Denver Preschool Program
Child Outcomes Study 2018–2019

Part A

Preschool Progress and Kindergarten Readiness Report, Cohort 11

Document Summary

This report describes short-term outcomes for Cohort 11 of the Denver Preschool Program Child Outcomes Study, focusing on progress made by students over the course of the preschool year and kindergarten readiness at the end of the preschool year. The Cohort is a random, stratified sample comprised of 234 preschool students who participated in DPP during the 2018-2019 school year. The outcomes focus on pre-academic skills (receptive vocabulary, early literacy, and math), social-emotional development, and cognitive skills. Subgroup comparisons by primary language, income level, race/ethnicity, and preschool provider type (Denver Public Schools vs. community sites) are included in addition to details on the relationship between high-quality preschool settings and child outcomes.

Clayton Early Learning Research and Evaluation Department

Executive Summary 2018-2019, Cohort 11

In 2018-2019, 234 preschool students participated in the Denver Preschool Program Child Outcomes Study. The students were directly assessed in the pre-academic domains of vocabulary, early literacy, and math skills in both the fall and spring semesters. Spanish-English dual-language learners completed direct assessments in both English and Spanish. Parents and teachers also completed surveys rating students' social-emotional development and executive function.

Do children make developmental progress while participating in DPP?

On average, preschool students enrolled in DPP developed at or above expectations over the course of the 2018-2019 school year. There were statistically significant increases from fall to spring in students' standard scores on vocabulary assessments administered in English and math assessments administered in Spanish, meaning that students developed in these areas to a greater degree than expected based on their age. For math and early literacy assessments administered in English and vocabulary and early literacy assessments administered in Spanish, students developed at a steady rate that is typical for their age. Students also exhibited statistically significant improvements in teacher ratings of social-emotional development (including self-regulation, initiative, and attachment) over the course of the year.

Do children with different demographic characteristics make similar developmental progress while in DPP?

There were significant differences in student progress by primary language, income level, and race/ethnicity, characteristics that were all strongly associated in this sample. Overall, English-monolingual learners scored higher on English vocabulary, early literacy, and math than did dual-language learners, but dual-language learners showed greater improvements in both vocabulary and math compared to English-monolingual learners. That is, dual-language learners outpaced English-monolingual learners on the vocabulary and math assessments in terms of growth, making developmental progress at a faster rate from fall to spring.

Similar gains in progress were found for students from lower income tiers and Hispanic students, both of whom were over-represented among dual-language learners. Specifically, over 70% of dual-language learners were from the lowest two income tiers, and 87% of dual-language learners were Hispanic. Because of the significant overlap in these categories, it is not possible to fully disentangle the effects of primary language, income, and ethnicity, and the observed

To what extent and in what areas are DPP children ready for kindergarten?

The vast majority of children met the kindergarten readiness benchmark of a standard score of 85 or higher (as defined by test publishers). Specifically, the number of students who demonstrated kindergarten readiness in at least one language on direct assessments of vocabulary, early literacy, and math administered in Spring 2019 ranged from 83% to 92%. Nationally, 84% of students are expected to score in the typical range (85 or above) on these assessments.

The likelihood of demonstrating kindergarten readiness on assessments administered in English was strongly associated with children's primary language, with the proportion of dual-language learners scoring in the typical range being significantly lower than the proportion of English-monolingual learners for all English assessments. For instance, 97% of English-monolingual learners scored 85 or above on English vocabulary compared to 63% of dual-language learners. An even more pronounced pattern of differentiated results emerged when a score of 100 (average score based on national samples) was used as the cutoff. Language development is expected to progress at a different pace for children learning more than one language compared to those learning only one language. When dual-language learners' highest scores in either language for each assessment domain were used, the differences between the proportions of English-monolingual and dual-language learners scoring in the typical range (85 or above) for early literacy and math were eliminated, and a significantly *greater* proportion of dual-language learners scored above average (100 or above) for early literacy.

A large majority of children also exhibited kindergarten readiness in social-emotional development and executive function, as rated by parents and teachers. Similar to the direct assessments, nationally 84% of children are expected to be rated in the "typical" or "no concern" ranges of scores. In Cohort 11, parents and teachers rated 88% to 97% of students as demonstrating kindergarten readiness in the spring in social emotional domains (Total Protective Factors and Behavioral Concerns). For executive function, 87% of children were rated in the adaptive range for Working Memory (e.g., remembering instructions, remembering several things asked to do) by parents compared to 79% rated by teachers. Parents rated 74% and teachers rated 79% of children in the adaptive range for Inhibition (e.g., thinking before acting, being able to stop an activity when asked).

Does the classroom environment make a difference in progress and kindergarten readiness for DPP children?

While it is known from the research literature that classroom quality is essential to promoting positive outcomes for children, it is difficult to examine the associations between classroom quality and child outcomes in this sample because the majority of programs have moderate to high quality ratings. These consistently high ratings provide little variability to statistically explore the relationships between classroom quality and child outcomes. No significant associations were observed between any of the child outcome measures and the Colorado Shines quality ratings of DPP programs. For CLASS® observation ratings, there were a handful of small but positive correlations which point to the positive impact that high quality likely has on child outcomes. Specifically, Emotional Support predicted higher English literacy, Spanish literacy, and Spanish vocabulary scores. Classroom Organization predicted higher English literacy scores and parent ratings of inhibition.

Overall, Denver Preschool Program students in Cohort 11 were ready for kindergarten in pre-academic, social-emotional development, and executive function domains. The majority of dual-language learners demonstrated kindergarten readiness in at least one language and showed accelerated progress over the school year.

Introduction

The Denver Preschool Program (DPP) is a taxpayer-funded initiative, created in 2006 and reauthorized by voters in 2014, that champions, funds, and increases access to high-quality preschools in Denver. DPP supports every family in Denver with a 4-year-old by offering tuition credits to access a high-quality preschool of their choosing and gives approved providers quality improvement resources. The vision of DPP is that children in Denver enter kindergarten ready to reach their full potential.

The Clayton Early Learning Research and Evaluation Department conducted a cross-sequential study, which gathers year-of and longitudinal data for each DPP cohort. The study started during the 2008-2009 school year and has continued every year, uninterrupted. The study helps DPP understand student progress over the course of the preschool year, kindergarten readiness, and the impact of receiving tuition credits on academic outcomes through the end of high school. During the 2018-2019 school year, the research team welcomed Cohort 11 and Cohort 1 was expected to be enrolled in ninth grade (see Table 1).

This report focuses on the short-term outcomes of DPP participation: progress made during the preschool year and kindergarten readiness in the areas of vocabulary, early literacy, math, social-emotional development, and executive functioning. A companion report *Denver Preschool Program Child Outcome Evaluation 2018–2019: Part B: Longitudinal Follow-up, Evaluation Cohorts 1-11 Elementary Report*, describes the long-term outcomes.

Table 1. DPP Evaluation Cohorts and Expected Grade Levels by School Year

	08-09	09-10	10-11	11-12	12-13	13-14	14-15	15-16	16-17	17-18	18-19
Cohort 1	P	K	1st	2nd	3rd	4th	5th	6th	7th	8th	9th
Cohort 2		P	K	1st	2nd	3rd	4th	5th	6th	7th	8th
Cohort 3			P	K	1st	2nd	3rd	4th	5th	6th	7th
Cohort 4				P	K	1st	2nd	3rd	4th	5th	6th
Cohort 5					P	K	1st	2nd	3rd	4th	5th
Cohort 6						P	K	1st	2nd	3rd	4th
Cohort 7							P	K	1st	2nd	3rd
Cohort 8								P	K	1st	2nd
Cohort 9									P	K	1st
Cohort 10										P	K
Cohort 11											P

Short-Term Child Outcome Questions

1.   **Do students make developmental progress while participating in DPP?**

2.   **Do students with different demographic characteristics make similar progress in their development while participating in DPP?**

3.   **To what extent and in what areas are DPP students ready for kindergarten?**

4.   **Does the classroom environment make a difference in progress and kindergarten readiness for DPP students?**

Data Collection and Measurement

The study involved a tremendous amount of coordination, partnerships, and collaboration to collect direct child assessments, parent and teacher surveys, and classroom observations for a sample of children enrolled in DPP in 2018-2019. To ensure that the sample was representative of all students enrolled in DPP, the research team stratified the pool of all enrollees at the beginning of the school year based on student race, income level, primary language, region of Denver, and quality ratings of the preschools students were enrolled in. Students were then randomly selected from each of the strata to ensure that the demographic composition of the sample would reflect the demographic composition of all students enrolled in DPP in 2018-2019. Parents of selected students were contacted by the research team and invited to participate in the study. The researchers also invited lead teachers of the selected students to participate and provide information on the student and the classroom environment. Parents and teachers were compensated for their time with a gift card and received a brief report on the results of their children's assessments.

Direct Child Assessments

The research team used standardized direct assessments to assess receptive vocabulary, early literacy, and math abilities for Standard English and Spanish languages (see Table 2). Because of the rapid growth that happens during the first five years of life and the group atmosphere of the classroom, it is often difficult to know how an individual child is developing based on parent and teacher observations alone. Direct child assessments completed one-on-one with a student by a trained research professional provide valuable insight into a child's pre-academic abilities.

Direct child assessments with students were conducted once in the fall and once in the spring. Highly trained research assistants (assessors) conducted the child assessments. The assessments were typically conducted at the student's school in an area designated by school staff, typically in a quiet space at a child-size table outside the classroom. The direct assessments were spaced about four to five months apart (Cohort 11: M = 4.6; SD = .26). Direct child assessments were conducted in English for all children regardless of their primary language. Spanish-English dual-language learners were also assessed in Spanish in both the fall and spring.

For each assessment, students were given a standard score based on a nationally normed sample. Standard scores are helpful for understanding student progress over the course of the year because these scores are based on age. Standard scores are based on a representative sample for each age and range from 0 to 160, with an average score of 100 and standard deviation of 15. Scores that are steady

Table 2. Direct Child Assessments

Domain	Dimension	Assessment	Abbreviation	Language
Pre-Academic	Vocabulary	Peabody Picture Vocabulary Test-4[1]	PPVT	English
	Early Literacy	Woodcock-Johnson IV Achievement Battery, Letter-Word Identification Subtest[2]	WJ-LWI	
	Math	Woodcock-Johnson IV Achievement Battery, Applied Problems Subtest[2]	WJ-AP	
Pre-Academic	Vocabulary	Batería III Woodcock-Muñoz, Vocabulario[3]	WM	Spanish
	Early Literacy	Batería III Woodcock-Muñoz, Identificación de Letras y Palabras[3]	WM-ILP	
	Math	Batería III Woodcock-Muñoz, Problemas Aplicados[3]	WM_PA	

(i.e., do not change) from fall to spring indicate that the student had a steady rate in development (i.e., developed at an expected rate based on their age). However, if a student learned more than what was expected based on typical development, their scores would increase from fall to spring, indicating accelerated development.

Standard scores are also useful for understanding school readiness in the spring before students enter kindergarten. A student's proficiency level for their age is determined based on the standard deviation from the score of 100. In this study, five categories of readiness expectations were derived from student scores on vocabulary, early literacy, and math: Lagging, Approaching, Meeting, Exceeding, and Excelling (see Table 3). Students with scores in the lagging or approaching categories may need additional supports in kindergarten to "catch up and keep up" with their peers. Students who are meeting, exceeding, or excelling above expectations (based on national averages) are primed for learning and generally need fewer supports to engage in academic instruction.

Table 3. Readiness Categories Indicating Kindergarten Readiness Expectations Level

Readiness Category	Standard Score Range	Standard Deviation
Lagging	0 – 84	< -1
Approaching	85 – 92.4	-.5 to -1
Meeting Expectations	92.5 – 107.4	-. 5 to +.5
Exceeding	107.5 – 114	+.5 to +1
Excelling	115 – 150	> +1

---

[1] Dunn, L. M., & Dunn, D. M. (2007). *PPVT-4: Peabody Picture Vocabulary Test.* London, UK: Pearson Assessments.
[2] Schrank, F. A., McGrew, K. S., Mather, N., Wendling, B. J., & LaForte, E. M. (2014). *Woodcock-Johnson IV Tests of Achievement: Form A.* Itasca, IL: Riverside Publishing.
[3] Muñoz-Sandoval, A. F., Woodcock, R. W., McGrew, K. S., & Mather, N. (2005). *Batería III Woodcock-Muñoz: Pruebas de aprovechamiento.* Itasca, IL: Riverside Publishing.

Parent and Teacher Surveys

Understanding a child's social-emotional development and thinking and memory skills requires in-depth and background knowledge of the child that is hard to capture in a one-time direct child assessment. For these domains, the researchers relied on surveys about students' social-emotional development and executive function (see Table 4) completed by important adults in the child's life. Parents and teachers were asked to fill out surveys during the fall and spring around the same time that the direct child assessments occurred.

The Devereux Early Childhood Assessment (DECA) was completed by parents and teachers and used to measure students' social-emotional development. T-scores were computed for each subscale of the DECA. An average T-score is 50 and has a standard deviation of 10. Similar to standard scores used in the direct child assessments, the interpretation of these T-scores allows for examination of growth over time and comparison with readiness benchmarks. Readiness on this measure is defined as scoring in the "Typical" or "Strength" categories as indicated by the publisher. For Protective Factors (composite of the Initiative, Self-Regulation, and Attachment subscales), children with T-scores greater than 40 fall into the "Typical" and "Strength" categories. For the Behavioral Concerns subscale, higher scores indicate greater levels of behavioral concerns, so children with T-scores below 60 are considered in the "Typical" range. Students who struggle to develop age-appropriate levels of initiative, self-regulation, or attachment in their preschool years are at greater risk of developing mental, emotional, and behavioral disorders or learning difficulties.[4]

Parents and teachers also completed the Childhood Executive Function Inventory (CHEXI; appropriate for ages 4-12) as a measure students' executive function in terms of thinking and

Table 4. Teacher and Parent Surveys

Domain	Questionnaire	Abbreviation	Subscales	Language
Social-emotional development	Devereux Early Childhood Assessment[5]	DECA	Total Protective Factors (Self-Regulation, Attachment, Initiative)	English
			Behavior Concerns	
Executive Function	Child Executive Function Inventory[6]	CHEXI	Working Memory (Working Memory, Planning)	English
			Inhibition (Regulation, Inhibition)	

---

[4] Nesheiwat, K. M., & Brandwein, D. (2011). Factors related to resilience in preschool and kindergarten students. *Child Welfare*, *90*(*1*), 7-24. Retrieved from http://libproxy.lib.unc.edu/login?url=https://search-proquest-com.libproxy.lib.unc.edu/docview/918461956?accountid=14244

[5] LeBuffe, P.A. & Naglieri, J.A. (2012). *Devereux Early Childhood Assessment for Preschoolers, Second Edition.* Lewisville, NC: Kaplan Early Learning Company.

[6] Thorell, L. B., & Nyberg, L. (2008). The childhood executive functioning inventory (CHEXI): A new rating instrument for parents and teachers. *Developmental Neuropsychology, 33(4)*, 536-552.

memory skills. This is a relatively new instrument, and it has been validated with a number of cross-cultural samples.[7] The CHEXI yields four subscale scores (see Table 4) that combine into two executive function factor scores: *Working Memory* and *Inhibition*. Scores range from 13 to 65 on the *Working Memory* factor (e.g., remembering instructions, remembering several things asked to do) and from 11 to 55 on the *Inhibition* factor (e.g., thinking before acting, being able to stop an activity when asked). Lower scores indicate greater levels of executive function (desired levels) in each of these areas. Scores lower than 35 for Working Memory and lower than 33 for Inhibition fall in the adaptive or typical range, and scores that exceed these cutoffs indicate more concerning levels of executive functioning.

Classroom Observations and Preschool Quality Ratings

Highly trained and reliable observers conducted classroom observations using the Pre-K CLASS® tool in classrooms that held a DPP student included in the study. When available, the research team used retrospective Pre-K CLASS® scores (completed in the same school year) for the classroom observation data. The research team also used secondary data from Colorado Shines as a measure of preschool classroom quality (see Table 5).

Table 5. Classroom Quality Observation Tools

Tool	Constructs Measured
Pre-K CLASS®8	Classroom Assessment Scoring System: Emotional Support, Classroom Organization, and Instructional Support
Colorado Shines (Quality Rating Improvement System)	QRIS level that is comprised of examination of learning environment, family partnerships, training and education average ratio, class size, and accreditation

Cohort 11 Snapshot

Cohort 11 was comprised of 234 DPP students who participated in the DPP Child Outcomes Study. Students in the sample averaged 4.5 years of age (range: 4.0 – 5.0) in the fall of the last year before kindergarten, October 2018 and were 48.8% female. Full sample demographics are provided in Appendix A.

To maximize the conclusions that can be drawn about both community and Denver Public Schools (DPS) preschool providers that participate in DPP, the researchers stratified the sample by type of provider and recruited roughly equal numbers of students from each provider type. The result was two subsamples: 116 students attending community preschool programs (49.6% of total sample) and 118 students attending Denver Public Schools (DPS) pre-kindergarten programs (50.4% of total sample).

---

[7] Camerota, M., Willoughby, M. T., Kuhn, L. J., & Blair, C. B. (2018). The Childhood Executive Functioning Inventory (CHEXI): Factor structure, measurement invariance, and correlates in US preschoolers. *Child Neuropsychology*, *24*(*3*), 322–337. https://doi-org.libproxy.lib.unc.edu/10.1080/09297049.2016.1247795

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

Both of these samples were representative of the population of children in each type of preschool at the time of sampling.

The total sample was distributed across 116 DPP preschool providers (59 community sites, 57 DPS sites) and 178 teachers (87 community teachers, 91 DPS teachers). Over half of the sample was enrolled in full-day preschool, about a quarter was enrolled in extended-day preschool, and the remaining students were enrolled in half-day preschool (see Table 6).

Table 6. 2018-2019 Sample by Enrollment Type

Enrollment Type	Percent of Sample	Definition
Half Day	17%	At least 2.5 hours/day and 5 hours/week
Full Day	57%	At least 5 hours/day and 25 hours/week
Extended Day	26%	At least 8 hours/day and 33 hours/week

Data were collected for 233 students in the fall and 232 in the spring. Two students (0.1% of the sample) who were assess in the fall dropped out before the spring data collection because they were no longer attending DPP preschools. Data were collected for the 231 students returning from the fall, and one additional student was recruited in the spring to make up for this attrition (see Table 7). Direct assessments were conducted in English with all sample students, and Spanish-speaking students (n = 45) were also assessed in Spanish. Teacher and parent reports were obtained for 69% and 80% of students, respectively, in the fall, and for 71% and 78% of students, respectively, in the spring. Classroom observations were obtained for 70% of all students.

Table 7. Sample Sizes by Data Collection Activity and Period

Data Collection Activity	Fall 2018	Spring 2019
Direct Child Assessments - English	233	232
Direct Child Assessments - Spanish	45	45
Teacher Reports	161	164
Parent Reports	187	182
Classroom Observations[1]	163 (70% of all students, *n* = 234)	

[1]This figure represents the number of children for whom classroom observations were obtained.

Because the percentages of students enrolled in community versus DPS preschools in the full population of DPP students in 2018-2019 were not equal (37% at community sites vs. 63% at DPS sites[9]), sampling weights were applied to all analyses on the sample as a whole so that the results would be representative of the population of students enrolled in DPP over the course of the full school year. For analyses comparing DPS and community preschools, weights were not applied.

Preschool Progress – Full Cohort 11 Sample

Do students make progress in their development while participating in DPP (i.e., vocabulary, early literacy, mathematics, social-emotional development, and executive function)?

---

[9] Final 2018-2019 enrollment numbers obtained from MetrixIQ.

Students' mean scores on each of the assessments and survey reports from the fall and the spring are reported in Tables 8 and 9. Paired *t*-tests were conducted to examine the progress that DPP students made over the course of their preschool year in vocabulary, early literacy, and math skills (in both English and Spanish) and in social-emotional development and executive function (reported by both teachers and parents).

Table 8. 2018-2019 Weighted Descriptive Statistics for the Directly Assessed Child Outcome Measures

English Direct Child Assessments[1]	Fall 2018				Spring 2019			
	n	Mean	SD	Range	*n*	Mean	SD	Range
Vocabulary	233	105.56	20.78	48-146	232	108.34	18.41	47-145
Early Literacy	231	93.80	13.27	62-145	231	93.61	14.10	57-144
Math	228	99.61	16.04	47-136	229	100.15	14.83	43-134
Spanish Direct Child Assessments[2]	*n*	Mean	SD	Range	*n*	Mean	SD	Range
Vocabulary	44	81.51	18.70	32-105	44	79.81	20.12	24-110
Early Literacy	45	101.08	12.70	80-132	45	101.43	17.38	75-152
Math	44	95.37	11.93	63-111	45	97.67	11.70	62-129

[1]All students were assessed in English, regardless of primary language.
[2]Children additionally learning Spanish (94% of the sample's dual-language learners) were assessed in Spanish in addition to English.

Table 9. 2018-2019 Weighted Descriptive Statistics for Parent and Teacher Surveys[10]

Social-Emotional Development	Fall 2018				Spring 2019			
Parent Ratings	*n*	Mean	SD	Range	*n*	Mean	SD	Range
Total Protective Factors	187	51.97	9.13	28-72	182	51.61	9.20	28-72
Initiative	187	52.23	9.39	28-68	182	51.97	8.95	29-68
Self-Regulation	187	51.47	8.64	28-72	182	52.19	8.69	28-72
Attachment	187	51.56	8.42	28-66	180	50.77	9.03	28-66
Behavioral Concerns	187	49.13	8.47	28-69	182	48.01	8.56	28-69
Teacher Ratings	*n*	Mean	SD	Range	*n*	Mean	SD	Range
Total Protective Factors	161	53.31	8.44	30-72	164	55.31	7.93	28-72
Initiative	161	52.72	9.33	29-72	164	54.06	8.77	29-72
Self-Regulation	160	53.89	8.38	31-69	164	55.05	8.63	28-70
Attachment	160	52.27	7.48	32-71	164	54.31	7.60	33-71
Behavioral Concerns	160	45.53	8.95	29-68	164	44.92	8.81	29-72
Executive Function	Fall 2018				Spring 2019			
Parent Ratings	*n*	Mean	SD	Range	*n*	Mean	SD	Range

[10] Some teachers and parents left items blank on the DECA and CHEXI questionnaires. Scores were calculated in at least 75% of the items were present. This resulted in some missing data for the Parent and Teacher Ratings.

Working Memory	186	26.96	7.46	13-60	181	26.11	7.09	13-51
Inhibition	184	28.69	6.20	12-52	181	27.87	6.36	11-46
Teacher Ratings	***n***	**Mean**	**SD**	**Range**	***n***	**Mean**	**SD**	**Range**
Working Memory	160	28.17	9.55	13-62	164	27.20	10.18	13-63
Inhibition	159	25.63	7.90	11-50	164	25.56	8.79	11-54

Vocabulary, early literacy, and math progress. Students had steady standard scores across the year in English early literacy and math (see Table 10). Because standard scores are adjusted for age, this indicates that students developed at a rate that is typical for their age. A statistically significant increase of almost four points (small effect size[11]) was found for English vocabulary scores. This suggests that DPP students learned English vocabulary at a faster rate from fall to spring than what is expected with typical maturation.

Analyses of Spanish assessments showed steady scores for vocabulary and early literacy, and a statistically significant increase in math scores (see Table 10). This means that Spanish-English dual-language learners learned vocabulary and early literacy in Spanish as expected based on typical development, and their math skills in Spanish developed at a faster rate than expected with an increase of over 2 points (small effect size).

Table 10. 2018-2019 Preschool Gains in Vocabulary, Early Literacy, and Math Standard Scores (Weighted)

English Direct Child Assessments	***n***	**Mean Diff (SD)**	***t***	**Cohen's D**
Vocabulary	231	2.70 (10.40)	**3.98****	**0.26**
Early Literacy	229	-0.16 (7.11)	-0.34	0.02
Math	226	0.93 (10.01)	1.51	0.09
Spanish Direct Child Assessments	***n***	**Mean Diff (SD)**	***t***	**Cohen's D**
Vocabulary	43	-1.88 (9.67)	-1.35	0.20
Early Literacy	45	0.35 (11.14)	0.22	0.03
Math	44	2.62 (8.16)	**2.24***	**0.32**

*$p < .05$, **$p < .001$

Social-emotional development and executive function progress. When rated by parents, no statistically significant changes were observed for students' Total Protective Factors or Behavioral Concerns (see Table 11). A small but statistically significant improvement from fall to spring was observed for teachers' ratings of Total Protective Factors, and no significant changes were observed on teachers' ratings of Behavioral Concerns.

No significant changes were observed from fall to spring on parent and teacher ratings of working memory and inhibition, both factors of executive function (see Table 12).

---

[11] Effect size guidelines: 0.20 – 0.40 small effect; 0.41 – 0.70 medium effect; > 0.70 large effect.

Table 11. 2018-2019 Preschool Gains in Social-emotional development (Weighted)

Parent Ratings	n	Mean Diff (SD)	t	Cohen's D
Total Protective Factors (t-score)	147	-0.20 (7.42)	-0.33	0.03
Behavioral Concerns (t-score)	147	-0.85 (9.11)	-1.29	0.10
Teacher Ratings	***n***	**Mean Diff (SD)**	***t***	**Cohen's D**
Total Protective Factors (t-score)	134	1.28 (6.86)	**2.24***	**0.19**
Behavioral Concerns (t-score)	133	0.20 (7.44)	0.32	0.03

*p < .05

Table 12. 2018-2019 Preschool Gains in Executive Function (Weighted)

Parent Ratings	n	Mean Diff (SD)	t	Cohen's D
Working Memory Factor (t-score)	145	-0.29 (6.51)	-0.54	0.04
Inhibition Factor (t-score)	144	-0.58 (5.72)	-1.24	0.10
Teacher Ratings	***n***	**Mean Diff (SD)**	***t***	**Cohen's D**
Working Memory Factor (t-score)	133	-0.13 (7.29)	-0.21	0.02
Inhibition Factor (t-score)	132	0.58 (6.36)	1.09	0.09

Preschool Progress – Differences by Primary Language

Because all students were assessed in English, regardless of their primary language, it is useful to examine how students' scores on the English assessments differed based on whether students spoke English as their primary language or whether they were dual-language learners. In both the fall and spring, there were statistically significant differences on all English assessment scores between English-monolingual learners and dual-language learners (students' whose primary language was any other language than English or a combination of languages). As shown in Table 13, dual-language learners consistently scored lower on the English assessments than English-monolingual students.

Table 13. Weighted English Direct Child Assessment Scores by Student Primary Language and Data Collection Period

|	English-Monolingual		Dual-Language			
Fall 2018	***n***	**Mean (SD)**	***n***	**Mean (SD)**	**Mean Diff**	***t***
Vocabulary	184	112.49 (14.69)	48	81.85 (21.38)	30.63	**9.75***
Early Literacy	182	96.09 (12.75)	48	86.26 (12.13)	9.84	**4.98***
Math	182	103.09 (13.77)	45	87.80 (16.91)	15.29	**6.60***
Spring 2019	***n***	**Mean (SD)**	***n***	**Mean (SD)**	**Mean Diff**	***t***
Vocabulary	184	113.63 (14.04)	47	89.99 (20.15)	23.64	**7.89***
Early Literacy	183	95.87 (13.45)	47	85.96 (13.60)	9.91	**4.65***
Math	183	102.90 (90.96)	45	90.96 (14.15)	11.93	**5.44***

**p < .001

Although the gap in English assessment scores between dual-language and English-monolingual learners persisted from fall to spring, as it has in past years, there could be differences in the amount of progress that students make over the course of the year. Students who begin the year scoring lower than average on assessments have more ground to gain by the end of the year.

Analyses were conducted with primary language predicting scores over time on assessments administered in English. There were significant interactions between primary language and time on vocabulary and math scores.[12] Although dual-language learners scored lower than English-monolingual learners in both the fall and spring, the spring vocabulary and math scores of dual-language learners were significantly greater than their fall scores (see Charts 1 & 2).[13] The vocabulary and math scores of English-monolingual learners did not change significantly from fall to spring.[14] In other words, dual-language learners made progress in English vocabulary and math at a faster rate than expected based on typical development, whereas English-monolingual learners progressed at an expected rate based on their age. The interaction between primary language and time on early literacy scores was not significant.[15]

In interpreting these findings, it is important to note that primary language is strongly associated with other student characteristics, namely income tier and ethnicity. In this year's sample, over 70% of dual-language learners were from the lowest two income tiers compared to less than 30% of English-monolingual learners (see Chart 3).[16] Additionally, the majority of dual-language learners were Hispanic (87%), and the majority of Hispanic students were in the lowest two income tiers (67%). Because of the significant overlap in these categories, it is impossible to completely disentangle the effects of income, primary language, and ethnicity, and any effects observed are possibly the result of the combination of these factors.

Chart 1. Change in Receptive Vocabulary Scores over Time by Student Primary Language (Assessed in English).[17]

---
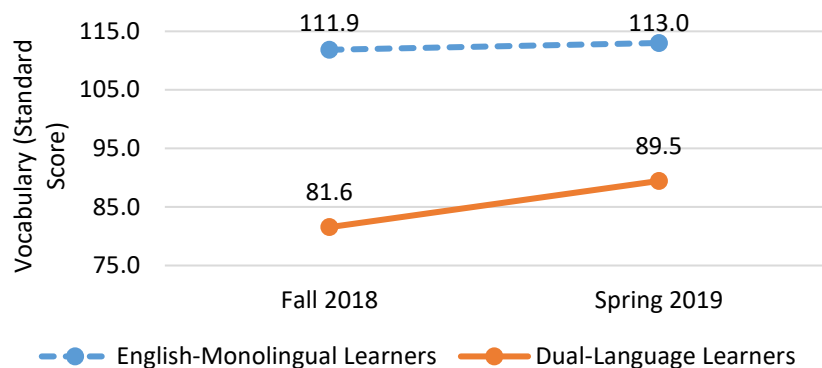
[12] Vocabulary: $F(1, 228) = 16.64$, $p < .001$; Math: $F(1, 223) = 5.22$, $p < .05$
[13] Vocabulary: $F(1, 228) = 28.67$, $p < .001$; Math: $F(1, 223) = 7.18$, $p < .01$
[14] Vocabulary: $F(1, 228) = 2.38$, *n.s.* ; Math: $F(1, 223) = 0.07$, *n.s.*
[15] Early Literacy: $F(1, 226) = 1.17$, *n.s.*
[16] $\chi^2(4) = 35.66$, $p < .001$
[17] Solid lines indicate significant change over time (p < .05). Dashed lines indicate no significant change.

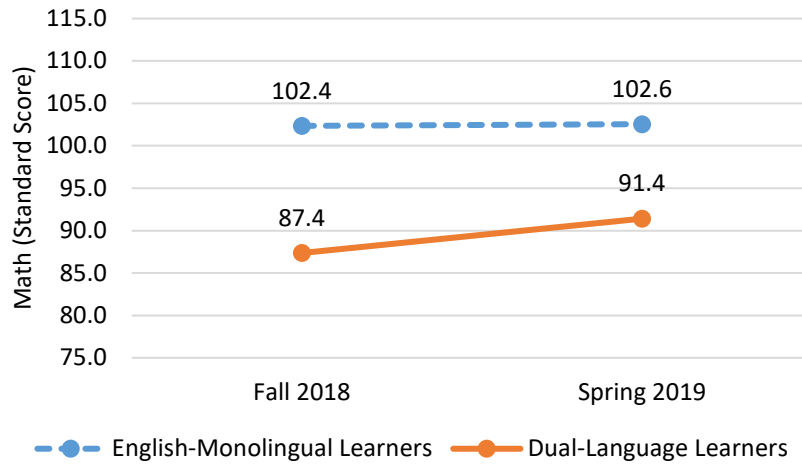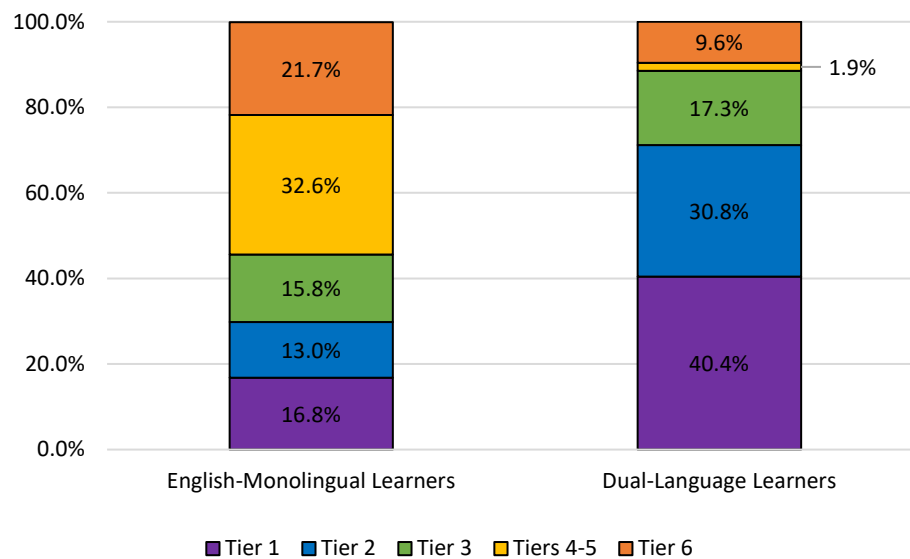*Chart 2. Change in Math Scores over Time by Student Primary Language (Assessed in English).[16]*

Chart 3. Income tier groups by student primary language.

Preschool Progress – Other Group Comparisons

Do students with different demographic characteristics make similar progress in their development while participating in DPP?

Group comparisons of student progress on standardized assessment scores provide insight into how participation in DPP may be more or less effective for different students. In addition to primary language, group comparisons related to income tier, race/ethnicity, gender, and provider type (DPS vs. community) were examined. No significant differences by gender were found on direct child assessments in English or Spanish. Differences by income tier, race/ethnicity, and provider type are reported below.

Change over time by income tier. The percentages of Cohort 11 children represented in each income tier are reported in Table 14. Because there were only 12 students in Tier 4, Tiers 4 and 5 were collapsed into a single income tier group prior to analyses. Parents in Tier 6 opted out of reporting income and were automatically assigned the lowest tuition credit amount.

Comparisons by income tier revealed that students in lower income tiers made significantly greater gains in English vocabulary scores compared to students in higher income tiers.[18] Although the scores of

students in the higher tiers in both the fall and Spring, the scores of students in Tiers 1, 2, and 3 increased significantly from fall to spring (see Chart 4).[19] Students in Tier 2 made the largest gains with an increase of 6.2 points.[20] The scores of students Tiers 4, 5, and 6 did not increase significantly from fall to spring.[21] In other words, students in Tiers 1-3 made progress in English vocabulary from fall to spring at a faster rate than expected based on typical development, whereas students in the higher tiers progressed at an expected rate based on their age.

Table 14. Percentage of Sample by Income Tier

Income Tier (Federal Poverty Threshold)	Percentage of Sample (Weighted)
Tier 1 (<100%)	22%
Tier 2 (100%-185%)	17%
Tier 3 (185%-300%)	16%
Tier 4 (300%-400%)	5%
Tier 5 (>400%)	21%
Tier 6 (Opt out)	19%

The statistically significant gains in vocabulary by students in Tiers 1, 2, and 3 are consistent with the significant progress made by dual-language learners, 89% of whom are in the lowest 3 tiers. No significant differences in progress by income tier were found for early literacy assessed in English or Spanish and vocabulary and math assessed in Spanish.[22]

---

[18] Income tier x Time interaction: $F_{(4, 226)} = 2.55$, $p < .05$

[19] Tier 1 Fall 2018 vs. Spring 2019: $F_{(1, 226)} = 4.53$, $p < .05$; Tier 2 Fall 2018 vs. Spring 2019: $F_{(1, 226)} = 14.29$, $p < .001$; Tier 3 Fall 2018 vs. Spring 2019: $F_{(1, 226)} = 4.41$, $p < .05$

[20] $\eta_p^2 = .22$, small effect size

[21] Tiers 4-5 Fall 2018 vs. Spring 2019: $F_{(1, 226)} = 0.58$, *n.s.*; Tier 6 Fall 2018 vs. Spring 2019: $F_{(1, 226)} = 0.06$, *n.s.*

[22] For analyses of assessments administered in Spanish, a two-level income tier group was used comparing students in Tier 1 with students in Tiers 2 and 3. The Tiers 4-5 and Tier 6 categories were omitted because there were no children in Tiers 4 and 5 who were assessed in Spanish and only 3 children assessed in Spanish in Tier 6. Tiers 2 and 3 were combined into a single category because there were only 7 children assessed in Spanish in Tier 3.

Chart 4. Change in Vocabulary Scores (Assessed in English) over Time by Income Tier.[23]
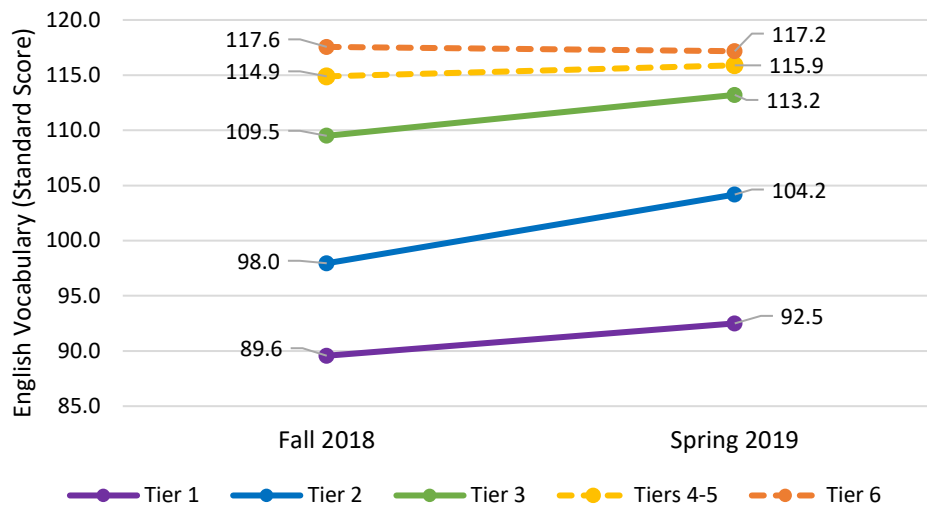
Change over time by race/ethnicity. Change in English vocabulary, early literacy, and math scores were also examined across different race/ethnicity groups.[24] Groups with fewer than 20 students (Asian or Pacific Islander, Multi-Racial, Other) were combined into a single group prior to analyses. Race/ethnicity interactions with time revealed that the amount of progress students made on vocabulary and math assessed in English varied by race/ethnicity.[25] Specifically, Hispanic students made significantly greater progress on vocabulary and math assessed in English than expected based on typical development (see Charts 5 & 6).[26] Students belonging to other race/ethnicity groups demonstrated steady vocabulary and math scores, indicating that they progressed at an expected rate based on their age. In other words, Hispanic students, who had the lowest scores on vocabulary and math assessed in English in the fall, were "catching up" or learning at a faster rate than their peers from other race/ethnicity groups over the course of the school year. No differences by race/ethnicity group were found for student progress on early literacy assessments in English.[27]

---

[23] Solid lines indicate significant change over time (p < .05). Dashed lines indicate no significant change.

[24] Race/ethnicity differences in progress on Spanish vocabulary, early literacy, and math were not examined because 92% of students who completed Spanish assessments were Hispanic.

[25] Race/ethnicity x Time interaction on English vocabulary: $F(3, 227) = 5.62$, $p < .01$; Race/ethnicity x Time interaction on English math: $F(3, 222) = 2.68$, $p < .07$.

[26] Hispanic students Fall 2018 vs. Spring 2019 vocabulary: $F(1, 227) = 28.54$, $p < .001$; Hispanic students Fall 2018 vs. Spring 2019 math: $F(1, 222) = 7.80$, $p < .01$

[27] Race/ethnicity x Time interaction on English early literacy: $F(3, 225) = 0.63$, *n.s.*

Chart 5. Change in Vocabulary Scores (Assessed in English) over Time by Race/Ethnicity.[28]
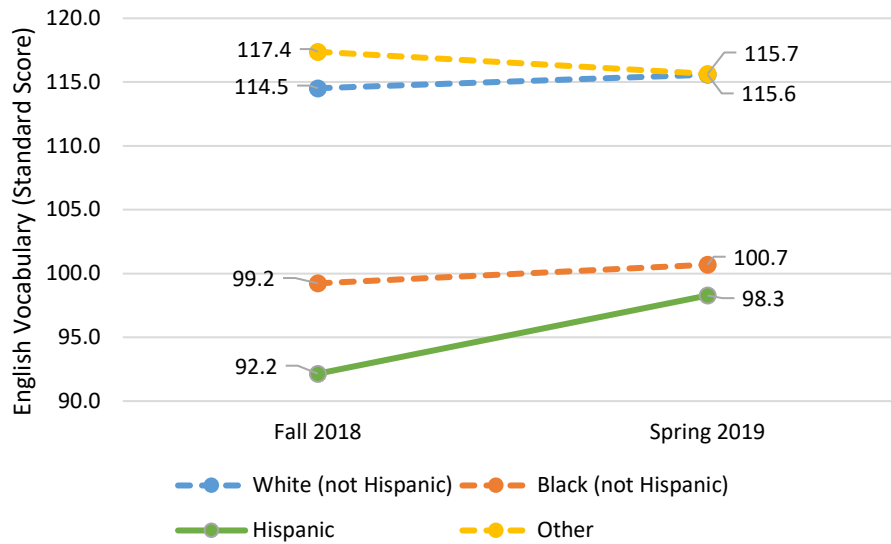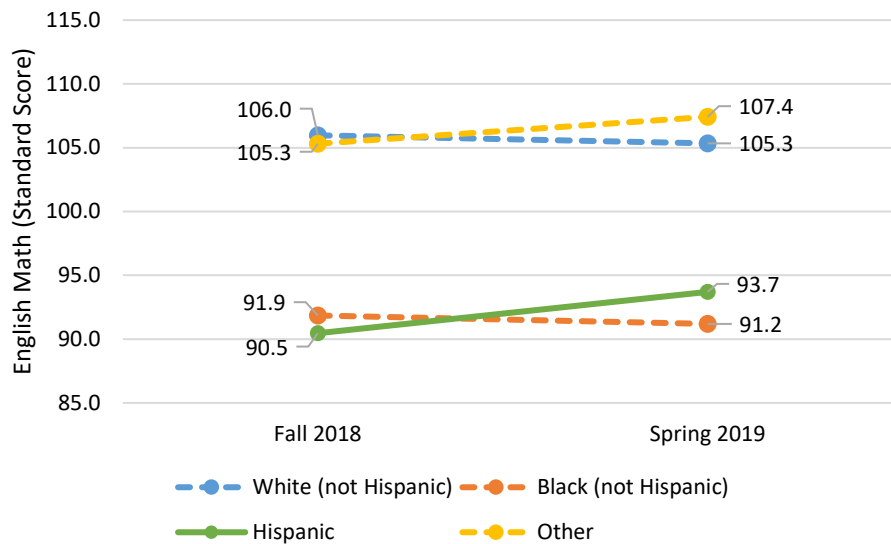

Chart 6. Change in Math Scores (Assessed in English) over Time by Race/Ethnicity.

The significant gains in vocabulary and math among Hispanic students mirror the pattern of significant gains made by dual-language learners. This is to be expected, given that the majority of dual-language learners were Hispanic. However, almost half (48%) of all Hispanic students in the sample were English-monolingual learners. When comparisons were made between progress made on vocabulary and math assessments by English-monolingual and dual-language Hispanic students, the significant progress effects persisted for both groups. In other words, Hispanic students made statistically significant gains in

---

[28] Solid lines indicate significant change over time ($p < .05$). Dashed lines indicate no significant change.

progress whether or not they were dual-language learners.[29] These findings underscore the fact that the effects of primary language, income, and race/ethnicity cannot be disentangled and that observed differences may be a result of a combination of these factors.

Change over time by provider type. Comparisons of students at DPS versus community sites revealed no differences in progress on vocabulary and math (in English or Spanish) and Spanish early literacy.[30] For early literacy assessed in English, there was a significant difference in the progress students made at DPS versus community sites.[31] Specifically, students at DPS sites showed a marginally significant increase over time,[32] whereas students at community sites showed a significant decrease over time.[33]

Cohort 11 Kindergarten Readiness

To what extent and in what areas are DPP students ready for kindergarten?

Analyses were conducted to determine how ready for kindergarten DPP students in Cohort 11 appeared to be at the end of their last preschool year before kindergarten. Readiness was examined in several ways. Analyses were first conducted to examine whether students scored in the typical range as defined by the test publishers, namely a standard score of 85 or above. A standard score below 85 indicates being in the risk range for the assessment. While not being at risk when entering kindergarten is important, it is also useful to examine whether students meet a higher standard of kindergarten readiness. Thus, additional analyses were conducted to examine whether students scored at or above 100, the national population mean, on each direct assessment. In the general population, about 84% of preschool students are expected to score 85 or above and 50% of students are expected to score 100 or above. Chart 7 presents the percentages of all students scoring 85 or above and the percentages of all students scoring 100 or above on each of the assessments in English and in Spanish administered in the spring.

---

[29] Primary language x Time interactions were non-significant for Hispanic students' English vocabulary and math scores, $F(1, 76) = 3.01$, *n.s.*, and $F(1, 72) = 0.35$, *n.s.*

[30] Provider type (DPS vs. Community) x Time (Fall 2018 vs. Spring 2019) interactions were non-significant for English vocabulary, $F(1, 229) = 1.27$, *n.s.*, English math, $F(1, 224) = 0.02$, *n.s.*, Spanish vocabulary, $F(1, 41) = 0.01$, *n.s.*, Spanish early literacy, $F(1, 43) = 3.85$, *n.s.*, and Spanish math, $F(1, 42) = 0.33$, *n.s.*

[31] $F(1, 227) = 14.69$, $p < .001$

[32] $F(1, 227) = 3.03$, $p < .10$, Cohen's $d = .08$ (small effect)

[33] $F(1, 227) = 13.51$, $p < .001$, Cohen's $d = .20$ (small effect)

Chart 7. Weighted Percentages of Students Scoring in the Typical and Above Average Ranges on All Spring Direct Standardized Assessments, 2018-2019
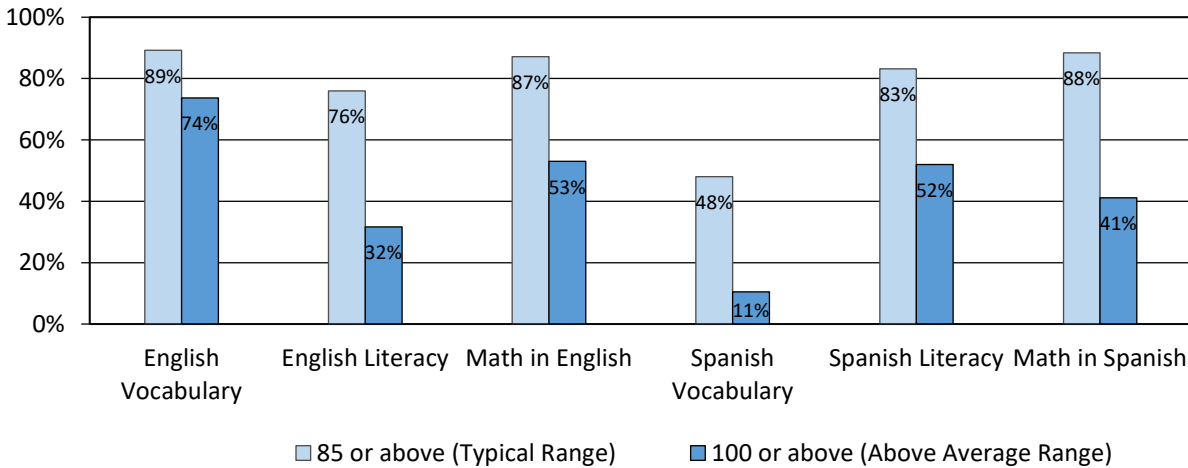
Further analyses were conducted to examine the proportion of students reaching each readiness benchmark by primary language. Using the same national standards on the English-administered assessments, Charts 8 and 9 take the first three bar sets in Chart 7 and break them down by whether students' primary language was English only or whether they were dual-language learners. Chart 8 shows the percentages of students scoring 85 or above on each English assessment by primary language, and Chart 9 shows the percentages of students scoring 100 or above by primary language. Students in the dual language learner group included those additionally learning Spanish, Arabic, Tamil and other languages.

Dual-language learners typically have a different developmental trajectory than children who are learning only one language.[34] Thus, it was not surprising that the likelihood of scoring 85 or above on the spring assessments was strongly associated with students' primary language in this sample. The proportions of English-monolingual learners who scored 85 or higher on each English assessment were significantly greater than the proportions of dual-language learners who scored 85 or higher (see Chart 8).[35] An even more pronounced pattern of differentiated results emerged when a score of 100 was used as the cutoff, with many fewer dual-language learners scoring 100 or above compared to English-monolingual learners (see Chart 9).[36]

---

[34] Kalia, V., Daneri, M.P., Wilbourn, M.P. (2019). Relations between vocabulary and executive functions in Spanish-English dual-language learners. *Bilingualism: Language and Cognition*, *22*(*1*), 1-14. https://doi-org.libproxy.lib.unc.edu/ 10.1017/S1366728917000463

[35] English Vocabulary: $\chi^2(1) = 49.88$, $p < .001$; English Literacy: $\chi^2(1) = 15.61$, $p < .001$; Math in English: $\chi^2(1) = 8.43$, $p < .05$

[36] English Vocabulary: $\chi^2(1) = 62.29$, $p < .001$; English Literacy: $\chi^2(1) = 8.41$, $p < .05$; Math in English: $\chi^2(1) = 20.25$, $p < .001$

*Chart 8. Weighted Percentages of Students Scoring in the Typical Range (**85 or Above**) on Spring English Assessments by Primary Language, 2018-2019*
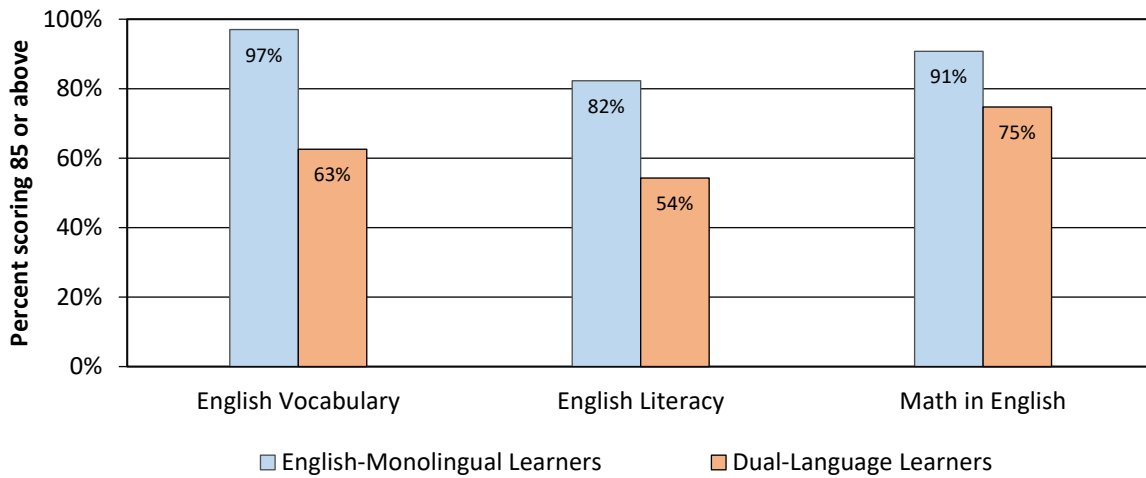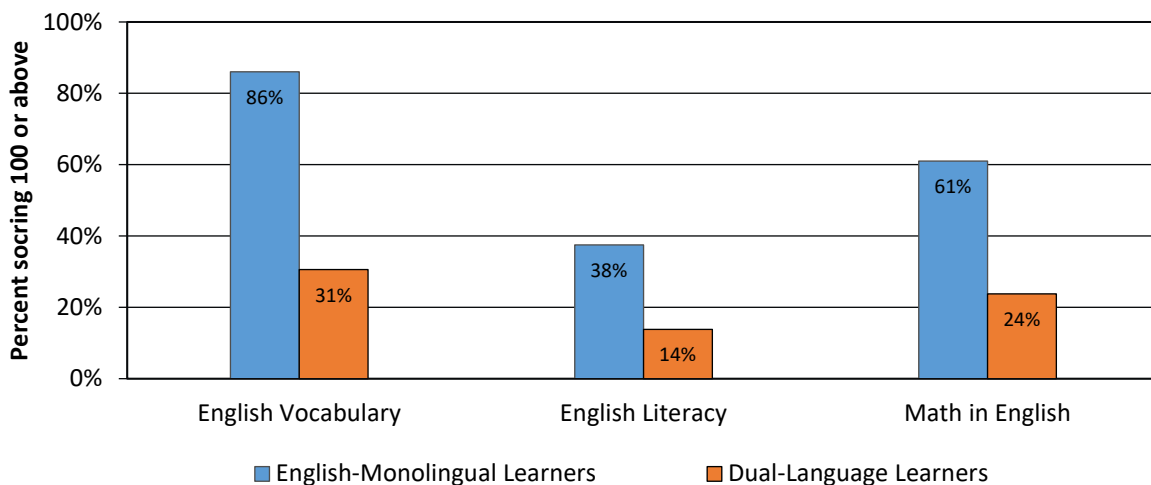

*Chart 9. Weighted Percentages of Students Scoring in the Above Average Range (**100 or Above**) on Spring English Assessments by Primary Language**, 2018-2019*

Kindergarten Readiness on Standardized Assessments in Either Language

Language development for children learning two or more languages is expected to progress at a different pace than for children learning one language. All of the assessments used in this study were normed with monolingual children who were learning only one language, and thus do not take into account typical development for dual-language learners. One way to address this limitation and better understand the progress of dual-language learners is to examine dual-language learners' scores in both languages.

In Charts 10 and 11, *students' highest score in either English or Spanish was used as the measure of kindergarten readiness* for each assessment domain (vocabulary, early literacy, and math).

Students who were Spanish-English dual-language learners contributed their highest score on either the Spanish or English assessment. Students who were assessed in English only had only one score for each domain to use for their highest score.

Using students' highest score in either language, the differences in proportions of students reaching each readiness benchmark by primary language were markedly smaller (see Chart 10). On early literacy and math assessments, 86% and 89% of dual-language learners, respectively, scored 85 or above, compared to 54% and 75% when only scores in English were considered. Furthermore, ***the differences between English-monolingual and dual-language learners for early literacy and math were eliminated when students' highest scores in either language were used.***[37]

The percentage of dual-language learners scoring 85 or above in vocabulary using their highest score in either language (76%) was also higher compared to using English scores only (67%), but was still significantly lower than the percentage of English-monolingual learners scoring 85 or above on vocabulary (97%).[38]

Greater percentages of dual-language learners also scored 100 or higher in each assessment domain when their highest score in either language was taken into account compared to using their English scores only (see Chart 11). Using their highest score in either language, the percentages of dual-language learners scoring 100 or higher on vocabulary and math assessments (39% and 46%, respectively) were still lower than the percentages of English-monolingual learners scoring 100 or higher (86% and 61%).[39] However, the percentage of dual-language learners who scored 100 of higher on early literacy in at least one language (54%) was significantly *greater* than the percentage of English-monolingual learners who scored 100 or higher on early literacy assessed in English (37%).[40]

*Chart 10. Weighted Percentages of Students Scoring in the Typical Range (**85 or Above**) on Spring Assessments Using Highest Score in English or Spanish, 2018-2019*
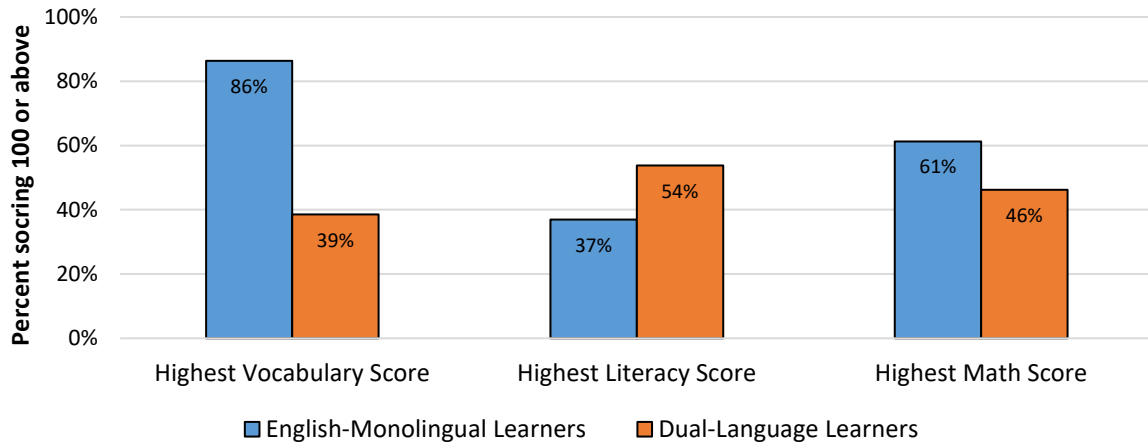

---
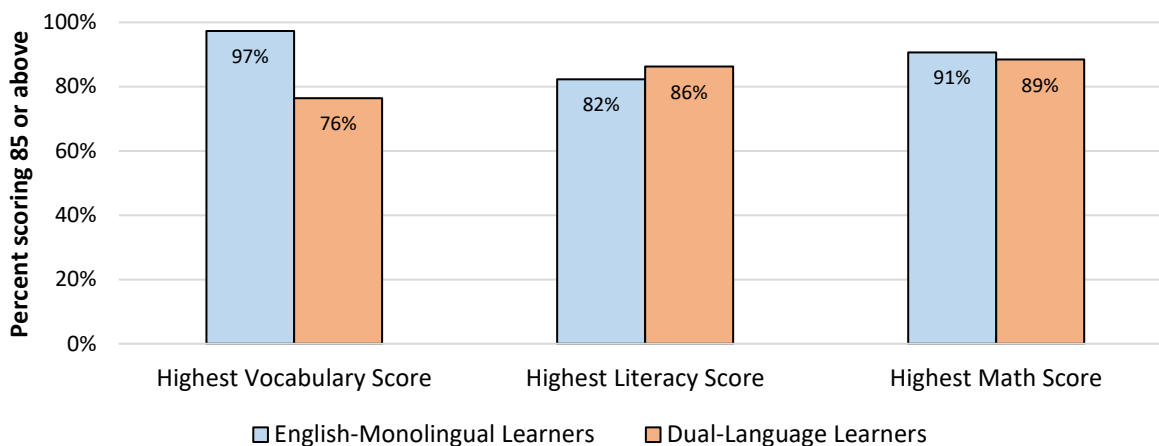
[37] Highest Literacy Score: $\chi^2(1) = 0.45$, *n.s.*; Highest Math Score: $\chi^2(1) = 0.22$, *n.s.*
[38] $\chi^2(1) = 25.44$, *p* < .001
[39] Highest Vocabulary Score: $\chi^2_1 = 50.16$, *p* <.001; Highest Math Score: $\chi^2_1 = 3.82$, *p* < .06
[40] $\chi^2_1 = 4.74$, *p* < .05

*Chart 11. Weighted Percentages of Students Scoring in the Average Range (**100 or Above**) on Spring Assessments Using Highest Score in English or Spanish, 2018-2019*

Understanding the percentages of students reaching readiness benchmarks defined by test publishers and based on national averages provides important information about how ready for kindergarten DPP students are in the broader national context. Beyond examining the proportions of students who fall on either side of these cutoffs, it is also useful to look at the full spectrum of how DPP students score in each assessment domain. Using students' highest score in either language, the proportions of students fall into each of the five readiness categories (described earlier) were examined. Taking into account dual-language learners highest score in either language, 85% of Cohort 11 students met or exceeded expectations in vocabulary in at least one language of administration. In early literacy, 61% of students met or exceeded expectations, and 79% of students met or exceeded expectations in math (see Chart 12).

Chart 12. Highest Score in Either English or Spanish by Spring Proficiency Level, 2018-2019 (Cohort 11)
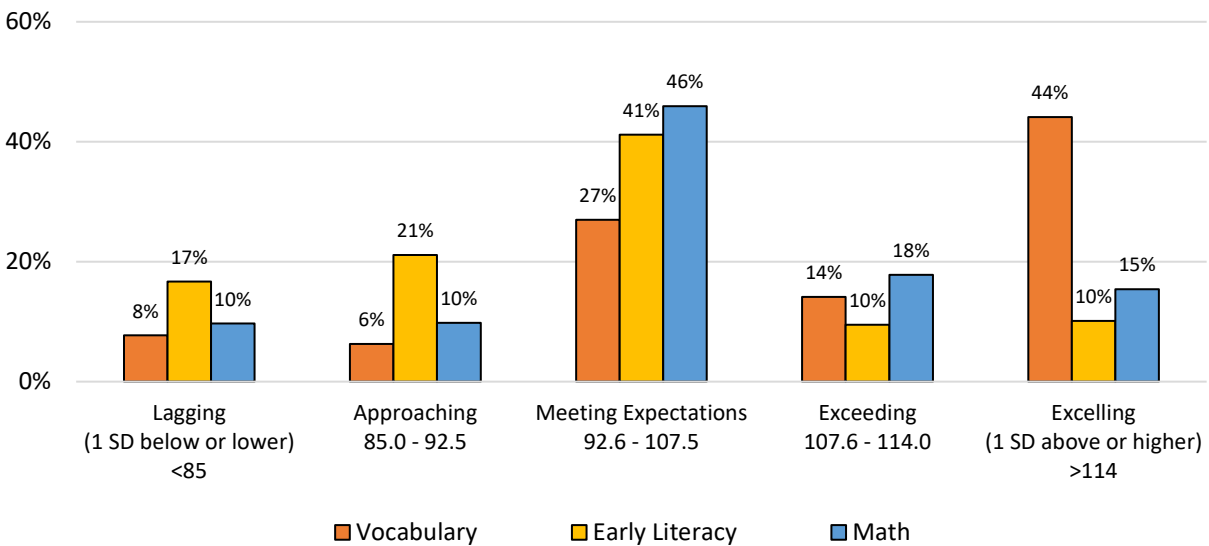
All together these benchmarks and measures provide an overview of students' school readiness at different levels. As discussed in the data collection and measurement section of this report, these categories have practical implications in terms of students' kindergarten readiness and relative needs for support. Students with scores in lagging or approaching categories may need additional support in kindergarten to "catch up and keep up" with their peers who scored in the higher categories. Students who are meeting, exceeding, or excelling above expectations are primed for learning and generally need fewer supports to engage in academic instruction.

Kindergarten Readiness - Social-emotional development & Executive Function

Analyses were conducted to examine the percentage of students falling in the typical and strength ranges for social-emotional development (measured with parent and teacher ratings on the DECA). In the general population, about 84% of children are expected to score in these ranges.[41] Spring ratings showed that DPP students exceeded these expectations, with more than 84% students rated by both parents and teachers in the typical and strength ranges on all indicators of social-emotional development (see Chart 13).

Chart 13. Weighted Percentages of Students Scoring in the Typical Range or Above on Spring Social-emotional development Surveys by Rater, 2018-2019
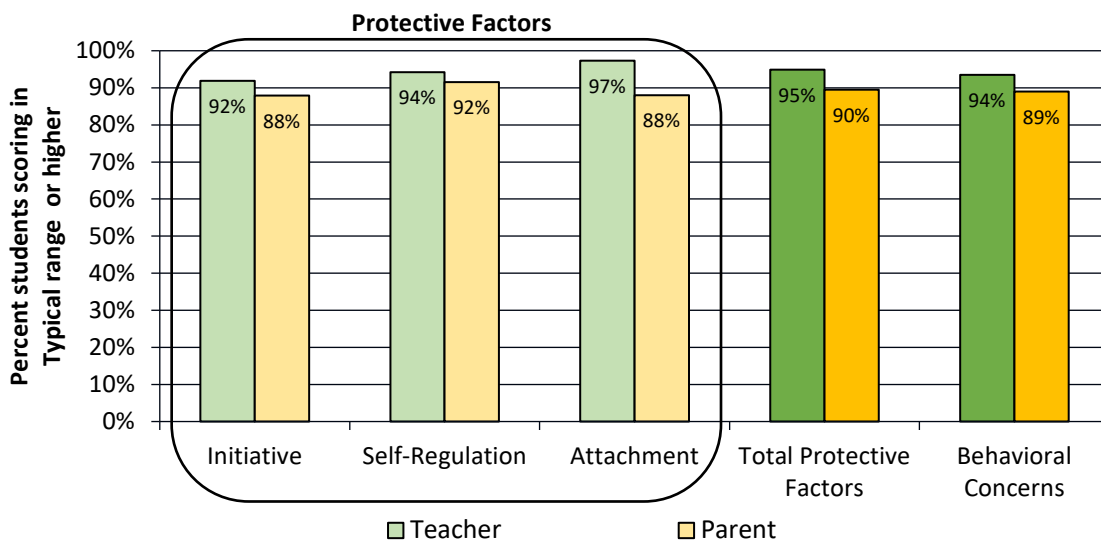

Chart 14 shows the proportion of students rated by parents and teachers, respectively, who were categorized into the adaptive executive function range in the spring. Teachers rated 79% of students in the adaptive (no concern) range for both working memory and inhibition. Parents rated 87% of students in the adaptive range for working memory and 74% for inhibition.

---

41 LeBuffe, P.A., & Naglieri, J.A. (1999). Technical manual for the Devereux Early Childhood Assessment (DECA). Villanova, PA: Devereux Foundation.

Chart 14. Weighted Percentages of Students in the Adaptive Executive Function Range on Spring Executive Function Surveys by Rater, 2018-2019
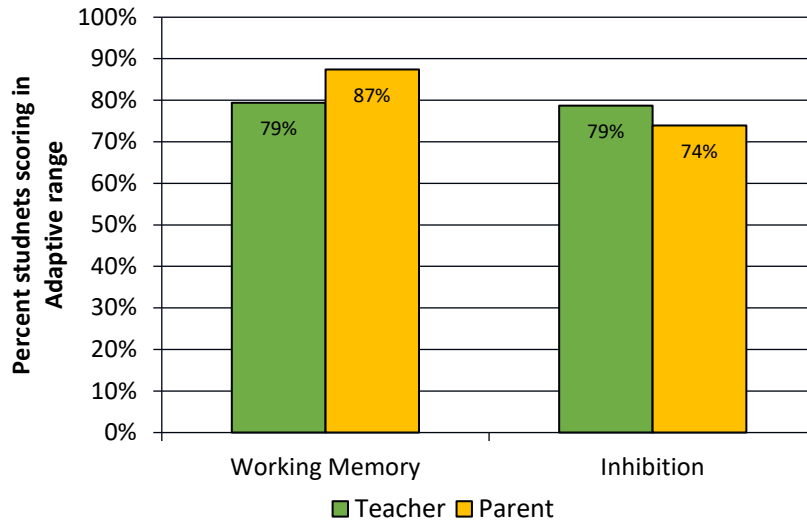
Kindergarten Readiness – No Differences by Participation Type

A comparative analysis by participation type (half-day, full-day, and extended-day) was conducted for all kindergarten readiness outcomes. A series of ANCOVAs (analysis of covariance, controlling for income tier) revealed no statistically significant differences in pre-academic, social-emotional, or executive function outcomes based on participation type.[42] The effect sizes were very small and not meaningfully different (ranging from .002 to .015).

Cohort 11 Preschool Quality in 2018-2019

Preschool Quality

Information regarding the quality of the 116 preschools where sample students were enrolled was obtained from two sources: a) the Colorado Shines Quality Rating Improvement System (adopted in January 2015; Colorado Shines QRIS) and b) classroom observations using the Pre-K CLASS® (Classroom Assessment Scoring System) tool that were conducted specifically for this evaluation project combined with observations that were conducted during the same program year for the ongoing DPP CLASS® ratings requirements.

Colorado Shines Ratings

Early learning programs are rated through Colorado Shines on a scale of 1 to 5:[43]

---

[42] ANCOVAs by participation type could not be conducted for Spanish assessments because only 6 students were enrolled in Spanish attended preschool half-day and only 8 students assessed in Spanish were enrolled in extended-day preschool.

[43] Colorado Departments of Human Services and Education. Retrieved from http://coloradoshines.force.com/ColoradoShines/programs?p=Your-Program-Colorado-Shines

Level 1: Program currently licensed with the State of Colorado

**Level 2:** Program is licensed and in good standing, plus:

- Has a quality improvement plan in place
- Has conducted the Level 2 Quality Indicator Program Assessment
- Has registered staff in the Colorado Shines Professional Development Information System (PDIS)
- Has completed Colorado Shines Level 2 E-learning Courses

Levels 3-5: Program is licensed and in good standing, plus:

- Has completed the Level 2 requirements
- Has been assessed and rated by a Colorado Shines Quality Ratings Assessor based on points in five categories (workforce qualifications, family partnerships, administration, learning environment, child health)

Within the sample, data were available from the Colorado Department of Education for all program sites (*n* = 116). Chart 15 presents the percentage of programs at each quality level. Nearly 10% of programs were rated at a Level 3. Over 80% of programs were rated Level 4. About 5% of programs were rated as Level 5. Only three preschools were rated at a Level 1 designation (licensed) and 5 preschools earned a Level 2, indicating that very few programs were of the lowest quality.

Chart 15. Colorado Shines Ratings for Classrooms with DPP Study Participants
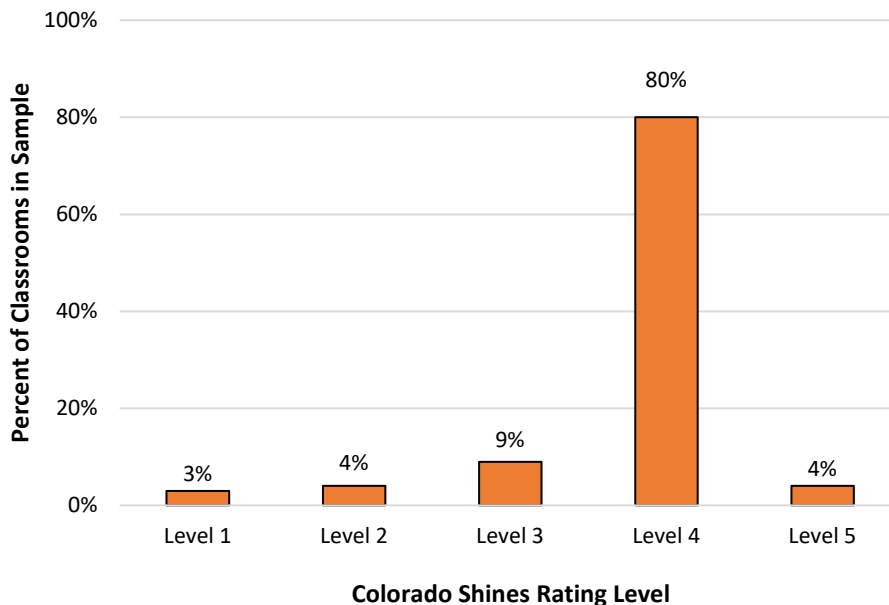
Due to DPS' participation in the alternative pathways for Colorado Shines, through which all DPS Pre-K sites receiving a Level-4 rating,[44] analyses could not be conducted to test whether there were differences in Colorado Shines ratings by provider type (DPS vs. community) was associated with the level of Colorado Shines rating.[45]

Classroom Observations

Chart 16 displays mean scores for the 120 classrooms (representing 172 students) that were observed using the Pre-K CLASS® tool. This tool is used to asses three domains of teacher-child interaction quality, each with a total possible score of 7. *Emotional Support* evaluates relationship aspects including positive climate, negative climate, teacher sensitivity, and regard for student perspectives. *Classroom Organization* assesses the interactions in terms of behavior management, productivity, and instructional learning formats used by the teacher. *Instructional Support* examines concept development, quality of feedback, and language modeling.

Studies using the Pre-K CLASS® conducted by the Office of Head Start (OHS) have consistently shown average preschool classroom scores are higher in the domains of Emotional Support (typical range: 5.5-6.5) and Classroom Organization (typical range: 4.5-6.0) than in the domain of Instructional Support (typical range: 2-3).[46] Consistent with these national findings, average Emotional Support and Classroom Organization scores for DPP sample classrooms were high, while average scores for Instructional Support were in the low middle-range of the total possible score of 7. These scores are shown in Chart 16, along with scores from on-site reviews of a national sample of Head Start classrooms.[47] Average scores for Emotional Support and Classroom Organization were slightly higher than average scores from the OHS review (and slightly higher for this year's Cohort than for the previous Cohort's classrooms[48]).

Chart 17 provides information about the variability in these domain scores. For Emotional Support, all but one classroom scored in the high range (scores above 5). For Classroom Organization, all but 6 classrooms scored in the high range. For Instructional Support, a little over half of classrooms scored in the low range, and 46% scored in the middle range, and no classrooms scored in the high range.

---

[44] https://www.coloradoshines.com/programs?p=Frequently-Asked-Questions-Program
[45] Moder, K. (2010). Alternatives to F-Test in One Way ANOVA in case of heterogeneity of variances (a simulation study). *Psychological Test and Assessment Modeling, 52*(4), 343-353.
[46] Office of Head Start, National Center on Early Childhood Development, Teaching and Learning. *CLASS® brief: Understanding and using CLASS® for program improvement.* https://eclkc.ohs.acf.hhs.gov/sites/default/files/pdf/class-brief-understanding-using-class-program-improvement.pdf
[47] Office of Head Start, National Center on Early Childhood Development, Teaching and Learning. *A national overview of grantee CLASS® Scores in 2018.* Retrieved from https://eclkc.ohs.acf.hhs.gov/data-ongoing-monitoring/article/national-overview-grantee-class-scores-2018
[48] Green, S., Mangels, D., & Reale, M. (2018). *Denver Preschool Program Child Outcomes Study 2017-2018: Part A: Preschool Progress and School Readiness Report, Cohort 10.* Denver, CO.

Chart 16. Average CLASS® *Domain Scores (n=120 classrooms)*

Average CLASS Domain Scores:

Domain	DPP KRS 1819	OHS CLASS® 2018
Emotional Support	6.54	6.08
Classroom Organization	6.09	5.8
Instructional Support	3.13	2.96

Legend: DPP KRS 1819, OHS CLASS® 2018
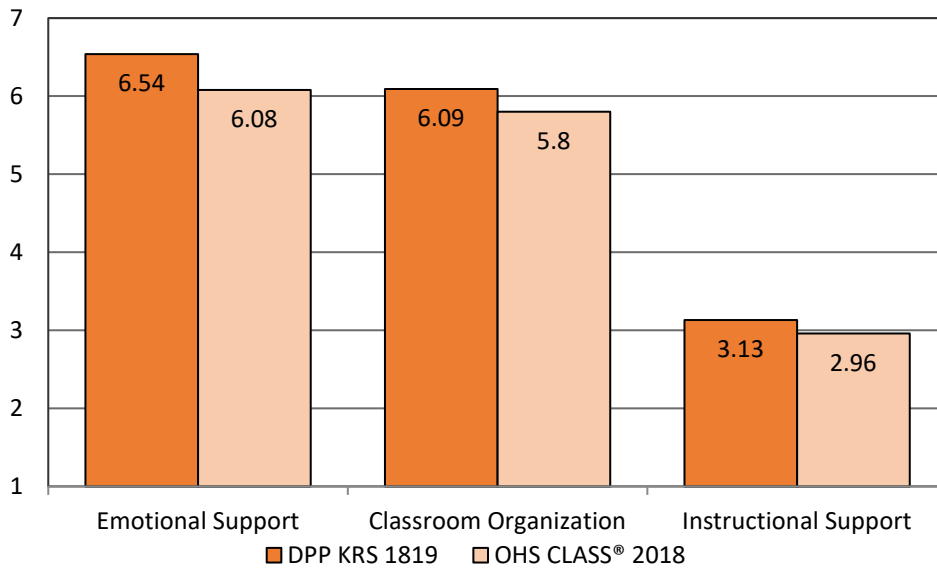
Chart 17. Distribution of CLASS® *Scores by Domain (n = 120 classrooms)*

Percent Classrooms in Sample:

Domain	Low (below 3)	Middle (3-5)	High (above 5)
Emotional Support		1%	99%
Classroom Organization		5%	95%
Instructional Support	46%	51%	3%

Legend: Low (below 3), Middle (3-5), High (above 5)
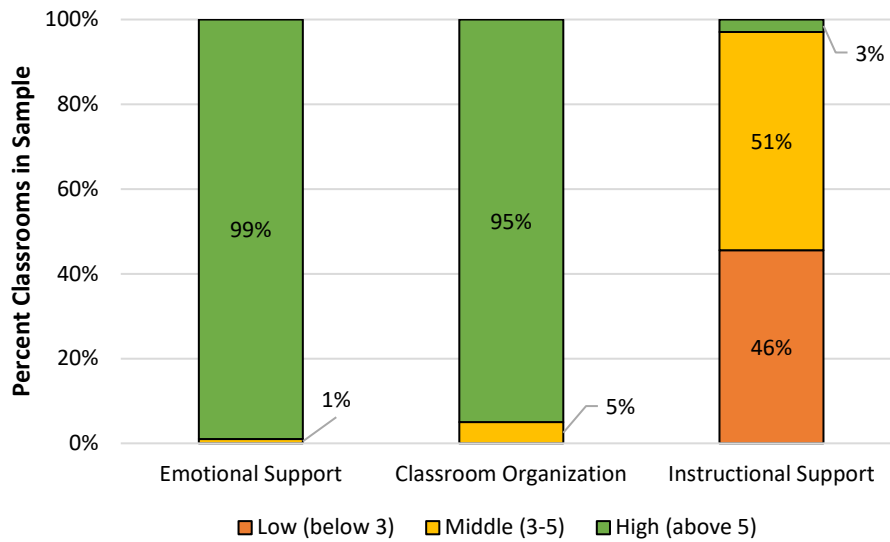
Analyses were conducted to test for differences in CLASS® domain scores by provider type. Average scores for all CLASS® subscales were significantly different by provider type, with DPS classrooms scoring higher than classrooms at community sites in each domain (see Chart 18).[49]

Chart 18. CLASS® Domain Scores by Provider Type[50]

Chart showing CLASS® Domain Scores by Provider Type:
- Emotional Support: Community = 6.41, DPS = 6.62
- Classroom Organization: Community = 5.84, DPS = 6.26
- Instructional Support: Community = 2.91, DPS = 3.28

Legend: Community (n=49), DPS (n=71)
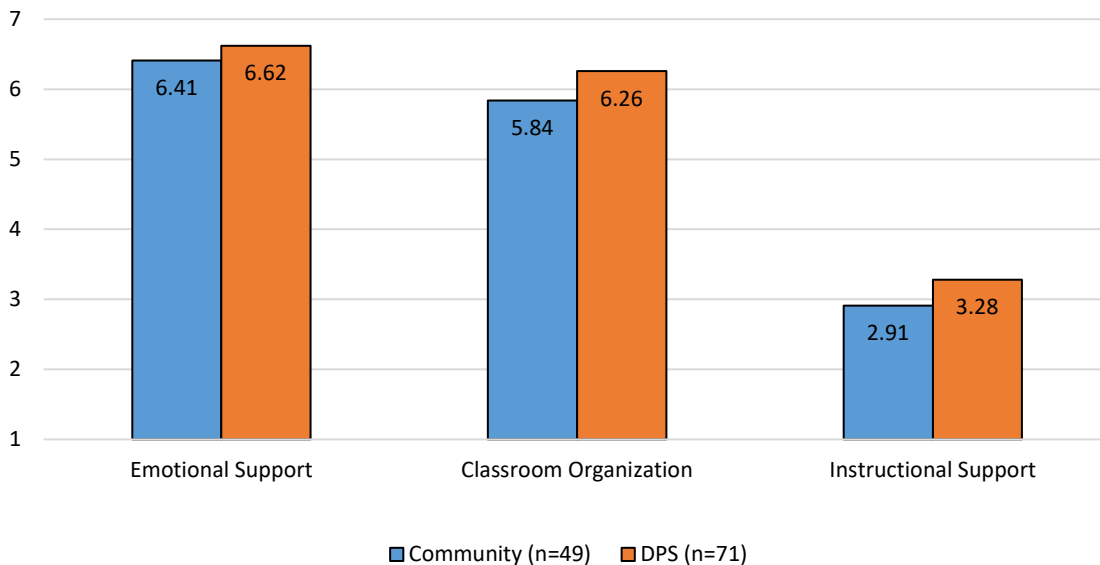
Does quality impact child outcomes?

The relationships between preschool quality and child outcomes were examined using CLASS Observation scores as indicators of preschool quality.[51] Because DPP strives for high quality, there is limited variability in quality levels that is observable with these tools (see Chart 17). A natural-log transformation was conducted on the Emotional Support and Classroom Organization scores prior to analyses to reduce skewness.[52]

To examine the relationship between CLASS scores and child outcomes, partial correlations were computed between all spring assessment scores and scores in each of the CLASS domains, controlling for fall assessment scores. There were significant positive correlations between Emotional Support and early literacy scores in both English and Spanish, $r = .21$, $p < .01$, and $r = .37$, $p < .05$, respectively. Emotional Support was also positively correlated with Spanish vocabulary scores, $r = .33$, $p < .05$. Classroom Organization scores were positively correlated with early literacy (assessed in English), $r = .20$, $p < .05$, and parent ratings of inhibition (executive function factor), $r = .19$, $p < .05$. In other words, students in classrooms with higher Emotional Support scores made greater gains in English literacy and

---

[49] Emotional Support: $t(71.44) = 2.81$, $p < .05$, Hedges' $g = .58$ (medium effect); Classroom Organization: $t(70.97) = .3.85$, $p < .001$, Hedges' $g = .78$ (large effect); Instructional Support: $t(118) = 2.78$, $p < .05$, Hedges' $g = .51$ (medium effect).

[50] Standard deviations: Emotional Support: Community = 0.48, DPS = 0.29; Classroom Organization: Community = 0.69, DPS = 0.41; Instructional Support: Community = 0.68, DPS = 0.75

[51] Because of the limited variability in Colorado Shines rating levels (86% of sample students were attending schools with a rating of 4), associations between Shines ratings and student outcomes were not examined.

[52] Emotional Support and Classroom Organization scores were negatively skewed, with the vast majority of scores in the High range (5-7) and only a handful of scores below 5.

Spanish literacy and vocabulary, on average, and students in classrooms with higher Classroom Organization scores made greater gains in English literacy and inhibition, on average. There were no significant correlations between Instructional Support scores and any of the spring assessment scores (including English and Spanish direct assessments and parent and teacher ratings on DECA and CHEXI).

Conclusion

As a representative sample of the full population of students enrolled in DPP in 2018-2019, Cohort 11 students demonstrated that overall DPP students are ready for kindergarten in pre-academic, social-emotional development, and executive function domains. In at least one language, 85% of Cohort 11 students met or exceeded expectations in vocabulary, 61% met or exceeded expectations in early literacy, and 79% met or exceeded expectations in math. Dual-language learners in particular demonstrated accelerated growth in vocabulary and math over the course of the preschool year.

Social-emotional development and executive function are valuable components of school readiness and promote a greater understanding of the whole child and the learning approaches DPP students take into kindergarten. The Cohort 11 sample demonstrated that DPP students fare well with regard to initiative, attachment, self-regulation, behavioral concerns, working memory, and inhibition – all of which are critical elements in growth and learning.

This evaluation study has shown year-to-year that students benefit from attending high-quality preschools. These results also demonstrate that there continue to be opportunities for growth in pre-academic, social-emotional, and executive function domains among Denver's preschool students. DPP can use these findings to partner with school districts and independent providers to identify students who are most likely to benefit from additional supports as they enter kindergarten.

Appendix A. Sample Characteristics 2018-2019

Characteristic	Entire Sample, weighted[1]	By Provider Type, Unweighted		
		DPS	Community	Significance of Difference by Provider Type[2]
Sex				$\chi^2(1) = 0.83$, *n.s.*
Female	48.8%	46.6%	52.6%	
Male	51.2%	53.4%	47.4%	
Race/Ethnicity				$\chi^2(5) = 4.43$, *n.s.*
Asian or Pacific Islander	2.6%	2.5%	2.6%	
Black (not of Hispanic origin)	10.4%	8.5%	13.8%	
Hispanic	34.8%	39.0%	27.6%	
White (not of Hispanic origin)	43.1%	40.7%	47.4%	
Multi-Racial	3.4%	3.4%	3.4%	
Other	5.7%	5.9%	5.2%	
Primary Language				$\chi^2(3) = 8.29$, $p < .05$
English	77.4%	72.0%	85.3%	
Spanish	19.7%	24.6%	11.2%	
Other Language	2.6%	3.4%	2.6%	
Not Reported	0.3%	0.0%	0.9%	
DPP Income Tier[3]				$\chi^2(5) = 11.74$, $p < .05$
Tier 1	22.3%	17.8%	30.2%	
Tier 2	17.0%	17.8%	15.5%	
Tier 3	15.9%	18.6%	11.2%	
Tier 4	5.1%	5.1%	5.2%	
Tier 5	20.7%	17.8%	25.9%	
Tier 6 (Opt out)	18.9%	22.9%	12.1%	
Region of Denver				$\chi^2(4) = 1.64$, *n.s.*
Central	13.6%	14.4%	12.1%	
Northeast	29.7%	31.4%	26.7%	
Northwest	17.1%	16.1%	19.0%	
Southeast	10.6%	9.3%	12.9%	
Southwest	29.0%	28.8%	29.3%	

[1] The weighted sample results are representative of the population of children enrolled in DPP in 2018-2019.

[2] There were significant differences between community and DPS sites among proportions of children by primary language and income tier.

[3] DPP income tiers are determined using family income and family size.