Sample in English (Pre-K–1st Grade)[a]

Measure	Pre-K				Kindergarten				1st Grade			
	Fall		Spring		Fall		Spring		Fall		Spring	
	Mean	(SD)	Mean	(SD)	Mean	(SD)	Mean	(SD)	Mean	(SD)	Mean	(SD)
Language & Literacy												
WJ-III Picture Vocabulary[b,c]	83.39	(2.43)	86.68	(2.40)	89.01	(2.41)	90.13	(2.43)	90.53	(2.44)	90.52	(2.45)
WJ-III Sound Awareness[b,c]	81.96	(3.48)	90.54	(3.39)	97.31	(3.42)	102.34	(3.46)	105.70	(3.48)	107.47	(3.51)
WJ-III Letter-Word Identification[b,c]	92.85	(3.59)	101.27	(3.56)	106.72	(3.57)	109.68	(3.59)	110.64	(3.61)	110.09	(3.63)
WJ-III Word Attack[b,c]	110.11	(3.42)	112.49	(3.24)	113.35	(3.24)	113.02	(3.26)	111.80	(3.29)	110.00	(3.31)
WJ-III Passage Comprehension[b,c]	-	-	-	-	-	-	-	-	96.82	(3.09)	95.43	(3.06)
Math												
WJ-III Calculation[b,c]	-	-	-	-	-	-	-	-	109.30	(3.24)	108.70	(3.18)
WJ-III Applied Problems[b,c]	92.61	(2.63)	99.59	(2.57)	103.62	(2.59)	105.29	(2.61)	105.19	(2.63)	103.91	(2.65)
Executive Function												
Forward Digit Span[d]	-	-	-	-	-	-	-	-	4.08	(4.19)	4.19	(.21)
Backward Digit Span[d]	-	-	-	-	-	-	-	-	2.54	(.15)	2.70	(.15)
Behavior Skills												
SSiS Social Skills[b]	92.54	(3.21)	97.37	(3.10)	100.19	(3.14)	101.50	(3.17)	101.82	(3.17)	101.65	(3.19)
SSiS Problem Behaviors[b]	95.84	(2.21)	94.84	(2.13)	94.07	(2.13)	93.53	(2.15)	93.22	(2.18)	93.15	(2.20)

TABLE B3 (continued). Child Adjusted Mean Outcomes for DLL Sub-Sample in English (2nd–4th Grade)[a]

Measure	2nd Grade				3rd Grade				4th Grade			
	Fall		Spring		Fall		Spring		Fall		Spring	
	Mean	(SD)	Mean	(SD)	Mean	(SD)	Mean	(SD)	Mean	(SD)	Mean	(SD)
Language & Literacy												
WJ-III Picture Vocabulary[b,c]	90.40	(2.47)	90.47	(2.49)	91.03	(2.52)	92.40	(2.60)	94.88	(2.87)	98.76	(3.50)
WJ-III Sound Awareness[b,c]	107.73	(3.55)	106.55	(3.60)	104.01	(3.67)	100.18	(3.86)	95.15	(4.42)	88.99	(5.67)
WJ-III Letter-Word Identification[b,c]	108.49	(3.67)	106.35	(3.71)	104.13	(3.76)	102.33	(3.85)	101.43	(4.09)	101.92	(4.64)
WJ-III Word Attack[b,c]	107.94	(3.34)	105.93	(3.38)	104.27	(3.44)	103.28	(3.56)	103.28	(3.89)	104.56	(4.67)
WJ-III Passage Comprehension[b,c]	93.43	(3.06)	91.44	(3.09)	90.08	(3.15)	89.98	(3.25)	91.75	(3.96)	96.01	(6.42)
Math												
WJ-III Calculation[b,c]	105.01	(3.16)	100.76	(3.19)	98.47	(3.24)	100.67	(3.36)	109.90	(4.63)	128.67	(8.68)
WJ-III Applied Problems[b,c]	102.02	(2.69)	100.12	(2.73)	98.80	(2.79)	98.63	(2.92)	100.20	(3.31)	104.11	(4.18)
Executive Function												
Forward Digit Span[d]	4.28	(.21)	4.38	(.21)	4.49	(.22)	4.62	(.22)	4.80	(.33)	5.04	(.64)
Backward Digit Span[d]	2.86	(.14)	3.02	(.15)	3.18	(.15)	3.34	(.16)	3.49	(.17)	3.65	(.18)
Behavior Skills												
SSiS Social Skills[b]	101.51	(3.21)	101.92	(3.25)	103.37	(3.31)	106.39	(3.55)	111.48	(4.28)	119.16	(5.88)
SSiS Problem Behaviors[b]	93.31	(2.21)	93.69	(2.22)	94.32	(2.27)	95.17	(2.36)	96.25	(2.55)	97.57	(2.85)

[a] Growth models were adjusted for child sex, IEP, ethnicity, race, English proficiency, poverty status, public preschool site, class size, and classroom quality.
[b] Indicates standard score on norm-referenced measure with mean=100, SD=15.
[c] Scores reflect use of updated normative tables (2007).
[d] Possible range=1-8.

TABLE B4. Child Adjusted Mean Outcomes for DLL Sub-Sample in Spanish (Pre-K–1st Grade)[a]

Measure	Pre-K				Kindergarten				1st Grade			
	Fall		Spring		Fall		Spring		Fall		Spring	
	Mean	(SD)	Mean	(SD)	Mean	(SD)	Mean	(SD)	Mean	(SD)	Mean	(SD)
Language & Literacy												
WJ-III Picture Vocabulary[b,c]	70.44	(4.76)	66.21	(4.71)	62.69	(4.71)	59.89	(4.72)	57.83	(4.73)	56.51	(4.75)
WJ-III Sound Awareness[b,c]	74.36	(4.11)	78.58	(3.99)	82.29	(3.99)	85.33	(3.99)	87.55	(3.99)	88.78	(4.00)
WJ-III Letter-Word Identification[b,c]	89.87	(4.08)	85.76	(3.98)	82.68	(3.98)	80.53	(3.99)	79.17	(4.02)	78.49	(4.06)
WJ-III Word Attack[b,c]	110.38	(3.36)	104.15	(3.14)	98.98	(3.11)	94.80	(3.11)	91.49	(3.13)	88.96	(3.16)
WJ-III Passage Comprehension[b,c]	-	-	-	-	-	-	-	-	79.64	(4.33)	79.21	(4.27)
Math												
WJ-III Calculation[b,c]	-	-	-	-	-	-	-	-	99.77	(5.00)	101.00	(4.89)
WJ-III Applied Problems[a,b]	86.91	(3.80)	90.87	(3.75)	92.90	(3.74)	93.33	(3.74)	92.51	(3.74)	90.81	(3.74)

TABLE B4 (continued). Child Adjusted Mean Outcomes for DLL Sub-Sample in Spanish (2nd–4th Grade)[a]

Measure	2nd Grade				3rd Grade				4th Grade			
	Fall		Spring		Fall		Spring		Fall		Spring	
	Mean	(SD)	Mean	(SD)	Mean	(SD)	Mean	(SD)	Mean	(SD)	Mean	(SD)
Language & Literacy												
WJ-III Picture Vocabulary[b,c]	55.94	(4.77)	56.13	(4.80)	57.10	(4.84)	58.86	(4.96)	61.40	(5.35)	64.76	(6.30)
WJ-III Sound Awareness[b,c]	88.86	(4.04)	87.63	(4.08)	84.93	(4.12)	80.60	(4.30)	74.49	(4.91)	66.43	(6.34)
WJ-III Letter-Word Identification[b,c]	78.35	(4.12)	78.66	(4.20)	79.27	(4.28)	80.07	(4.48)	80.94	(5.00)	81.75	(6.19)
WJ-III Word Attack[b,c]	87.13	(3.22)	85.89	(3.28)	85.15	(3.35)	84.81	(3.50)	84.78	(3.94)	84.96	(4.95)
WJ-III Passage Comprehension[b,c]	76.03	(4.23)	71.91	(4.23)	68.68	(4.26)	68.16	(4.38)	72.16	(5.50)	82.50	(9.29)
Math												
WJ-III Calculation[b,c]	99.57	(5.00)	97.11	(4.83)	95.20	(4.88)	95.46	(5.04)	99.50	(6.54)	108.93	(11.45)
WJ-III Applied Problems[a,b]	88.56	(3.76)	86.12	(3.77)	83.84	(3.79)	82.06	(3.86)	81.14	(4.13)	81.42	(4.82)

[a] Growth models were adjusted for child sex, IEP, ethnicity, race, English proficiency, poverty status, public preschool site, class size, and classroom quality.
[b] Indicates standard score on norm-referenced measure with mean=100, SD=15.
[c] Scores reflect use of updated normative tables (2007).

Appendix C. HLM Coefficient Tables for Moderation Analyses

TABLE C1. Full Sample HLM Results with Moderation for Language Proficiency Level — Language and Literacy Skills

	Literacy and Language				
	Picture Vocabulary (PK-4)	Sound Awareness (PK-4)	Letter-Word (PK-4)	Word Attack (PK-4)	Passage Comprehension (1-4)
	Est (SE)	Est (SE)	Est (SE)	Est (SE)	Est (SE)
Ethnicity					
Hispanic	-2.48 (1.36)	0.45 (2.49)	-1.09 (1.83)	4.00 (2.73)	2.57 (1.97)
Race					
White	-1.22 (0.60)	-1.66 (1.12)	3.96 (0.88)	0.82 (0.96)	0.99 (0.83)
Black	-0.45 (0.94)	2.70 (1.72)	6.85 (1.37)	5.12 (1.43)	3.74 (1.34)
Multi-Other	--	--	--	--	--
Moderation by Time					
English Proficiency					
Level 1	5.19 (0.84)	-1.98 (1.57)	2.39 (0.95)	-3.97 (1.72)	0.54 (1.73)
Level 2 or 3	1.26 (0.51)	-1.35 (0.94)	-0.21 (0.57)	-1.50 (0.91)	-2.18 (1.04)
Level 4	1.44 (0.42)	0.97 (0.79)	0.68 (0.48)	0.66 (0.72)	-0.69 (0.87)
Level 5	--	--	--	--	--
Moderation by Time*Time					
English Proficiency					
Level 1	-1.01 (0.22)	0.81 (0.42)	-0.51 (0.25)	0.77 (0.39)	-0.29 (0.86)
Level 2 or 3	-0.39 (0.14)	0.65 (0.25)	0.25 (0.15)	0.43 (0.21)	0.68 (0.52)
Level 4	-0.34 (0.11)	-0.01 (0.21)	-0.02 (0.13)	-0.05 (0.17)	0.38 (0.43)
Level 5	--	--	--	--	--
Moderation by Time*Time*Time					
English Proficiency					
Level 1	0.06 (0.02)	-0.06 (0.03)	0.03 (0.02)	-0.05 (0.03)	0.02 (0.11)
Level 2 or 3	0.03 (0.01)	-0.05 (0.02)	-0.02 (0.01)	-0.03 (0.01)	-0.08 (0.07)
Level 4	0.02 (0.01)	-0.01 (0.02)	0.00 (0.01)	0.00 (0.01)	-0.06 (0.06)
Level 5	--	--	--	--	--

Note: Significance levels are *p< .05, **p< .01, ***p< .001.
Note: Growth models were adjusted for child sex, IEP, ethnicity, race, English proficiency, poverty status, public preschool site, class size, and classroom quality.

TABLE C2. Full Sample HLM Results with Moderation for Language Proficiency Level — Math, Executive Function, and Social Skills

	Math		Executive Function		Social Skills	
	Calculation (1-4)	Applied Problems (PK-4)	Forward Digit Span (1-4)	Backward Digit Span (1-4)	Social Skills (PK-4)	Problem Behaviors (PK-4)
	Est (SE)	Est (SE)	Est (SE)	Est (SE)	Est (SE)	Est (SE)
Ethnicity						
Hispanic	3.02 (2.50)	-0.07 (1.72)	0.04 (0.16)	-0.03(0.1)	3.49 (2.57)	-2.51 (2.25)
Race						
White	2.31 (1.04)	-5.87 (0.76)	0.20 (0.07)	-0.01(0.04)	-1.35 (1.13)	-0.29 (1.03)
Black	5.99 (1.68)	0.68 (1.16)	0.06 (0.11)	0.1(0.07)	0.56 (1.77)	-2.36 (1.58)
Multi-Other						
Moderation by Time						
English Proficiency						
Level 1	3.18 (2.91)	7.47 (1.13)	-0.25 (0.19)	0.03(0.03)	7.01 (1.87)	-0.98 (0.81)
Level 2 or 3	0.86 (1.75)	1.73 (0.67)	0.26 (0.12)	0.02(0.02)	1.01 (1.15)	-0.51 (0.5)
Level 4	-0.32 (1.46)	1.30 (0.57)	0.03 (0.1)	0.01(0.02)	3.02 (0.96)	-0.28 (0.41)
Level 5						
Moderation by Time*Time						
English Proficiency						
Level 1	-1.12 (1.46)	-1.30 (0.30)	0.21 (0.1)		-1.65 (0.51)	0.12 (0.09)
Level 2 or 3	0.35 (0.88)	-0.29 (0.18)	-0.13 (0.06)		-0.24 (0.31)	0.06 (0.05)
Level 4	0.39 (0.73)	-0.22 (0.15)	-0.01 (0.05)		-0.85 (0.26)	0.07 (0.04)
Level 5						
Moderation by Time*Time*Time						
English Proficiency						
Level 1	0.09 (0.19)	0.07 (0.02)	-0.03 (0.01)		0.10 (0.04)	
Level 2 or 3	-0.10 (0.12)	0.02 (0.01)	0.02 (0.01)		0.02 (0.02)	
Level 4	-0.06 (0.10)	0.01 (0.01)	0.00 (0.01)		0.06 (0.02)	
Level 5						

Note: Significance levels are *p< .05, **p< .01, ***p< .001.
Note: Growth models were adjusted for child sex, IEP, ethnicity, race, English proficiency, poverty status, public preschool site, class size, and classroom quality.

TABLE C3. Full Sample HLM Results with Moderation for Race and Ethnicity — Language and Literacy Skills

	Literacy and Language				
	Picture Vocabulary (PK-4)	Sound Awareness (PK-4)	Letter-Word (PK-4)	Word Attack (PK-4)	Passage Comprehension (1-4)
	Est (SE)	Est (SE)	Est (SE)	Est (SE)	Est (SE)
Ethnicity					
Hispanic	-2.48 (1.36)	0.45 (2.49)	-1.09 (1.83)	4.00 (2.73)	2.57 (1.97)
Race					
White	-1.22 (0.60)	-1.66 (1.12)	3.96 (0.88)	0.82 (0.96)	0.99 (0.83)
Black	-0.45 (0.94)	2.70 (1.72)	6.85 (1.37)	5.12 (1.43)	3.74 (1.34)
Multi-Other	--	--	--	--	--
Moderation by Time					
Ethnicity					
Hispanic	-0.80 (0.93)	0.26 (1.71)	1.32 (1.05)	-1.13 (1.85)	-2.88 (1.96)
Race					
White	0.08 (0.38)	-0.53 (0.71)	-0.64 (0.44)	-0.10 (0.62)	-2.24 (0.79)
Black	-0.63 (0.62)	1.17 (1.15)	-0.58 (0.7)	-1.91 (0.94)	-2.32 (1.31)
Multi-Other	--	--	--	--	--
Moderation by Time*Time					
Ethnicity					
Hispanic	0.23 (0.25)	-0.19 (0.45)	-0.26 (0.28)	0.21 (0.42)	0.63 (0.96)
Race					
White	-0.09 (0.10)	0.10 (0.19)	-0.02 (0.11)	-0.20 (0.15)	0.46 (0.39)
Black	0.16 (0.17)	-0.51 (0.31)	-0.13 (0.19)	0.25 (0.23)	0.75 (0.65)
Multi-Other	--	--	--	--	--
Moderation by Time*Time*Time					
Ethnicity					
Hispanic	-0.02 (0.02)	0.02 (0.03)	0.02 (0.02)	-0.01 (0.03)	-0.02 (0.13)
Race					
White	0.00 (0.01)	-0.01 (0.01)	0.00 (0.01)	0.02 (0.01)	-0.04 (0.05)
Black	-0.01 (0.01)	0.04 (0.02)	0.02 (0.01)	-0.01 (0.02)	-0.08 (0.09)
Multi-Other	--	--	--	--	--

Note: Significance levels are *p< .05, **p< .01, ***p< .001.
Note: Growth models were adjusted for child sex, IEP, ethnicity, race, English proficiency, poverty status, public preschool site, class size, and classroom quality.

TABLE C4. Full Sample HLM Results with Moderation for Race and Ethnicity — Math, Executive Function, and Social Skills

	Math		Executive Function		Social Skills	
	Calculation (1-4)	Applied Problems (PK-4)	Forward Digit Span (1-4)	Backward Digit Span (1-4)	Social Skills (PK-4)	Problem Behaviors (PK-4)
	Est (SE)	Est (SE)	Est (SE)	Est (SE)	Est (SE)	Est (SE)
Ethnicity						
Hispanic	3.02 (2.50)	-0.07 (1.72)	0.04 (0.16)	-0.03 (0.1)	3.49 (2.57)	-2.51 (2.25)
Race						
White	2.31 (1.04)	-5.87 (0.76)	0.20 (0.07)	-0.01 (0.04)	-1.35 (1.13)	-0.29 (1.03)
Black	5.99 (1.68)	0.68 (1.16)	0.06 (0.11)	0.10 (0.07)	0.56 (1.77)	-2.36 (1.58)
Multi-Other	--	--	--	--	--	
Moderation by Time						
Ethnicity						
Hispanic	-0.90 (3.3)	0.20 (1.23)	-0.36 (0.22)	0.03 (0.04)	-0.94 (2.04)	-0.68 (0.90)
Race						
White	-1.91 (1.34)	1.71 (0.51)	0.07 (0.09)	-0.01 (0.02)	-0.16 (0.88)	0.22 (0.38)
Black	-1.08 (2.22)	0.23 (0.82)	0.30 (0.15)	0.00 (0.03)	0.81 (1.45)	0.38 (0.63)
Multi-Other	--	--	--	--	--	--
Moderation by Time*Time						
Hispanic	-0.26 (1.63)	-0.15 (0.33)	0.20 (0.11)		0.06 (0.55)	0.01 (0.10)
Race						
White	0.59 (0.67)	-0.35 (0.14)	-0.05 (0.04)		-0.11 (0.24)	0.02 (0.04)
Black	0.62 (1.11)	0.09 (0.22)	-0.15 (0.07)		-0.36 (0.39)	-0.05 (0.07)
Multi-Other	--	--	--	--	--	--
Moderation by Time*Time*Time						
Ethnicity						
Hispanic	0.10 (0.21)	0.02 (0.02)	-0.03 (0.01)		0.01 (0.04)	
Race						
White	-0.07 (0.09)	0.02 (0.01)	0.01 (0.01)		0.01 (0.02)	
Black	-0.09 (0.15)	-0.01 (0.02)	0.02 (0.01)		0.03 (0.03)	
Multi-Other	--	--	--	--	--	--

Note: Significance levels are *p< .05, **p< .01, ***p< .001.
Note: Growth models were adjusted for child sex, IEP, ethnicity, race, English proficiency, poverty status, public preschool site, class size, and classroom quality.

TABLE C5. Annual Gains in Spanish Outcomes Pre-K–4th Grade (DLL Subsample)

	Pre-K Gains		Kindergarten Gains		1st Grade Gains		2nd Grade Gains		3rd Grade Gains		4th Grade Gains	
Spanish Outcomes	Est	(SE)	Est	(SE)	Est	(SE)	Est	(SE)	Est	(SE)	Est	(SE)
Language & Literacy												
Bat-III Picture Vocabulary	-4.41***	(0.46)	-2.71***	(0.27)	-1.01***	(0.26)	0.70	(0.45)	2.0	(0.78)	-43.85	(10.80)
Bat-III Sound Awareness	6.22***	(0.62)	3.69***	(0.35)	1.16***	(0.31)	-1.36	(0.56)	-4.12	(0.93)	60.75	(13.42)
Bat-III Letter-Word Identification	-3.22***	(0.50)	-2.03***	(0.30)	-0.83*	(0.27)	0.36	(0.45)	0.75	(0.85)	-50.01	(12.10)
Bat-III Word Attack	-5.90***	(0.56)	-4.27***	(0.32)	-2.64***	(0.22)	-1.01	(0.41)	-0.28	(0.68)	-67.34	(13.00)
Bat-III Passage Comprehension					-2.36	(0.93)	-2.72*	(0.90)	-0.54	(1.07)	11.07	(11.32)
Math												
Bat-III Calculation	-	-	-	-	0.00	(1.30)	-1.82	(1.28)	0.30	(1.34)	21.83	(14.19)
Bat-III Applied Problems	3.31***	(0.39)	1.08***	(0.20)	-1.15***	(0.18)	-3.38***	(0.35)	-1.52	(0.57)	69.48	(8.45)
Executive Function												
Forward Digit Span	-	-	-	-	0.12	(0.05)	0.03	(0.05)	0.24	(0.07)	-	-
Backward Digit Span	-	-	-	-	0.13	(0.05)	0.12	(0.05)	0.1	(0.02)	-	-

Note: Significance levels are *p< .05, **p< .01, ***p< .001.

Note: Growth models were adjusted for child sex, IEP, ethnicity, race, English proficiency, poverty status, public preschool site, class size, and classroom quality.

Appendix D. Classroom Quality Analyses

TABLE D1. Mean Classroom Assessment Scoring System (CLASS) Scores by Year (Pre-K–1st Grade)

CLASS	Pre-K (n=199)			Kindergarten (n=296)			1st Grade (n=296)		
Domain/ Dimension	Mean	(SD)	Range[a]	Mean	(SD)	Range[a]	Mean	(SD)	Rang
CLASS Total Score	4.6	(0.6)	3.0-6.5	4.3	(0.7)	2.2-5.9	4.2	(0.7)	1.7-6
Emotional Support	5.7	(0.7)	3.6–6.9	5.2	(0.8)	1.7–6.9	5.1	(0.7)	2.7-6
Positive climate	5.9	(0.9)	3.0–7.0	5.3	(1.1)	1.8–7.0	5.1	(1.0)	1.8-7
Negative climate[b]	1.2	(0.4)	1.0–3.4	1.4	(0.7)	1.0–5.8	1.3	(0.5)	1.0-4
Teacher sensitivity	5.5	(1.0)	2.2–7.0	5.1	(1.1)	1.3–7.0	4.9	(1.1)	1.4-7
Regard for student perspectives	4.7	(1.0)	1.6–6.8	3.7	(1.0)	1.5–6.4	3.8	(0.9)	1.0-6
Classroom Organization	5.5	(0.8)	3.1–6.9	5.3	(0.8)	2.7–6.9	5.1	(0.8)	1.3-6
Behavior management	5.8	(1.0)	2.8–7.0	5.6	(1.0)	1.6–7.0	5.4	(1.0)	1.2-7
Productivity	5.9	(0.7)	3.4–7.0	5.6	(0.9)	2.6–7.0	5.3	(0.9)	1.4-7
Instructional learning formats	4.7	(0.9)	2.0–6.8	4.8	(1.0)	2.0–7.0	4.5	(0.9)	1.2-6
Instructional Support	2.6	(0.7)	1.1–5.9	2.5	(0.8)	1.0–4.9	2.4	(0.7)	1.0-5
Concept development	2.5	(0.8)	1.0–5.6	2.4	(0.8)	1.0–4.6	2.2	(0.7)	1.0-5
Quality of feedback	2.6	(0.9)	1.0–6.0	2.6	(1.0)	1.0–5.8	2.5	(0.9)	1.0-5

TABLE D1 (continued). Mean Classroom Assessment Scoring System (CLASS) Scores by Year (2nd–4th Grade)

CLASS	2nd Grade (n=280)			3rd Grade (n=256)			4th Grade (n=228)		
Domain/ Dimension	Mean	(SD)	Range[a]	Mean	(SD)	Range[a]	Mean	(SD)	Range[a]
CLASS Total Score	4.2	(0.7)	2.5-5.9	4.4	(0.7)	2.3-5.8	4.7	(0.6)	2.1-6.1
Emotional Support	5.3	(0.8)	2.9-6.9	5.4	(0.6)	3.1-6.9	4.5	(0.8)	1.7-6.2
Positive climate	5.4	(1.0)	2.4-7.0	5.6	(1.0)	2.2-7.0	5.2	(1.0)	2.0-7.0
Negative climate[b]	1.3	(0.5)	1.0-3.4	1.1	(0.3)	1.0-2.8	1.2	(0.5)	1.0-4.6
Teacher sensitivity	5.2	(1.1)	2.0-7.0	5.5	(0.9)	2.0-7.0	5.5	(1.0)	2.0-7.0
Regard for student perspectives	3.9	(0.9)	1.0-6.8	3.8	(0.8)	1.4-6.8	2.8	(0.9)	1.0-5.2
Classroom Organization	5.2	(0.9)	2.9-6.8	5.5	(0.8)	2.4-6.9	6.3	(0.6)	3.1-7.0
Behavior management	5.5	(1.1)	2.2-7.0	5.9	(0.9)	2.8-7.0	6.1	(0.9)	2.0-7.0
Productivity	5.5	(0.9)	3.0-7.0	5.9	(0.8)	3.0-7.0	6.0	(0.8)	3.4-7.0
Instructional learning formats	4.5	(1.0)	1.2-6.8	4.7	(1.0)	1.4-7.0	4.9	(0.8)	2.0-6.8
Instructional Support	2.2	(0.7)	1.0-4.6	2.3	(0.7)	1.0-4.5	3.2	(0.8)	1.2-5.1
Concept development	2.1	(0.7)	1.0-4.4	2.2	(0.7)	1.0-5.4	NA	-	-
Quality of feedback	2.3	(0.8)	1.0-5.2	2.4	(0.8)	1.0-5.0	2.9	(0.9)	1.0-5.6

[a] Total and Domain scores could range from 1.0–7.0; Dimension scores could range from 1–7.
[b] Scoring is reversed for the Negative climate dimension before it is included in the calculation of the Emotional Support domain score.

TABLE D2. Frequencies of Classroom Assessment Scoring System (CLASS) Scores by Category

	CLASS Total Score	CLASS Domains		
		Emotional Support	Classroom Organization	Instructional Support
Classroom Quality Results: High (score = 6.0-7.0)				
Pre-K	26%	88%	78%	1%
Kindergarten	17%	67%	72%	0%
1st Grade	9%	61%	60%	1%
2nd Grade	13%	72%	64%	0%
3rd Grade	14%	79%	76%	0%
4th Grade	29%	30%	96%	1%
Classroom Quality Results: Mid (score = 3.0-5.9)				
Pre-K	74%	12%	22%	25%
Kindergarten	79%	32%	27%	28%
1st Grade	87%	38%	38%	15%
2nd Grade	80%	28%	36%	14%
3rd Grade	83%	21%	23%	16%
4th Grade	69%	66%	4%	61%
Classroom Quality Results: Low (score = 1.0-2.9)				
Pre-K	1%	0%	0%	74%
Kindergarten	4%	1%	1%	72%
1st Grade	4%	1%	2%	84%
2nd Grade	6%	0%	0%	86%
3rd Grade	2%	0%	0%	84%
4th Grade	1%	4%	0%	38%

TABLE D3. Full Sample HLM Results with Moderation for Classroom Quality — Language and Literacy Skills

	Literacy and Language				
	Picture Vocabulary (PK-4)	Sound Awareness (PK-4)	Letter-Word (PK-4)	Word Attack (PK-4)	Passage Comprehension (1-4)
	Est (S.E.)	Est (S.E.)	Est (S.E.)	Est (S.E.)	Est (S.E.)
Classroom Quality			*	*	
K-4 CLASS High	0.09 (0.44)	1.03 (0.85)	1.74 (0.54)	3.36 (0.78)	1.41 (0.82)
High CLASS*time	-0.27 (0.43)	-0.97 (0.81)	-1.91 (0.5)	-2.70 (0.7)	-1.66 (1.11)
High CLASS*time*time	0.12 (0.12)	0.25 (0.23)	0.50 (0.14)	0.61 (0.18)	0.78 (0.54)
High CLASS*time*time*time	-0.01 (0.01)	-0.02 (0.02)	-0.04 (0.01)	-0.04 (0.01)	-0.10 (0.07)

Note: * = a test of all levels of the variable.
Note: Growth models were adjusted for child sex, IEP, ethnicity, race, English proficiency, poverty status, public preschool site, class size, and classroom quality.

TABLE D4. Full Sample HLM Results with Moderation for Classroom Quality— Math, Executive Function, and Social Skills

	Math		Executive Function		Social Skills	
	Calculations (1-4)	Applied Problems (PK-4)	Forward Digit Span (1-4)	Backward Digit Span (1-4)	Social Skills (PK-4)	Problem Behaviors (PK-4)
	Est (S.E.)	Est (S.E.)	Est (S.E.)	Est (S.E.)	Est (S.E.)	Est (S.E.)
Classroom Quality						*
K-4 CLASS High	2.50 (1.21)	0.74 (0.59)	-0.07 (0.08)	0.06 (0.06)	0.11 (0.99)	2.18 (0.85)
High CLASS*time	-3.57 (1.86)	-0.90 (0.58)	0.06 (0.12)	0.00 (0.02)	0.50 (0.99)	-1.41 (0.45)
High CLASS*time*time	1.56 (0.92)	0.26 (0.16)	-0.02 (0.06)	--	-0.12 (0.28)	0.17 (0.05)
High CLASS*time*time*time	-0.19 (0.12)	-0.02 (0.01)	0.00 (0.01)	--	0.01 (0.02)	--

Note: * = a test of all levels of the variable
Note: Growth models were adjusted for child sex, IEP, ethnicity, race, English proficiency, poverty status, public preschool site, class size, and classroom quality.

Appendix E. Comparison Group Results

TABLE E1. Effect Estimates for Grade 3 and Grade 4 Child Outcomes Comparing Children Enrolled in GAPK With Children Who Had No Pre-K Prior to Kindergarten Entry

		Intercept	Mean Difference (SE) Between GAPK & Comparison	95% Confidence Limits
Language/Literacy	WJ Picture Vocabulary	95.53	***2.65 (0.51)	(1.65, 3.66)
	WJ Sound Awareness	103.55	***6.21 (0.93)	(4.39, 8.04)
	WJ Letter Word	104.26	1.34 (0.64)	(0.08, 2.59)
	WJ Word Attack	102.72	1.36 (0.63)	(0.12, 2.59)
	WJ Passage Comprehension	95.65	***2.15 (0.59)	(0.99, 3.3)
Math	WJ Calculations	101.53	0.49 (0.72)	(-0.92, 1.91)
	WJ Applied Problems	99.047	1.64 (0.72)	(0.23, 3.06)
Executive Function	FDS	4.56	**0.18 (0.05)	(0.08, 0.28)
	BDS	2.74	0.07 (0.05)	(-0.02, 0.16)
Behavior Skills	SSiS Social Skills	104.65	1.25 (1.01)	(-0.73, 3.24)
	SSiS Problematic Behaviors	99.091	-1.14 (0.92)	(-2.95, 0.66)

Note: Significance levels are *p< .05, **p< .01, ***p< .001.
Note: Growth models were adjusted for child sex, IEP, ethnicity, race, English proficiency, poverty status, public preschool site, class size, and classroom quality.

Appendix F. References

[i] Barnett, S.W., Jung, K., Youn, M., & Frede, G.C. (2015). Abbott Preschool Program Longitudinal Effects Study: Fifth Grade Follow-Up. *National Institute for Early Education Research at Rutgers-The State University of New Jersey.* Retrieved October 1, 2020, from www.nieer.org

[ii] Hill, C.J., Gormley, W.T., & Adelstein, S. (2015). Do the short-term effects of a high-quality preschool program persist? *Early Childhood Research Quarterly, 32*, 60-79.

[iii] Lipsey, M.W., Farran, D.C., & Durkin, K. (2018). Effects of the Tennessee Prekindergarten Program on children's achievement and behavior through third grade. *Early Childhood Research Quarterly, 45*, 155-176. https://doi.org/10.1016/j.ecresq.2018.03.005

[iv] Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological), 57*(1), 289-300.

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

[vi] Muñoz-Sandoval, A. F., Woodcock, R. W., McGrew, K. S., & Mather, N. (2005). *Batería III Pruebas de aprovechamiento*. Itasca, IL: Riverside Publishing.

[vii] Gathercole, S. E., & Pickering, S. J. (2000). Working memory deficits in children with low achievements in the national curriculum at 7 years of age. *British Journal of Education Psychology, 70*, 177-194.

[viii] Gresham, F., & Elliott, S. (2008). *Social Skills Improvement System*. Minneapolis, MN: Pearson.

[ix] Pianta, R. C., La Paro, K. M., Hamre, B.K. (2008) *Classroom Assessment Scoring System* (CLASS) Pre-K Manual. Baltimore, MD: Paul H. Brookes Publishing Co.

[x] Pianta, R. C., La Paro, K. M., Hamre, B.K. (2008) *Classroom Assessment Scoring System* (CLASS) K-3 Manual. Baltimore, MD: Paul H. Brookes Publishing Co.

[xi] Pianta, R.C., Hamre, B.K., Mintz, S. (2012) *Classroom Assessment Scoring System* (CLASS) Upper Elementary

[xii] Friedman-Krauss, A. H., Barnett, W. S., Garver, K. A., Hodges, K. S., Weisenfeld, G. G., & Gardiner, B. A. (2021). The State of Preschool 2020: State Preschool Yearbook. *National Institute for Early Education Research*. Available Online: https://nieer.org/wp-content/uploads/2021/08/YB2020_Full_Report_080521.pdf

[xiii] Bleses, D., Makransky, G., Dale, P. S., Højen, A., & Ari, B. A. (2016). Early productive vocabulary predicts academic achievement 10 years later. *Applied Psycholinguistics, 37(6)*, 1461-1476. DOI: 10.1017/S0142716416000060

[xiv] Hammer, C. S., Burchinal, M., Hong, S. S., LaForett, D. R., Páez, M., Buysse, V., ... & López, L. M. (2020). Change in language and literacy knowledge for Spanish–English dual language learners at school-entry: Analyses from three studies. *Early Childhood Research Quarterly, 51*, 81-92. DOI: 10.1016/j.ecresq.2019.07.001

[xv] Ansari, A., & Pianta, R. C. (2018). Variation in the long-term benefits of child care: The role of classroom quality in elementary school. *Developmental Psychology, 54(10),* 1854. DOI: 10.1037/dev0000513