Denver Preschool Program: Report on Child Outcomes

2013-14 School Year

Prepared for the Denver Preschool Program
By Sheridan Green, Ph.D., Caroline Ponce, M.S., MPH, and Diana Mangels, M.A.
Clayton Early Learning
Based on original analytic design and report template by Mary Klute, Ph.D.

CONTENTS

EXECUTIVE SUMMARY

The Denver Preschool Program (DPP) is a taxpayer-funded initiative aimed at increasing access to high-quality preschool for Denver's 4-year old children. DPP operates on the premise that preschool plays an important role in the academic and social-emotional development of children and that participating in a high-quality preschool experience, even for only one year, can have a positive impact on a child.

The program encourages families to enroll their children in preschool by providing tuition credits to parents to offset the cost of preschool. The size of the tuition credit each family receives is determined by the family's income, the size of the family, and the quality rating of the preschool the child attends. DPP also provides funding for preschools serving children who live in Denver to obtain a DPP quality rating. Participating programs also receive access to professional development opportunities (e.g., training and coaching) and quality improvement grants to assist them in their efforts to improve their quality.

Clayton Early Learning Institute collaborates with Augenblick, Palaich and Associates to complete an annual evaluation of DPP. This report details the work completed by the Institute at Clayton Early Learning, which is focused on questions related to the development of children enrolled in DPP both during their preschool year and beyond.

QUESTION 1: DO CHILDREN MAKE PROGRESS IN THEIR DEVELOPMENT WHILE IN DPP EARLY CHILDHOOD ENVIRONMENTS?

Children did make significant progress in their academic and social-emotional development during their preschool year. With respect to academic skills, assessments of all children in English demonstrated that children made progress in the areas of vocabulary, literacy, and math skills. Spanish-speaking children also made progress in their vocabulary, literacy, and math skills assessed in Spanish over the course of their preschool year. The gains observed were above and beyond what would be expected based on normal development. Progress was observed in social-emotional development as well. Over the course of the preschool year, teachers reported that children demonstrated significantly more protective factors and significantly fewer behavioral concerns.

QUESTION 2: TO WHAT EXTENT AND IN WHAT AREAS ARE CHILDREN ENROLLED IN DPP READY FOR KINDERGARTEN?

Results of the evaluation suggest that the vast majority of children are ready for school, both academically and social-emotionally. When considering both languages of assessment, we concluded that relatively few children had scores in the risk range (below 85) on assessments of their vocabulary, literacy and math skills. These standardized assessments are scaled such that 84% of the general population would be expected to score above the at-risk range (a score of 85 or above). Scores for literacy and math in this sample clearly exceed that threshold. Vocabulary scores in this sample fell just below this threshold. Meeting a more stringent criterion[1] (a score of 100 or above) when both languages of assessment were considered, we found 60% for vocabulary and about three-

[1] We also considered a more stringent criterion to examine readiness, namely scores that met or exceeded the population average (a score of 100). The assessments are scaled such that half of children in the general population would be expected to meet or exceed this threshold

quarters for literacy and math. When teachers rated children's behaviors, their ratings of protective factors were high for most children. Protective factors were rated as an area of concern by teachers for 14% of children in this sample. Teachers' ratings of behavioral concerns were low on average. Teachers identified behavioral concerns as an area of concern for about 9% of children. Parents identified protective factors as an area of concern for about 20% of children and behavioral concerns as an area of concern for about a third of children. The DECA, the social-emotional assessment we used, provides T-scores, which are scaled such that nearly 16% of the general population would be expected to be identified as having a concern. All of the teachers' ratings are within the threshold. Parents' ratings of protective factors, however, exceeded that threshold. Parents identified more below-threshold protective factors. They additionally reported above-threshold behavioral concerns for about twice as many children as would be expected based on the way in which the assessment is scored.

QUESTION 3: DO CHILDREN FROM DIFFERENT INCOME LEVELS AND WITH DIFFERENT PRIMARY LANGUAGES MAKE SIMILAR PROGRESS IN THEIR DEVELOPMENT WHILE IN DPP EARLY CHILDHOOD ENVIRONMENTS?

Results of this year's study revealed a consistent pattern of results for income tier (defined by income adjusted for family size). In all cases, the income tier by time interaction was non-significant, indicating that children progressed similarly in these areas over the course of their preschool year, regardless of their income tier. Child of lower income tiers started and ended the year with lower scores on the assessments than their more wealthy counterparts. Likewise, analyses of primary language groups revealed that children whose primary language is not English started the year lower than their primarily English-speaking counterparts on English vocabulary and the social-emotional assessments. Unlike last year, the children whose primary language was not English did not increase at a greater rate than the primarily English-speakers; in fact, the persistence of low vocabulary scores was of concern.

QUESTION 4: DO CHILDREN WHO RECEIVED DPP TUITION CREDITS COMPARE FAVORABLY WITH THE DISTRICT AS A WHOLE ON ASSESSMENTS ADMINISTERED BY DENVER PUBLIC SCHOOLS IN KINDERGARTEN AND BEYOND?

Cohorts 2, 3, 4, and 5 were demographically similar to the populations of children in third grade, second grade, first grade and kindergarten, respectively, in terms of their gender and ethnic backgrounds. A smaller proportion of children from each cohort qualified for free or reduced lunch than in the district as a whole, but this was most pronounced for Cohort 3.

Cohort 2 children were compared to the population of third graders in DPS. DPP graduates scored proficient or advanced on the reading TCAP compared at a slightly higher rate (6%) than the district as a whole.

Cohort 3 children were compared to all second graders in DPS. Among children whose reading was assessed in English in second grade, DPP graduates were more likely to be reading at or above grade level than children in the district as a whole. Among children assessed in Spanish, DPP graduates were slightly less likely to be reading at grade level than the district as a whole.

Cohort 4 children were compared to the population of first graders in DPS. Among children whose reading was assessed in English, the proportion of children in Cohort 3 who were reading at or above grade level exceeded the proportion in the district as a whole. Among children assessed in Spanish at the end of the first grade year, the

proportion of DPP graduates reading at or above grade level was also greater than the proportion in the district as a whole.

Cohort 5 children were compared to the population of kindergarteners in DPS. For children assessed in both English and Spanish at the end of kindergarten, DPP graduates were much more likely to be reading at or above grade level than in the district as a whole.

QUESTION 5: IS ATTENDANCE AT HIGHER-RATED PRESCHOOL PROGRAMS ASSOCIATED WITH GREATER KINDERGARTEN READINESS AND LATER ACADEMIC SUCCESS?

Our ability to examine quality in conjunction with later academic success for cohorts 1 and 2 was limited by the lack of variability in the Qualistar rating. Very few children had been enrolled in preschools with less than a star 3 rating. In our analyses, we did not find a strong association between available measures of preschool quality and reading skill in second or third grade.

Starting with Cohort 3, with hopes for increased variability, we examined the association between CLASS observation scores and child outcomes. For the current preschool Cohort (6) there were no significant correlations between the English academic assessments (vocabulary, literacy, math, and sustained attention) and classroom quality. There were also no significant correlations for Spanish assessments. Similarly, there were no associations observed for the spring teacher DECA ratings and any of the CLASS domains after controlling for fall ratings.

Children in Cohort 5 who were enrolled in classrooms scoring higher on two of the three CLASS domains (i.e., *Classroom Organization* – assessing classroom routines and procedures and *Instructional Support* – assessing implementation of curriculum to support development) made greater gains over the preschool year in their literacy skills assessed in English. Children in classrooms that were rated higher on Instructional Support also demonstrated greater gains in Protective Factors. For Cohorts 3 and 4, there were no associations between CLASS observation scores and reading assessment scores in either language.

SUMMARY AND FUTURE DIRECTIONS

This evaluation described children's progress during the course of their DPP preschool year. In general, children progressed in their vocabulary, literacy, and math skills as assessed in both English and Spanish at a rate which exceeded what would be expected simply because of maturation, with the exception of Spanish-speakers English receptive vocabulary. Children demonstrated positive changes in their social-emotional functioning over time; teachers reported that children demonstrated more positive behaviors and fewer negative behaviors at the end of the school year than at the beginning.

Overall, children in this study were enrolled in DPP preschools that were of high quality and the children made excellent progress over the course of their preschool year, on average. The results of this study also suggest that DPP graduates tend to demonstrate similar or greater reading proficiency in kindergarten, first grade, and second grade than the district as a whole. The only exception to this was a group of Cohort 3 children assessed in Spanish in second grade (which matches the previous year's findings from this cohort). Results from future years of this annual evaluation will provide the opportunity to replicate these findings as well as to continue to follow these cohorts of children as they move through elementary school.

INTRODUCTION

The Denver Preschool Program (DPP) is a taxpayer-funded initiative aimed at increasing access to high-quality preschool for Denver's 4-year old children. DPP operates on the premise that preschool plays an important role in the academic and social-emotional development of children and that participating in a high-quality preschool experience, even for only one year, can have a positive impact on a child.

The program encourages families to enroll their children in preschool by providing tuition credits to parents to offset the cost of preschool. The size of the tuition credit each family receives is determined by the family's income, the size of the family, and the quality rating of the preschool the child attends. DPP provides funding for preschools serving children who live in Denver to obtain a DPP quality rating. Participating programs also receive access to professional development (e.g., training and coaching) and quality improvement grants to assist them in their efforts to improve their quality.

The Institute at Clayton Early Learning collaborates with Augenblick, Palaich and Associates to complete an annual evaluation of DPP. This report details the work completed by Clayton Early Learning, which is focused on questions related to the development of children enrolled in DPP both during their preschool year and beyond[2]. This portion of the evaluation was designed to address five questions relevant to children's development while enrolled in DPP and beyond:

1. Do children make progress in their development while in DPP early childhood environments (i.e., language, literacy, mathematics, and social-emotional development)?
2. To what extent and in what areas are children enrolled in DPP ready for kindergarten?
3. Do children from different income levels and with different primary languages make similar progress in their development while in DPP early childhood environments?
4. Do children who received DPP tuition credits compare favorably with the district as a whole on assessments administered by Denver Public Schools (DPS) in kindergarten and beyond?
5. Is attendance at higher-rated preschool programs associated with greater kindergarten readiness and long-term academic success (as measured by TCAP)?

The 2013-14 school year marks the seventh year of the DPP program. The cohort from the second year of DPP's operation was the first cohort for which we were able to implement our evaluation design. This cohort was expected to be enrolled in third grade during the 2012-13 school year, the first grade in which students take the TCAP. As such, this year's annual report represents the second time that we were able to address question 5.

[2] Augenblick, Palaich and Associates prepares a separate report detailing the growth of the DPP program over time, characteristics of enrolled children, the availability of quality preschool slots to families, and information relevant to participants' experience with the program.

METHODS

SAMPLE

The sample of children included in this report is drawn from six (6) cohorts of children who were enrolled in DPP during the year before they were eligible to attend kindergarten (see Table 1).

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

	School Year					
	08-09	09-10	10-11	11-12	12-13	13-14
Cohort 1	Preschool	Kindergarten	1st Grade	2nd Grade	3rd Grade	4th Grade
Cohort 2		Preschool	Kindergarten	1st Grade	2nd Grade	3rd Grade
Cohort 3			Preschool	Kindergarten	1st Grade	2nd Grade
Cohort 4				Preschool	Kindergarten	1st Grade
Cohort 5					Preschool	Kindergarten
Cohort 6						Preschool

COHORT 1

The 2008-09 school year was the second year of DPP's operation. This was the first year that we were able to carry out our evaluation as designed, including drawing a sample of children that was representative of the population of children enrolled in DPP at that time and assessing those children in the fall and spring of their preschool year.[3] Henceforward, this cohort of children will be referred to as Cohort 1. The total sample size for Cohort 1 was 207; 200 children were assessed in the fall and spring of the preschool year. We were able to obtain DPS IDs for 200 of these children (97% of the original sample).

Cohort 1 children were expected to be in the fourth grade during the 13-14 school year (see Table 1). No new data were obtained related to spring reading assessment data for the sample. (75% of the whole sample; 78% of those for whom we had obtained DPS IDs). All of these children were in third grade as expected.

COHORT 2

Starting with the 09-10 school year, we modified our approach to sampling slightly. To maximize the conclusions we can draw about both community DPP sites and those sites in Denver Public Schools (DPS), we stratified our sample by type of provider. The result was two samples: a sample of children in community sites and a sample of children in DPS sites. Both of these samples were representative of the population of children in each type of preschool at the time of sampling. For all analyses on the sample of 200 as a whole, sampling weights were applied so that the results would be representative of the population of children enrolled in DPP at the time of sampling. For analyses comparing DPS and community sites, weights were not applied. The total sample size for Cohort 2 was 201; 200 children were assessed in the fall and spring of the preschool year. We were able to obtain DPS IDs for all 201 of these children.

[3] For more information about this sample and results from the preschool year, readers are referred to the Annual Evaluation Report. Klute, M. M. (2009). *Denver Preschool Program: Report on Child Outcomes—2008-09 School Year.* Unpublished Report. Denver: Clayton Early Learning Institute.

Cohort 2 children were expected to be in third grade during the 13-14 school year (see Table 1). We obtained reading data for 124 children (62% of the sample; 68% of the sample when sampling weights were applied). All of the children were in third grade as expected.

COHORT 3

As explained above for Cohort 2, we stratified our sample for Cohort 3 by type of provider. The result was two samples: a sample of children in community sites and a sample of children in DPS sites. Both of these samples were representative of the population of children in each type of preschool at the time of sampling. For all analyses on the sample of 200 as a whole, sampling weights were applied so that the results would be representative of the population of children enrolled in DPP at the time of sampling. For analyses comparing DPS and community sites, weights were not applied. The total sample size for Cohort 3 was 204; 200 children were assessed in the fall and 199 were assessed in the spring of the preschool year. We were able to obtain DPS IDs for 200 of these children.

Cohort 3 children were expected to be in second grade during the 13-14 school year (see Table 1). We obtained reading data for 133 children (65% of the sample; 77% of children for whom we were able to obtain DPSIDs; 73% of the total sample when sampling weights were applied). All of the children were in second grade as expected.

COHORT 4

The sample for Cohort 4 was also stratified by type of provider. The result was two samples: a sample of children in community sites and a sample of children in DPS sites. Both of these samples were representative of the population of children in each type of preschool at the time of sampling. For all analyses on the sample of 200 as a whole, sampling weights were applied so that the results would be representative of the population of children enrolled in DPP at the time of sampling. For analyses comparing DPS and community sites, weights were not applied. The total sample size for Cohort 4 was 203; 200 children were assessed in the fall and the spring of the preschool year. We were able to obtain DPS IDs for 203 of these children.

Cohort 4 children were expected to be in first grade during the 13-14 school year (see Table 1). We obtained reading data for 144 children (71% of the sample; 81% of the sample when sampling weights were applied). All of the children were in first grade as expected.

COHORT 5

The sample for Cohort 5 was also stratified by type of provider. The result was two samples: a sample of children in community sites and a sample of children in DPS sites. Both of these samples were representative of the population of children in each type of preschool at the time of sampling. For all analyses on the sample of 200 as a whole, sampling weights were applied so that the results would be representative of the population of children enrolled in DPP at the time of sampling. For analyses comparing DPS and community sites, weights were not applied. The total sample size for Cohort 5 was 211; 200 children were assessed in the fall and the spring of the preschool year. We were able to obtain DPS IDs for 208 of these children.

Cohort 5 children were expected to be in kindergarten during the 13-14 school year (see Table 1). We obtained reading data for 147 children (70% of the sample; 77% of the sample when sampling weights were applied). All of the children were in kindergarten as expected.

COHORT 6

SAMPLING PLAN

As with previous years, we stratified our sample for Cohort 6 by type of provider. The result is two samples: a sample of children in community sites and a sample of children in DPS sites. During the DPP enrollment process, parents were asked if they would be willing to be contacted about participation in the evaluation study.[4] In August 2013, a sample of 100 children enrolled in DPS sites was drawn from the group of families that volunteered to participate (henceforth referred to as "volunteers"). In September 2013, a sample of 100 children enrolled in community sites was drawn from the group of families that volunteered. Prior to drawing each of these samples, volunteers and those who refused to be contacted about the evaluation (henceforth referred to as "non-volunteers") were compared on the following demographic characteristics: sex of the child, ethnicity, Qualistar rating of the preschool program, home language, child language, and region of the city in which the child lives. DPP income tier, which takes into account both family size and income, was also examined. It is comprised of six levels, with tier 1 representing the lowest income. More detail on how income tier is determined can be found in the appendix. Volunteers and non-volunteers were also compared on whether they declined to provide income information.

COMMUNITY SITES

In community sites, there were significant differences between the 762 volunteers and the 346 non-volunteers for variables. There was a significant difference between volunteers and non-volunteers only on primary language.[5] Follow-up analyses revealed differences between volunteers and non-volunteers who children spoke primarily Spanish. Seventy-six percent of Spanish-speaking parents volunteered compared with 67% of non-Spanish-Speakers. To adjust for these differences, the sampling frame was stratified by primary language (Spanish vs. other). The proportion of children drawn from each stratum was adjusted to match the proportions in the population of children enrolled in DPP at the time of sampling. The result was a sample of 100 that was representative of the community site population as a whole in September 2013 with respect to the variables examined. The sample was drawn with replacement; if a selected child was deemed ineligible for the study,[6] a selected family was unable to be contacted to obtain informed consent to participate in the study, or if a selected family refused to participate in the study, a replacement child was randomly drawn from the same stratum.

DPS SITES

In DPS sites, significant differences were detected between the 1,471 volunteers and 787 non-volunteers on three variables.[7] First, a significant difference was detected for ethnicity.[8] Follow-up analyses revealed that this was due to a difference between volunteers and non-volunteers from two ethnic groups: Asian and multi-ethnic. Parents of Asian and multi-ethnic children were significantly less likely to volunteer than other parents. Forty-one percent of parents of Asian children volunteered compared with 66% of parents of children from other ethnic groups. Fifty-two percent of parents of multi-ethnic children volunteered compared with 66% of parents of

[4] Information about the evaluation was provided on the DPP application, which was available in both English and Spanish.
[5] $\chi^2_4=18.60$, $p<.01$
[6] Typically children were deemed ineligible because they were no longer enrolled in a DPP preschool at the time the family was contacted for participation.
[7] Because of the very large sample size and associated statistical power, a p-value of .01 was used for determining statistical significance.
[8] $\chi^2_5=34.62$, $p<.0001$

children from other ethnic groups. In contrast to last year, parents of white children were not significantly more likely to volunteer than parents of children from other ethnic groups (percent who volunteered: 69% white, 63% non-white).

In addition, the likelihood of volunteering varied significantly by home language and child primary language.[9] Parents of children with primary languages and home languages that were not English were more likely to volunteer than parents of children with English primary languages and home languages. Seventy-six percent of parents of children with home languages other than English volunteered compared with about 63% of parents of children with English as their primary language. Seventy-two percent of parents of children with primary languages other than English volunteered compared with about 64% of parents of children with English as their primary language.

To adjust for these differences, the sampling frame was stratified by ethnicity (Asian and multi-ethnic), home language (English vs. other), and primary language (English vs. other). The proportion of children drawn from each stratum was adjusted to match the proportions in the population of children enrolled in DPP at the time of sampling. The result was a sample of 100 that was representative of the DPS site population as a whole in August 2013 with respect to the variables examined. As with the sample from community sites, the sample was drawn with replacement; if a selected child was deemed ineligible for the study, a selected family was unable to be contacted to obtain informed consent to participate in the study, or if a selected family refused to participate in the study, a replacement child was randomly drawn from the same stratum.

SAMPLING WEIGHTS

At the time of sampling, about one third of children enrolled in DPP were attending community sites and the remaining two-thirds were attending DPS sites.[10] The sample was divided evenly between community sites and DPS sites. As a result, the sampling design involved oversampling children from community sites. When analyzing data for the sample of 200 as a whole, it was important to weight the sample so that both program types had weights in the analysis that are comparable to each group's proportion of the total population. The result is an analysis of data that are representative of the DPP population as a whole.

SAMPLE CHARACTERISTICS

Characteristics of the fall sample are summarized in Table 2.[11] The sample was approximately equally split between boys and girls. Hispanics represented less than half of the sample (44%); the next most common ethnic group was Whites. African-Americans made up slightly more than a tenth of the sample. About 56% of children spoke English as their primary language and in a similar percentage of their homes, English was the primary language spoken. In terms of income, over half of the children in the sample were from the lowest two income tiers. The upper bound for Tier 1 is equivalent to the Federal Poverty Guideline for 2011. The upper bound for Tier 2 is equivalent to 185% of the Federal Poverty Guideline for 2011, which is also the cutoff for free or reduced Lunch. The next most common income tier was Tier 5. About sixteen percent of families were assigned to the highest tier, Tier 6, because they opted out of the requirement to provide their income.

[9] Home language: χ^2_2=23.73, p<.01; child primary language: χ^2_2=12.30, p<.01

[10] A small number of children were enrolled in more than one DPP site. We used the site that was named as their primary preschool in the Metrix database to determine their provider type.

[11] Sample characteristics for the spring sample, which were nearly identical, are presented in the appendix.

Nearly all (93%) of the children were enrolled in preschools with a 3 or 4 star rating. Over half of children were enrolled in star 3 preschools and over a third of children were enrolled in star 4 preschools. The highest proportion of children lived in northeast Denver, while the smallest proportion of children lived in southeast Denver. The right hand side of Table 2 presents demographic characteristics by provider type. With regard to the demographic compositions of the two samples, three of the differences were statistically significant – ethnicity, primary language, and geographic region of the city.

In the spring 2014, there were six children that were lost to follow-up for the following reasons: three children moved out of Denver before the spring round and three children withdrew from their preschool programs and did not enroll in another DPP preschool. An alternate from the same stratum was selected for each of these children and assessed during the spring round.[12] As a result, the total sample size for the 2012-13 school year was 206.

REPRESENTATIVENESS OF THE SAMPLE

Analyses were conducted to test whether the sample selected was representative of the population of children enrolled in DPP. These analyses were conducted separately for children enrolled in community sites and those enrolled in DPS sites. Because enrollment continued after the sample was drawn, two sets of analyses were conducted to address this question. First, each of the samples of 100 (community and DPS) was compared to the population of children from which it was drawn. Second, the spring sample for each of these groups was compared to the population of children enrolled as of the end of the 2012-13 school year. Each set of analyses are described in turn below.

[12] We "refreshed" the sample in the spring to maintain the total sample size of 200. This was done because we wanted to ensure that we had a sample of at least 200 to follow into the elementary school years.

Table 2: Cohort 6 Sample Characteristics Fall 2013[1]

	Entire Sample, weighted[2]	By Provider Type, Unweighted		
Characteristic		Community	DPS	Significance of Difference by Provider Type
Sex				χ^2_1=.021; ns.
Female	48.7%	48.0%	49.0%	
Male	51.3%	52.0%	51.0%	
Ethnicity				χ^2_5=16.5; *p*=.01
Hispanic	44.1%	26.7%	51.9%	
White (not of Hispanic origin)	35.9%	45.5%	31.7%	
African-American (not of Hispanic origin)	10.7%	12.9%	9.6%	
Multi-Racial	4.5%	9.9%	1.9%	
Asian/Pacific Islander	3.9%	4.0%	3.8%	
Native American	1.0%	1.0%	1.0%	
Child's Primary Language				χ^2_1=13.5; p<.01
English	68.8%	84.3%	60.6%	
Another Language	30.9%	14.7%	38.5%	
Not Reported	1.0%	1.0%	1.0%	
Home Language				χ^2_1=5.9; n.s.
English	54.9%	73.5%	46.2%	
Another Language	20.0%	19.6%	20.2%	
Not Reported	25.1%	6.9%	33.7%	
DPP Income Tier[3]				χ^2_5=3.6; n.s.
Tier 1	37.4%	37.3%	37.5%	
Tier 2	18.8%	13.7%	21.2%	
Tier 3	2.6%	3.9%	1.9%	
Tier 4	5.8%	7.8%	4.8%	
Tier 5	19.6%	22.5%	18.3%	
Tier 6—Income Not Reported	15.8%	14.7%	16.3%	
Star Level of Preschool				χ^2_2=5.23; n.s.
Not Yet Rated	0.0%	0.0%	0.0%	
Provisional	0.9%	2.9%	0.0%	
Star 1	0.0%	0.0%	0.0%	
Star 2	6.4%	7.8%	5.8%	
Star 3	57.9%	44.1%	64.4%	
Star 4	34.7%	45.1%	29.8%	
Region of the City				χ^2_4=12.7;p=.01
Central	12.5%	18.6%	9.6%	
Northeast	35.7%	21.6%	42.3%	
Northwest	17.6%	24.5%	14.4%	
Southeast	8.4%	7.8%	8.7%	
Southwest	25.8%	27.5%	25.0%	

[1]Some percentages do not sum to 100 because of rounding error.
[2]The weighted sample results are representative of the population of children enrolled in DPP in Fall 2013.
[3]DPP Income Tiers are determined using family income and family size. Tier 1 is the lowest income. Details on the income tiers can be found in the appendix.

FALL 2012

COMMUNITY SAMPLE

Children who were included in the community sample were compared to 1,120 children enrolled in DPP in community sites but not included in the sample on several key demographic characteristics: child gender, child ethnicity, income tier, Qualistar rating of the child's preschool, home language, child's primary language, and region of the city. The community sample was representative of the population of children enrolled in community sites in September 2013. [13]

DPS SAMPLE

Children who were included in the DPS sample were compared to 2,447 children enrolled in DPP in DPS sites who were not included in the sample. These two groups were compared on the same set of demographic characteristics described above. The DPS sample was representative of the population of children enrolled in DPS sites in August 2013. [14]

SUMMER 2013

COMMUNITY SAMPLE

Children who were included in the community sample were compared to 1,479 children enrolled in DPP by the end of the school year in community sites but not included in the sample on the same demographic characteristics described above. All analyses were non-significant, indicating that the community sample did not differ significantly from those not in the sample.[15] That is, the community sample was representative of the population of children enrolled in community sites at the end of the school year.

DPS SAMPLE

Children who were included in the DPS sample were compared to 3,467 children enrolled in DPS sites at the end of the school year who were not included in the sample. These two groups were compared on the same set of demographic characteristics described above. All tests were non-significant, indicating that the DPS sample did not differ significantly from those not in the sample.[16] That is, the DPS sample was representative of the population of enrolled children in DPS sites at the end of the school year.

---

[13] Gender: χ^2_1=.53, n.s.; ethnicity: χ^2_6=6.72, n.s.; home language: χ^2_1=4.72, n.s.; child primary language: χ^2_1=4.63, n.s.; income tier: χ^2_5=6.87, n.s.; Qualistar rating: χ^2_4=4.32, n.s.; region of the city: χ^2_4=2.01, n.s.

[14] Gender: χ^2_1=0.24, n.s.; ethnicity: χ^2_6=3.56, n.s.; income tier: χ^2_6=6.19, n.s.; Qualistar rating: χ^2_2=1.41, n.s.; home language: χ^2_1=2.46, n.s.; child primary language: χ^2_1=.98, n.s.; region of the city: χ^2_4=5.24, n.s.

[15] Gender: χ^2_1=.240, n.s.; ethnicity: χ^2_5=10.34, n.s.; income tier: χ^2_6=11.54, n.s.; Qualistar rating: χ^2_4=4.64, n.s.; home language: χ^2_1=5.76, n.s.; child primary language: χ^2_1=5.69, n.s.; region of the city: χ^2_4=1.40, n.s.

[16] Gender: χ^2_1=0.22, n.s.; Ethnicity: χ^2_5=4.37, n.s.; income tier: χ^2_6=12.12, n.s.; Qualistar rating: χ^2_2=.36, n.s.; home language: χ^2_1=1.11, n.s.; child primary language: χ^2_1=.66, n.s.; region of the city: χ^2_4=10.23, n.s.

PROCEDURES

Once parents or guardians of children selected for the study provided informed consent, children were assessed using standardized assessments at their preschools during normal school hours. Children who spoke Spanish were assessed twice by a bilingual assessor, once in English and once in Spanish, on different days. All children were assessed in English because most children are exposed to English during their DPP preschool experience and we wanted to understand their progress in English during their preschool year.

Table 3: Sample sizes by data collection type, Fall 2013 and Spring 2014

Data Collection Activity	Fall 2013	Spring 2014
Standardized Assessments—English	200	200
Standardized Assessments—Spanish	48	49
DECA—Teacher Report	196	196
DECA—Parent Report	197 (96% of the full sample *n*=206)	
Classroom Observations[1]	129 (63% of all children *n*=206)	

[1]This figure represents the number of children for whom we have a classroom observation.

After providing informed consent, teachers were asked to complete a survey about children's social-emotional development on two occasions. Assessors completed the consent process and left a survey with teachers at the time of the fall assessment. They returned approximately a week later to pick up the completed survey. In the spring, since most teachers had already completed the consent process, teachers were mailed the surveys ahead of time. Assessors picked up the completed surveys at the time of the assessment. Teachers were also asked to allow us to visit their classroom one time for a half-day observation. These observations took place throughout the school year.

Parents were mailed a survey about their children's social-emotional development in January 2014. Follow-up mailings and phone calls were used to boost response rates. Parents were asked to complete the survey just one time during the course of the school year. A Spanish version of the survey was available for parents and teachers who preferred to complete it in Spanish.

Table 3 presents the total sample sizes for each data collection activity. About one quarter of the children in the sample spoke Spanish and completed assessments in Spanish as well as English. Response rates for the teacher and parent surveys were excellent and the response rate for the classroom observations was acceptable.

MEASURES[17]

PRESCHOOL YEAR

STANDARDIZED ASSESSMENTS OF CHILDREN

Children were assessed using a battery of standardized assessments (see Table 4). Assessments included measures of children's receptive vocabulary, literacy skills, and mathematics skills. As described above, Spanish-English bilingual children were assessed in both languages. Assessments were chosen because they have been widely used in other similar studies of preschool-aged children, including two major studies of state-wide universal pre-kindergarten programs.[24] New this year to the assessment battery is a test designed to understand children's executive functioning, specifically their ability to sustain attention. The Leiter Attention Sustained (LAS) subtest is designed to measure children's ability to focus cognitive activity on specific stimuli (Roid et al., 2013). The ability to selectively focus attention on any given task is important for cognitive development, such as categorization, language comprehension, reasoning, and problem solving.

Table 4: Areas of Child Development Assessed

Area Assessed	Name of Assessment	Acronym	Language of Assessment
Sustained Attention	Leiter International Performance Scale-Revised,[18] Attention Sustained Subscale	LAS	Language free
Receptive Vocabulary	Peabody Picture Vocabulary Test-4[19]	PPVT	English
	Test de Vocabulario en Imagenes Peabody[20]	TVIP	Spanish
Literacy Skills	Woodcock-Johnson III Achievement Battery,[21] Letter-Word Identification Subtest	WJ LWI	English
	Batería III Woodcock-Muñoz,[22] Letter-Word Identification Subtest	WM LWI	Spanish
Math Skills	Woodcock-Johnson III Achievement Battery, Applied Problems Subtest	WJ AP	English
	Batería III Woodcock-Muñoz, Applied Problems Subtest	WM AP	Spanish
Social-emotional Development	Devereaux Early Childhood Assessment[23]	DECA	English or Spanish

[17] The measures described here were supplemented by information about demographic characteristics that was obtained from the contractor that handles enrollment and tuition payments for the Denver Preschool Program.
[18] Roid, G. H., Miller, L. J., Pomplun, M., Koch, C. (2013*) Leiter-3: Leiter International Performance Scale-Third Edition. Subtest: Attention Sustained*. Stoelting Company, Wood Dale, IL Cat. No. 34100M. www. Stoelingco.com
[19] Dunn, L. M., & Dunn, D. M. (2007). *Peabody Picture Vocabulary Test, Fourth Edition.* Minneapolis: Pearson Assessments.
[20] Dunn, L. M., Lugo, D. E., Padilla, E. R., & Dunn, L. M. (1986). *Test de Vocabulario en Imagenes Peabody (TVIP).* Minneapolis: Pearson Assessments.
[21] Woodcock, R. W., Schrank, F. A., Mather, N., & McGrew, K. S. (2007). *Woodcock-Johnson III, Tests of Achievement (Normative Update*). Rolling Meadows, IL: Riverside Publishing.
[22] Muñoz Sandoval, A. F., Woodcock, R. W., McGrew, K. S., & Mather, N. (2005). *Batería III Woodcock-Muñoz*. Rolling Meadows, IL: Riverside Publishing.
[23] LeBuffe, P. A., & Naglieri, J. A. (1999). *Devereux Early Childhood Assessment, User's Guide*. Lewisville, NC: Kaplan.
[24] Early, D. M., Barbarin, O., Bryant, D. M., Burchinal, M., Chang, F., Clifford, R. M., Crawford, G. M., Howes, C., Ritchie, S., Kraft-Sayre, M. E., Pianta, R. C., Barnett, W. S., & Weaver, W. (2005). *Pre-kindergarten in eleven states: NCEDL's Multi-State Study of Pre-Kindergarten and study of State-Wide Early Education Programs (SWEEP): Preliminary descriptive report.* Chapel Hill, NC: National Center for Early Development & Learning.

The structure of the Attention Sustained (AS) subtest is an individually administrated test designed to assess attention and interference in young children (ages 3-5) or in older children to adults (ages 6-75). This non-verbal assessment utilizes pantomime for instructions, allowing it to be used with children and adults regardless of verbal language skills or primary language.

PARENT AND TEACHER SURVEYS

The parent and teacher surveys consisted of a measure of children's social-emotional development called the Devereaux Early Childhood Assessment (DECA; see Table 4). The DECA is a 37-item measure with four subscales including three protective factors: Initiative, Self-Control, and Attachment, as well as a subscale devoted to Behavioral Concerns. In addition to the four subscales, there is also a Total Protective Factors scale which is the sum of the three protective factors. T-scores can be computed for all of the scales based on separate norms for parent and teacher reports. Based on T-scores, children can be categorized into 3 categories (area of concern, typical, and strength) for Protective Factors and two categories for Behavioral Concerns (area of concern and typical). In some cases, teachers or parents left some items blank on the survey. In these cases, scores were only computed if at least 75% of the items on the scale were completed.

CLASSROOM QUALITY

We supplemented archival information about classroom quality that was obtained from Qualistar (described above) with an additional observation of classrooms in which children who were part of our sample were enrolled. This additional observation was useful because Qualistar does not rate every classroom every year. In addition, while the Qualistar rating provides valuable information about global program quality, it does not shed as much light on what day-to-day experiences are like for children in the classroom. Finally, in previous years, there has been very little variability among DPP preschools on the Qualistar rating. The vast majority of sites have earned either a star 3 or star 4 rating. To address these issues, during the 2010-11 school year we added the CLASS (Classroom Assessment Scoring System), which is an observational measure of classroom quality that focuses on teacher-child interactions.[25] The 2013-14 school year was the fourth year that we observed classrooms using the CLASS. Observers visit the classroom and observe for up to six 30-minute cycles. Each cycle includes a 20-minute period of observation followed by a 10-minute period during which the observer rates the classroom using a 7-point scale on 10 dimensions. The 10 individual dimensions on the CLASS are organized into three broad domains: Emotional Support, Classroom Organization, and Instructional Support. The Emotional Support domain describes the tone of classroom climate and the extent to which the classroom is sensitive to the concerns and points of view of students. In previous large studies, classrooms have scored, on average, in the 4.5 to 5.5 range on Emotional Support. Classroom Organization describes the ways in which children's behavior, time and attention are managed and organized in the classroom. In previous large studies, classrooms have scored, on average, in the 4.5-5 range on this dimension. Finally, the Instructional Support dimension focuses on the extent to which a teacher structures learning activities and curriculum in a way that supports children's cognitive and language development. In previous large studies, classrooms have scored rather low on this dimension, on average, with scores in the 2-3 range.

[25] Pianta, R. C., LaParo, K. M., & Hamre, B. K. (2008). *Classroom Assessment Scoring System (CLASS) Manual, Pre-K.* Baltimore, MD: Brookes.

ELEMENTARY SCHOOL

Children's reading proficiency was measured using the Developmental Reading Assessment (DRA2)[26] and its Spanish language counterpart Evaluación del Desarrollo de la Lectura (EDL2).[27] Denver Public Schools administers these assessments in the spring of the kindergarten, first grade and second grade years. These assessments are criterion-referenced and part of an instructional system designed to help teachers pinpoint children's reading level and design differentiated instruction to meet the needs of all children in their classroom.[28] The assessment yields a reading level for each child. In kindergarten, a reading level of 4 is considered reading on grade level.[29] In first grade, a reading level of 16 is considered on grade level. In second grade, a reading level of 28 is considered on grade level. In third grade, a reading level of 38 is considered on grade level.

Beginning in third grade, students in Colorado take the Transitional Colorado Assessment Program (TCAP) tests.[30] These assessments are aligned with state standards and yield a score to indicate whether a student is performing at an Advanced, Proficient, Partially Proficient, or Unsatisfactory level. Students in third grade are assessed in reading, writing, and math. Reading TCAP scores are released several months before writing and math scores are released. As a result, only third grade reading TCAP scores are included in this report.

RESULTS: PRESCHOOL YEAR

PRELIMINARY ANALYSES

Table 5 presents descriptive statistics for fall and spring child outcome measures. The PPVT, TVIP, WJ and WM are all scaled such that 100 is an average score, with a standard deviation of 15. Scores within one standard deviation of the mean are considered in the average range (i.e., 85-115). All scores are adjusted for the child's age at the time of assessment. As such, one would expect a child who is developing at an average rate to have the same score over time. In both the fall and the spring, children, on average, scored in the average range for all of the standardized assessments. On average, scores for the PPVT and TVIP tended to be lower than those for the WJ and WM. It is noteworthy that for all of these assessments, there is considerable variability in children's scores, with some children scoring quite low and some scoring rather high.

The new Leiter Sustained Attention subtest of executive function is a criterion-referenced assessment (unlike the norm-referenced measures described in the paragraph above) that derives a scaled score using the raw correct and incorrect responses. A score of 10 is the national average, with any scores lower than a 7 indicating there may be some underlying attentional difficulties.

The DECA is scaled using T-scores, which have a mean of 50 and a standard deviation of 10. In both the fall and spring, teachers rated children, on average, fairly close to the national average of 50 on all of the subscales, with a slightly higher average score on self-control. Parents' ratings of children were, on average, close to the national average, with slightly lower scores on Attachment. Once again there was substantial variability in all of the scores.

---

[26] Beaver, J. M., & Carter, M. A. (2006). *The Developmental Reading Assessment, Second Edition (DRA2).* Upper Saddle River, NJ: Pearson.
[27] Ruiz, O.A. & Cuesta, V. M. (2007). *Evaluación del desarrollo de la lectura.* Parsippany, NJ: Pearson Education, Inc.
[28] *K-8 Technical Manual, Developmental Reading Assessment, Second Edition* (2009). Upper Saddle River, NJ: Pearson Education, Inc.
[29] Prior to the 2010-11 school year, a reading level of 3 was considered on grade level for kindergarten.
[30] For more information about TCAP, visit http://www.cde.state.co.us/assessment/GeneralInfo.asp

Since all children were assessed in English, regardless of their primary language, it is useful to consider whether children's scores on the English assessments differed based on whether children spoke English as their primary language. We performed *t*-tests to examine whether there were differences in PPVT, LWI, AP, and LAS by primary language group (i.e., English vs. any other language). Results for the fall round are presented in Table 6. In the fall round, there was a large difference in the scores on the PPVT by primary language. Children whose primary language was English scored nearly three standard deviations higher on the PPVT than their counterparts with another primary language. For LWI and AP, children whose primary language was English scored close to one standard deviation higher on average than their counterparts with a different primary language. All differences were statistically significant. No differences by language were observed for LAS. A similar pattern of findings was observed in the spring round (Table 7). For this round, once again, the differences were statistically significant, except in the case of LAS. Similar to the fall, the largest difference between the primary-language groups was observed for the PPVT, was over two standard deviations in magnitude. Differences between primary language groups for LWI and AP were slightly smaller than the fall, but still statistically significant. For LWI and AP, the difference between language groups was about one standard deviation in magnitude.

Table 5: Weighted Descriptive Statistics for Child Outcome Measures

Variable	**Fall 2013**				**Spring 2014**			
All Children	N	Mean	Standard Deviation	Range	N	Mean	Standard Deviation	Range
Standardized Assessments								
PPVT Standard Score	200	92.17	26.98	20-140	200	94.19	27.04	20-146
WJ LWI Standard Score	200	99.91	14.43	62-140	200	102.88	15.25	56-154
WJ AP Standard Score	200	104.70	15.47	71-140	200	104.63	14.22	67-136
LAS Scaled Score	200	7.52	4.37	0-16	200	9.57	4.56	0-20
Teacher-Rated DECA[1]								
Initiative T-Score	196	49.39	7.99	28-61	196	51.49	7.43	30-61
Self-Control T-Score	196	57.05	9.47	28-72	196	58.16	8.05	33-72
Attachment T-Score	196	50.62	9.23	28-72	196	52.88	9.30	30-72
Total Protective Factors T-Score	196	51.72	9.14	28-72	196	53.96	8.62	30-72
Behavioral Concerns T-Score	196	46.85	9.20	31-72	196	47.05	10.01	31-72
Parent-Rated DECA								
Initiative T-Score	--	--	--	--	197	46.27	7.30	28-56
Self-Control T-Score	--	--	--	--	197	54.13	9.53	28-72
Attachment T-Score	--	--	--	--	197	49.45	12.12	28-72
Total Protective Factors T-Score	--	--	--	--	197	48.76	9.57	28-70
Behavioral Concerns T-Score	--	--	--	--	197	46.85	9.20	28-72
Spanish-Speaking Children Only								
Standardized Assessments								
TVIP Standard Score	48	82.69	13.94	61-131	49	88.55	14.70	60-120
WM LWI Standard Score	48	95.06	11.94	70-117	49	103.42	11.49	71-127
WM AP Standard Score	48	93.47	9.65	76-115	49	94.88	7.52	78-115

[1]Some teachers and parents left items blank on the DECA. Scores were only calculated if at least 75% of the items were present. This resulted in some missing data for the DECA.

Table 6: Weighted English Assessment Scores by Child's Primary Language, Fall Round[1]

Assessment	Primary Language						*t*
	English			Another Language			
	N	Mean	SD	N	Mean	SD	
PPVT Standard Score	149	106.04	17.20	55	63.34	19.85	15.50*
WJ LWI Standard Score	149	105.36	12.92	55	88.57	10.35	9.81*
WJ AP Standard Score	149	110.92	12.65	55	92.16	12.52	9.77*
LAS Scaled Score	144	7.45	4.55	55	7.58	3.79	.207

*p<.001.
[1]When the child's primary language was missing, the variable was populated with Spanish, if the child was tested in Spanish. This was intended to maximize sample size.

Table 7: Weighted English Assessment Scores by Child's Primary Language, Spring Round[1]

Assessment	Primary Language						*t*
	English			Another Language			
	N	Mean	SD	N	Mean	SD	
PPVT Standard Score	143	108.23	16.51	55	64.93	21.00	15.79*
WJ LWI Standard Score	143	108.10	13.15	55	91.98	13.75	7.95*
WJ AP Standard Score	143	110.50	11.19	55	92.30	12.16	10.40*
LAS Scaled Score	143	9.54	4.80	55	9.57	4.02	.031

*p<.001.
[1]Information about the child's primary language was missing or ambiguous for 2 children in the sample.

PRESCHOOL QUALITY

The 206 children in the sample were enrolled in 112 different preschools. Information regarding quality of these preschools was gleaned from two sources: a) the Qualistar Rating and Accreditation information that DPP incorporates in its calculation of the tuition credit for each child, and b) the classroom observations using the CLASS tool that were conducted specifically for this evaluation project as well as observations that were conducted for the DPP quality rating.[31]

QUALISTAR RATING AND ACCREDITATION

Within the sample, 110 of the 112 preschools were Qualistar rated. Detailed information about the quality of these preschools was provided to Clayton Early Learning from Qualistar. Figure 1 presents the breakdown of programs by star level. Sixty-two percent of programs had earned 3 stars. Nearly 30% of programs had 4 stars. No preschools earned a provisional rating and no preschools had a rating of one star, indicating that very few programs were of the lowest quality. Two preschools had not yet been rated. One program earned a 4 through NAEYC accreditation.

[31] It is important to keep in mind that all of the preschool quality information provided here is based on only a sample of preschools where the children in the study sample were enrolled. For information on the quality of all preschool programs participating in DPP during the 13-14 school year, readers are referred to the annual evaluation report prepared by Augenblick, Palaich and Associates.

Figure 1: Star Level of Programs Attended by Children in the Sample (n=112 programs)

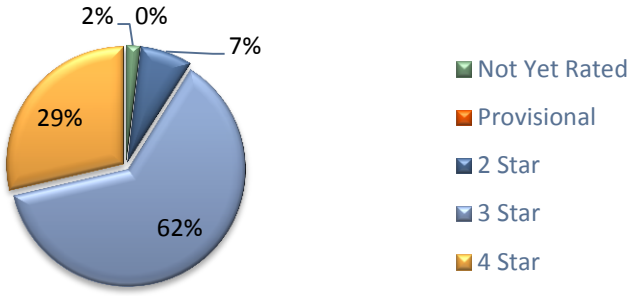


Table 8 presents descriptive statistics for the five component areas of the Qualistar rating for the 89 sites with expanded Qualistar rating data available.[32] Sites were strongest, on average, in the areas of Family Partnerships and Adult-to-Child Ratios and Group Size. Family Partnerships was a particularly strong area, with programs earning, on average, 90% of the possible points for this area. While scores in this area covered a wide range (0-10), very few programs earned very low scores on this component. Three programs earned no points for this area. Three other programs earned 4 points. The remainder earned between 8 and 10 points. For Adult-to-Child Ratios and Group Size, the average of the programs was relatively high, but there was still some variability around that mean, with scores ranging between 4 and 10. On average, programs earned about 60% of the possible points for Training and Education. There was considerable variability around this mean with some programs earning very few points and half earning 6 or 7 of the 10 points possible. Programs earned, on average, about two-thirds of the possible points for Learning Environment. Scores in this area ranged somewhat, with some programs earning 4 points and 3 earning all 10 of the possible points. Two programs earned the 2 points for having earned an accreditation.

Table 8: Qualistar Rating Components for Programs Attended by Children in the Sample (n=89 programs)

Component	Possible Range	Mean	SD	Range
Learning Environment	0-10	6.58	1.25	4-10
Family Partnerships	0-10	9.21	2.12	0-10
Training and Education	0-10	6.31	1.47	3-10
Adult-to-Child Ratios and Group Size	0-10	8.85	1.32	4-10
Accreditation	0-2	.04	0.30	0-2
Accreditation	0-2	.04	0.30	0-2

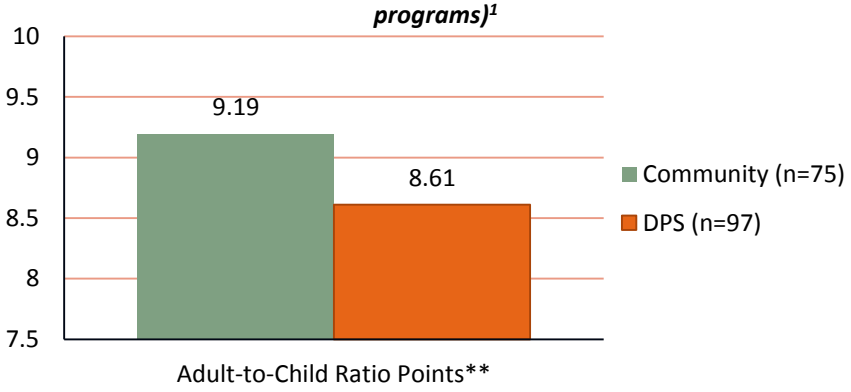
Analyses were conducted to test whether the type of provider (DPS vs. Community) was associated with the components of the Qualistar rating. The two types of programs only differed significantly in one area: Adult-to-Child Ratios (see Figure 2).[33] On average, community programs earned significantly more points than community preschools in this area.

[32] More information about the five component areas of the Qualistar rating is available at: http://www.qualistar.org/qualistar-rating-components.html. Data were provided for only 89 of our 110 rated sites.

[33] t=2.82, df=170, p<.01

Analyses were conducted to test whether any of the child and family background characteristics were associated with Total Qualistar Rating Points. Total Rating Points was not associated with tier level, region of the city, child primary language, home language or ethnicity.

Figure 2: Qualistar Rating Adult-to-Child Points for Programs Attended by Children in the Sample, by Provider Type (n=89) programs)[1]

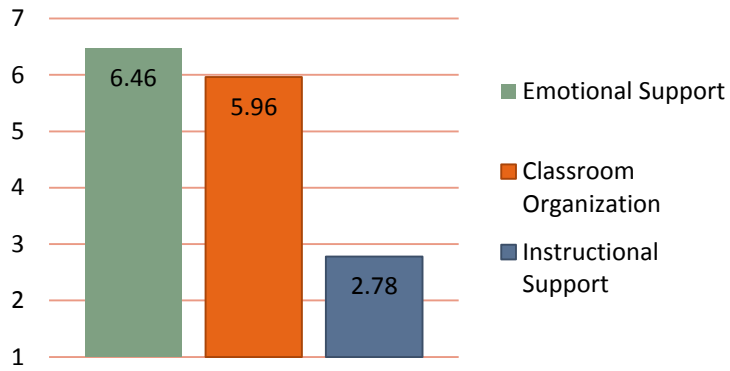


**p<.01
[1]Standard deviations: Community1.43 DPS=1.25

CLASS OBSERVATIONS

Figure 3 displays the mean scores for the 129 classrooms that were observed using the CLASS Observation. On average, scores for Emotional Support and Classroom Organization were high, while scores for Instructional Support were near the bottom of the middle-range. Average scores for Emotional Support and Classroom Organization were slightly higher than average scores from previous large studies. As described above, in previous large studies using this observation tool, average scores for Emotional Support tended to be in the 4.5-5.5 range and average scores for Classroom Organization tended to be in the 4.5-5 point range. Scores for Instructional Support were similar to what has been observed in previous large studies, which have been in the 2-3 range.

Figure 3: Average CLASS Domain Scores (n=129 Classrooms)



Figures 4, 5 and 6 provide information about the variability in these domain scores. For Emotional Support, the vast majority of classrooms scored in the high range (scores above 5) and the remainder scored in the middle-range (scores between 3 and 5). For Classroom Organization, a little less than two-thirds of classrooms scored in the high range, no classrooms scored in the low range (below 3), and the remainder scored in the middle-range. For Instructional Support, about two-thirds of classrooms scored in the low range, slightly more than a third scored in the middle range, and no classrooms scored in the high range.

Figure 4: Distribution of Scores for Emotional Support (n=129 Classrooms)

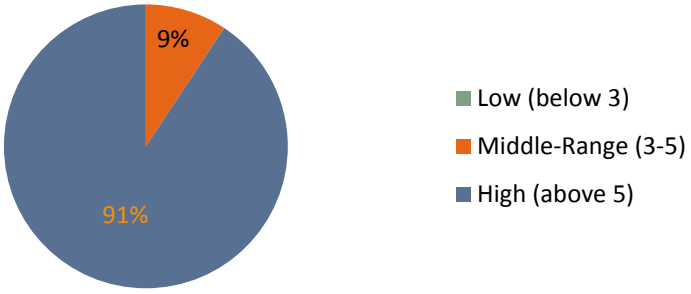


Figure 5: Distribution of Scores for Classroom Organization (n=129 Classrooms)

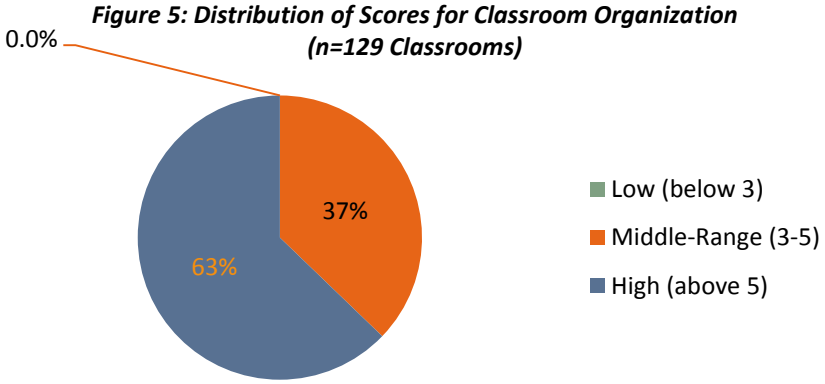
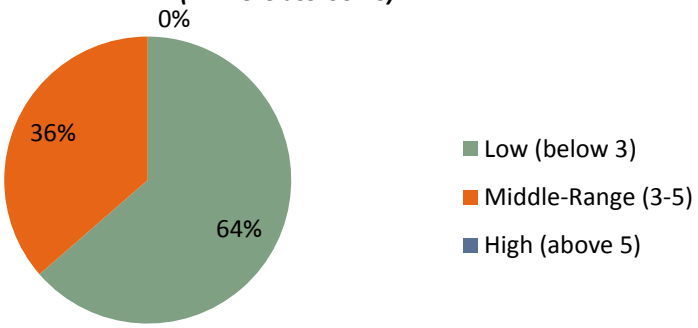


Figure 6: Distribution of Scores for Instructional Support (n=129 Classrooms)



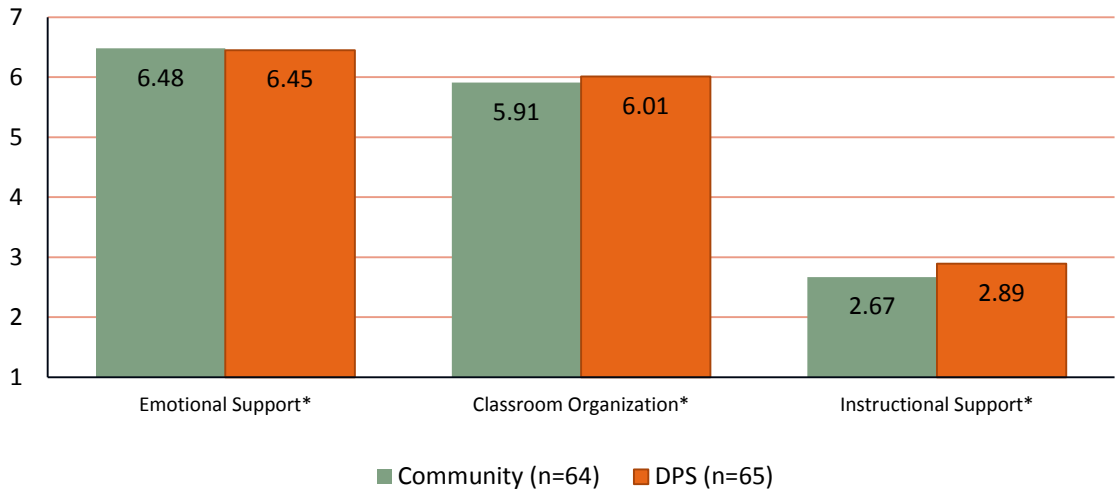
We also conducted analyses to test for differences in CLASS domain scores by provider type. The results of these analyses are presented in Figure 7. Scores for all CLASS subscales were not statistically different, on average, for DPS classrooms compared with community-based preschool classrooms.[34] This is a change from the previous year's analysis in which the two groups' mean ratings were statistically significantly different.

[34] Emotional Support—t=.504, df=127, p=n.s.; Classroom Organization—t=.913, df=127, p=n.s.; Instructional Support—t=1.52, df=127, p=n.s.

The ratio of children to adults during our observations varied widely. On average, classrooms had 6.4 children for every adult in the classroom. The smallest observed ratio was 2.0 children for every adult and the largest ratio was 13 children for every adult. We examined whether child to adult ratio during the observation was significantly associated with scores on the CLASS observation. These correlations were non-significant.[35]

As with the Qualistar Rating, we conducted analyses to test for associations between CLASS domain scores and child and family background characteristics. No significant associations were observed.

Figure 7: CLASS Domain Scores by Provider Type



*p=n.s.
^Standard deviations: Emotional Support—Community=0.45, DPS=0.37; Classroom Organization—Community=0.64, DPS=0.64; Instructional Support—Community=0.85, DPS=0.75

KINDERGARTEN READINESS

STANDARDIZED ASSESSMENTS

Analyses were conducted to determine how ready for kindergarten DPP participants appeared to be at the end of their preschool year. Readiness was examined in two ways. First, we examined whether children scored in the average range as defined by the tests' publishers, namely a standard score of 85 or above. A score of 85 or above can be interpreted as not 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 children meet a higher standard, defined as scoring at or above 100, the population mean, on the assessments used in the study. Figure 8 presents the percent of children scoring 85 or above and 100 or above on each of the assessments at the spring time point. In the general population, one would expect about 84% of children to score above 85 and 50% of children to score above 100.

[35] Ratio with Emotional Support, r=-.07; ratio with Classroom Organization, r=-.02; ratio with Instructional Support, r=-.09; all non-significant.

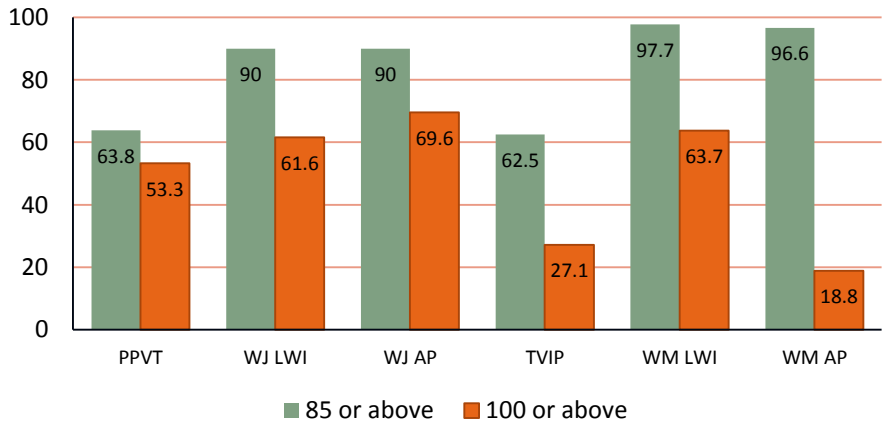
For the English assessments, the vast majority of children (over 95%) scored 85 or above on the WJ LWI and WJ Applied Problems assessments. About two-thirds of children scored 100 or above on WJ LWI and WJ Applied Problems. In contrast, only about 80% of children earned a score of 85 or above on the PPVT. Nearly half of the children scored 100 or above.

Not surprisingly, follow-up analyses revealed that the likelihood of scoring 85 or above on these assessments was strongly associated with children's primary language (with similar findings for home language). Over 88% of children whose primary language was English scored 85 or above on the PPVT as compared with 14% of children whose primary language was not English.[36] A different pattern was observed for WJ LWI and AP. Nearly all children (98%) whose primary language was English scored 85 or above on WJ LWI as compared with 73% of children whose primary language was not English.[37] Nearly all children whose primary language was English (98%) scored 85 or above on WJ AP compared with 73% of children with another primary language.[38]

A more pronounced pattern of results emerged when a score of 100 was used as the cutoff. For PPVT, three-quarters children whose primary language was English (74%) earned a score of 100 or greater as compared with just 11% of children with another primary language.[39] For WJ LWI, 76% of children whose primary language was English scored 100 or greater as compared with 30% of children with another primary language.[40] Finally, for WJ Applied Problems, 87% of children whose primary language was English earned scores of 100 or above compared with 30% of children whose primary language was something other than English.[41]

For assessments administered in Spanish, scores were once again stronger for LWI and Applied Problems than for vocabulary (TVIP). Almost 98% of children scored 85 or above on WM LWI and 97% of children scored 85 or above on Applied Problems. In contrast, just 63% of children scored 85 or above on the TVIP (see Figure 8). About a quarter of children scored 100 or above on the TVIP, about 64% scored 100 or above on the WM LWI, and less than 20% of all children scored 100 or above on WM Applied Problems. It is important to keep in mind that all of these assessments were normed with children learning only one language. Language development for children learning two languages is expected to progress at a different pace than for

Figure 8: Weighted Percent of Children Scoring in the Average Range or Above on Spring Standardized Assessments

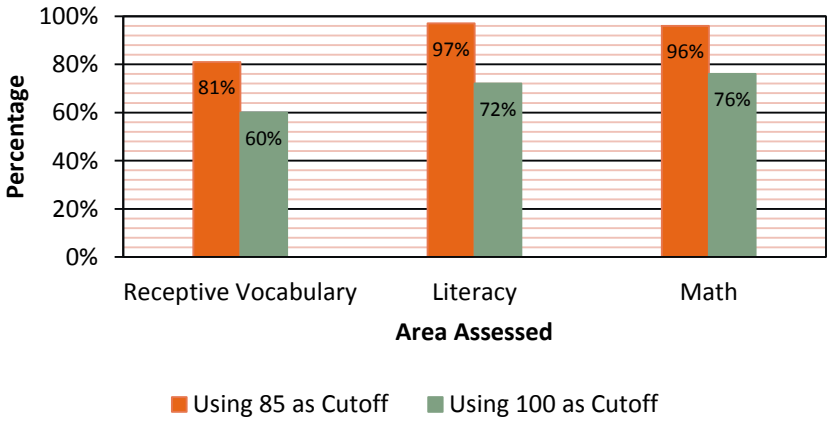


[36] χ^2_1=103.84, p<.0001
[37] χ^2_1=28.45, p<.0001
[38] χ^2_1=29.00, p<.0001
[39] χ^2_1=69.62, p<.0001
[40] χ^2_1=38.66, p<.0001
[41] χ^2_1=66.04, p<.0001

children learning one language. One way to address this issue is to jointly look at bilingual children's scores in both languages.

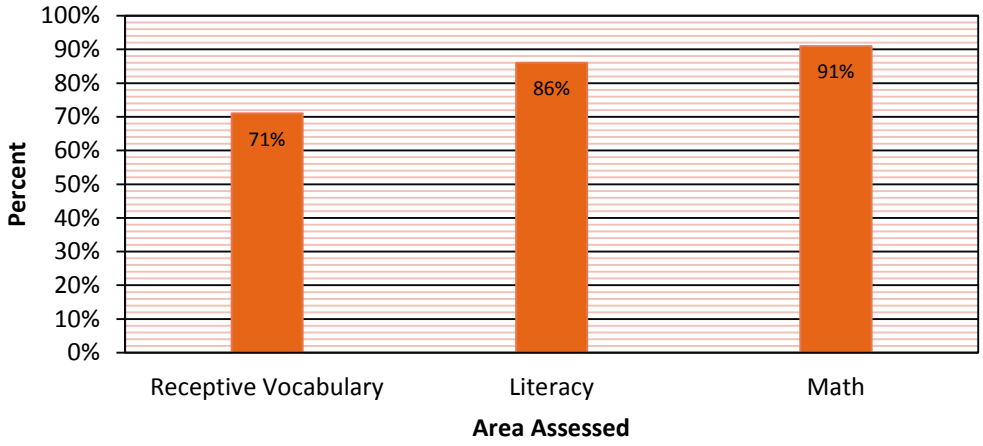
A variable was constructed to indicate whether children met or exceeded the two cutoff scores (85 and 100) in at least one language for each standardized test. Children who were bilingual could meet this criterion by meeting or exceeding the cutoff in either language. Children who were only assessed in English had only one opportunity to meet or exceed the cutoff. Results of this analysis are presented in Figure 9. A little over 80% of children met or exceeded the cutoff of 85 in at least one language in the area of receptive vocabulary (i.e., PPVT or TVIP). Nearly all children met or exceeded the cutoff of 85 in at least one language on the literacy assessment (WJ-LWI or WM-LWI) and the math assessment (WJ-AP or WM-AP). When a score of 100 was used as a cutoff, 60% of children met or exceeded this benchmark for vocabulary; approximately three-quarters met or exceeded this benchmark for both literacy and math.

Figure 9: Weighted Percentage of Children Scoring in the Average Range or Above on Spring Standardized Assessments in Spanish or English



When considering the analyses reported above, it is important to keep in mind the meaning of the two cutoffs used. A score of 85, one standard deviation below the mean, represents the lower bound of the "average range." Scores below 85 are quite low. In contrast, a score of 100 is the national average. As mentioned above, we would expect only half of children to score above this cutoff. DPP leadership struggled with the use of both of these cutoffs as indicators of whether children were ready for school. The cutoff of 85 was believed to be too low (i.e., that merely exceeding the threshold for being "at risk" should not constitute the definition of "ready for school"). Further, adopting the cutoff of 100 seemed too high (i.e., requiring that children be scoring "above average" seemed too stringent a criterion for defining "ready for school," as it is likely that children scoring slightly below average are ready for school). As a compromise, we considered the cutoff of one half of a standard deviation (i.e., a score of 92.5) below the mean for defining school readiness. Results using this cutoff are presented in Figure 10. In the general population, one would expect 69% of children to meet or exceed this threshold. For receptive vocabulary, nearly three-quarters of children met or exceeded this threshold in at least one language. For literacy and math, about 90% of children met or exceeded this threshold.

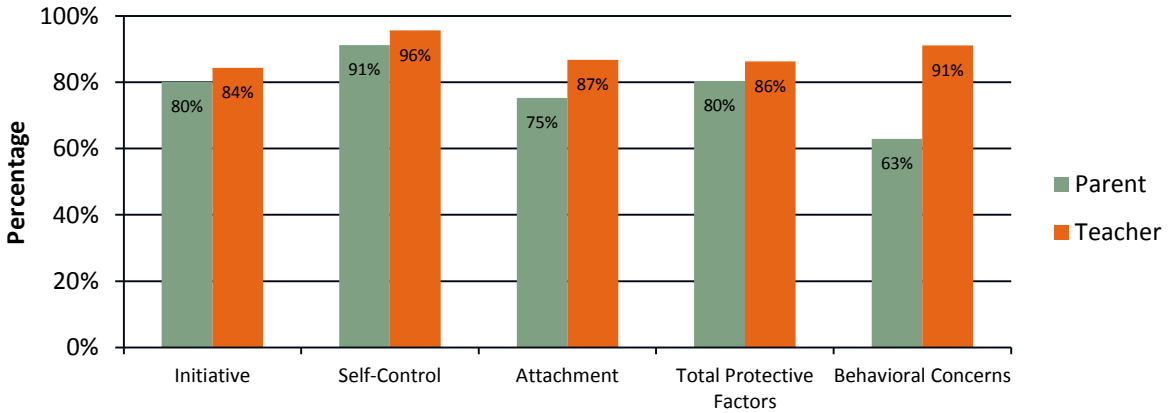
Figure 10: Weighted Percent of Children Scoring 92.5 or Above on Spring Standardized Assessments in Spanish or English



PARENT AND TEACHER SURVEYS

For the DECA, readiness is defined as being in the "Typical" or "Strength" categories as defined by the publisher. For Protective Factors, children with T-scores greater than 40 fall into these categories. For Behavioral Concerns, higher scores indicate greater levels of behavioral concerns, so children with T-scores below 60 are considered in the "Typical" range. In the general population, one would expect about 84% of children to fall within these ranges. As displayed in Figure 11, according to parents, a majority of children were in the typical or strength range for Initiative, Self-Control and Total Protective Factors (a combination of Initiative, Self-Control and Attachment). Parents rated about three quarters of children in the typical or strength range for Attachment and a little less than two-thirds in the typical range for Behavioral Concerns. Teachers also rated a majority of children in the typical or strength range for Initiative, Self-Control, Attachment and Total Protective Factors and over 90% in the typical range on Behavioral Concerns.

Figure 11: Weighted Percentage of Children Scoring in the Average Range or Above on Spring Parent and Teacher DECA Surveys.



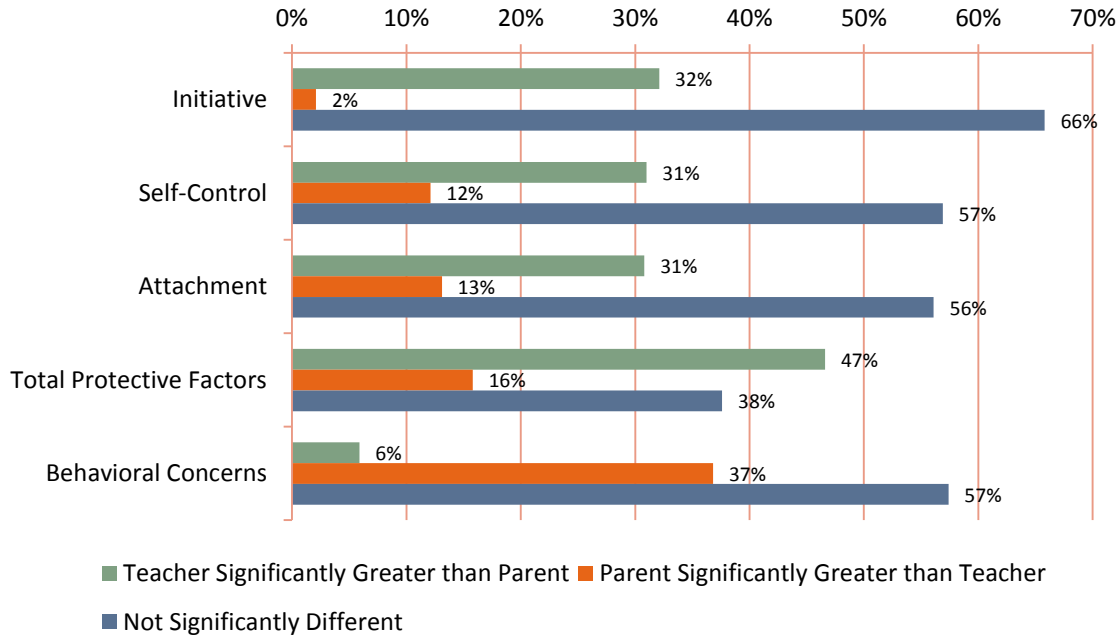
We examined the differences between teachers' and parents' ratings using guidelines from the authors of the DECA. The authors developed these guidelines to help users distinguish between differences in scores due to measurement error and differences that are likely due to a meaningful difference between scores. For Initiative, a difference of 10 is needed to conclude that there is a significant difference between the parent and teacher rating. The average difference between teacher and parent reports, 5.2 (*SD*=8.4), did not exceed this threshold, indicating that, on average, teacher and parent ratings did not differ. As displayed in Figure 12, for two-thirds of children, teacher and parent reports did not significantly differ. For about a third of the sample, the teacher rating was significantly greater than the parent rating. For only 2% of the sample, the parent rating was significantly greater than teacher rating.

For Self-Control, a difference of 10 is needed to conclude that there is a significant difference between the parent and teacher rating. The average difference between teachers' and parents' reports, 4.2 (*SD*=11.4), did not exceed this threshold, indicating that, on average, teacher and parent ratings did not differ. As displayed in Figure 12, for over half of children, teacher and parent reports did not significantly differ. For about a third of children, teachers rated children significantly higher than did parents. About 12% were rated higher by their parents than their teachers.

For Attachment, a difference between the teacher's and parent's score of 12 is needed to conclude that the scores are significantly different. On average, the difference between the parents' and teachers' scores, 3.6 (*SD*=14.7), did not exceed this threshold. Once again, over half of the children, the parent and teacher scores did not significantly differ. For those where the difference was significant, it was more common for the teacher's score to be higher than it was for the parent's score to be higher.

For Total Protective Factors, a difference of 7 points is needed to conclude that there is a significant difference between the parents' and teachers' ratings. Across the sample, the average difference between ratings for Protective Factors was 5.3 (*SD*=11.0), which was below that threshold. However, for 47% of children, the teacher's rating was significantly greater than the parent's rating (see Figure 12). For over a third of children, there was not a significant difference between raters. For 16% of children, the parent's rating was significantly higher than the teacher's rating.

Figure 12: Comparison of Parent and Teacher DECA Surveys, Weighted



For Behavioral Concerns, a difference of 14 points is needed to conclude that there is a significant difference between the parent's and teacher's ratings. The average difference in the sample was 8.7 (*SD*=11.9), which did not reach this threshold. For nearly 60% of children, there was not a significant difference between the parent's and teacher's rating (see Figure 12). For the remaining children, it was significantly more common for the parent to report more Behavioral Concerns than vice versa.

In sum, for Initiative, Self-Control and Attachment, teachers and parents made similar ratings for over half of the sample. For the other half, teachers rated children significantly higher than did parents more often than parents rated children significantly higher than teachers. For nearly half of the sample, teachers rated children significantly higher than parents on Total Protective Factors. For Behavioral Concerns, parents and teachers generally agreed; but when they differed, parents rated the child significantly higher than the teacher.

CHANGE IN ASSESSMENT SCORES OVER THE PRESCHOOL YEAR

A series of paired *t*-tests was conducted to test for change over time in standardized assessments in English and Spanish as well as teacher-rated DECAs. Results are presented in Table 9. There were significant increases in PPVT and in WJ LWI (small in magnitude, about a tenth of a standard deviation). For Applied Problems, there was no significant change. It is important to keep in mind that these scores are adjusted for age, so when increases are observed, they are above and beyond what one would expect due to typical maturation. The Leiter Attention Sustained scale was a criterion-referenced test (not adjusted for age), but rather uses a cutoff (criterion) for indicating adaptive levels of attention. A scaled score of seven or below may indicate attentional difficulties, and nationally, a scaled score of 10 is average. A statistically significant increase by spring was observed, bringing children to the national average. For assessments administered in Spanish, there was a moderate increase in TVIP scores over the course of the school year, over a third of standard deviation. No difference was observed for WM LWI and WM AP from fall to spring.

Table 9: Change in Child Outcome Variables Over the Course of the Preschool Year

Variable	N	Fall Mean (*SD*)	Spring Mean (*SD*)	*t*
Standardized Assessments—English				
PPVT	194	91.99 (27.19)	93.98 (27.12)	2.29*
WJ-LWI	194	99.89 (14.58)	102.62 (14.79)	4.03***
WJ-AP	194	104.71 (15.62)	104.53 (14.24)	.26
Leiter AS	194	7.45 (4.31)	9.97 (4.48)	6.51***
Standardized Assessments—Spanish				
TVIP	48	82.69 (13.94)	88.76 (14.67)	3.16**
WM-LWI	45	88.23 (27.31)	91.62 (11.77)	.89
WM-AP	48	93.47 (9.65)	91.63 (10.86)	1.34
Teacher Survey				
Initiative T-Score[1]	189	49.34 (8.06)	51.47 (7.48)	4.32***
Self-Control T-Score	189	57.20 (9.27)	58.21 (7.98)	1.79
Attachment T-Score	189	50.77 (9.29)	52.97 (9.23)	3.55***
Total Protective Factors T-Score	189	51.81 (9.18)	53.99 (8.58)	3.83***
Behavioral Concerns T-Score	189	46.80 (9.04)	47.02 (10.05)	.37

*p<.05, **p<.01, ***p<.001

[1]Some teachers and parents left items blank on the DECA. Scores were only calculated if at least 75% of the items were present. This resulted in some missing data for the DECA.

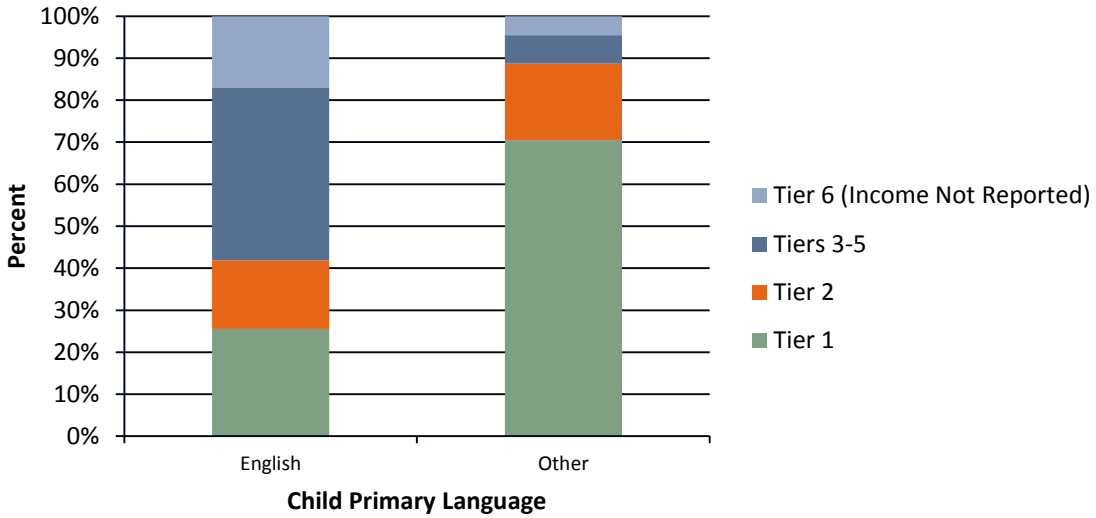
Significant improvements were also observed in three of the teachers' ratings on the DECA over the course of the school year. Change over time was significant and positive for the Protective Factors (Initiative and Attachment, as well as Total Protective Factors). No significant decreases in Behavioral Concerns over the course of the school year were reported.

CHANGE OVER TIME BY SUBGROUP

Further analyses were conducted to test whether the extent of the change over time varied by two background characteristics: income tier and children's primary language. Prior to conducting analyses by income tier, some data reduction was necessary since the number of participants from some of the income tiers was rather small (see Table 2). Income tier was collapsed into a new income tier group variable with 4 categories: Tier 1, Tier 2, Tiers 3-5 and Tier 6 (i.e., parents who opted out of the requirement to report income and instead elected to

automatically be assigned to the lowest tuition credit level).[42] It is important to note that these two background characteristics, income tier and child's primary language, are strongly associated (see Figure 13).[43] Nearly all children whose primary language is not English are from Tiers 1 or 2 whereas only about 42% of the children whose primary language is English are from these lowest two tiers. As a result, in this sample, it will be impossible to disentangle the effects of income and primary language and any effects observed are possibly the result of the co-occurrence of these two factors.

Figure 13: Income Tier Groups, by Child Primary Language



INCOME TIER

A series of Repeated Measures ANOVAs[44] was conducted with income tier group predicting scores over time on assessments administered in English and Spanish as well as teacher-rated DECA. Unlike last year's findings, there were no significant interactions (i.e., tiers didn't change at different rates on average) between income tier group and time for any of the scales (PPVT, WJ LWI, WJ Applied Problems, Leiter AS, WM LWI, WM Applied Problems, or TVIP), including the teacher-rated DECA subscales. In all cases, the income tier by time interaction was non-significant, indicating that children progressed similarly in these areas over the course of their preschool year, regardless of their income tier.

[42] For analyses of assessments administered in Spanish, a two-level income tier group variable was used omitting the category 'tiers 3-5' and 'tier 6' because only one child assessed in Spanish fell into tiers 3-5 and three children assessed in Spanish fell into tier 6.

[43] χ^2_3=39.88, p<.0001

[44] ANOVA (Analysis of Variance) is a statistical technique that compares mean scores for specified groups. Repeated Measures ANOVAs take into account scores at multiple points in time. This analysis compares the amount of change over time for specified groups.

CHILDREN'S PRIMARY LANGUAGE

A series of Repeated Measures ANOVAs was conducted with primary language predicting scores over time on assessments administered in English and teacher-rated DECA.[45] As with income tier, there were no significant interactions between primary language group and time for any of the assessments including the teacher-rated DECA, meaning that the rate of change by language did not differ by language.

PRESCHOOL QUALITY AND CHILD OUTCOMES

Because of the lack of variability in Qualistar data, we focused on the CLASS Observation data when examining the association between preschool quality and child outcomes. In addition, since there was very little variability in the Emotional Support domain (see Figure 4), we restricted our focus to Classroom Organization and Instructional Support. To examine the association between quality and child outcomes we computed partial correlations between spring assessment scores and CLASS domain scores, controlling for fall assessment scores. These analyses, while not specifically focused on change over time (i.e., the actual difference between fall and spring scores), examine "residualized gain," which can be understood as how children score in the spring after taking into account the differences between them in the fall.

For the English academic assessments (vocabulary, literacy, math, and sustained attention), there were no significant correlations with classroom quality. There were also no significant correlations for Spanish assessments. Similarly, there were no associations observed for the spring teacher DECA ratings and any of the CLASS domains after controlling for fall ratings.

RESULTS: ELEMENTARY SCHOOL

For the 2013-2014 analysis, the results below present data from following Cohorts 2-5 into their elementary years. Cohort 2 was in Kindergarten and Cohort 5 was in 3rd grade in 2013-2014. Cohort 1 DPP students were in 4th Grade during the 2013-2014 school year and elementary data were not provided, as the study follows children through the 3rd grade.

HOW SIMILAR ARE DPP GRADUATES TO THE POPULATION OF CHILDREN IN THE DISTRICT AS A WHOLE?

Prior to making comparisons between reading scores for DPP graduates and DPS as a whole, it is important to consider whether the samples of DPP graduates are similar demographically to the district as a whole. The Colorado Department of Education (CDE) provides demographic data on school districts in Colorado in the fall of each school year.[46] CDE provides information about free or reduced lunch status for the district as a whole as well as gender and race/ethnicity for each grade level.

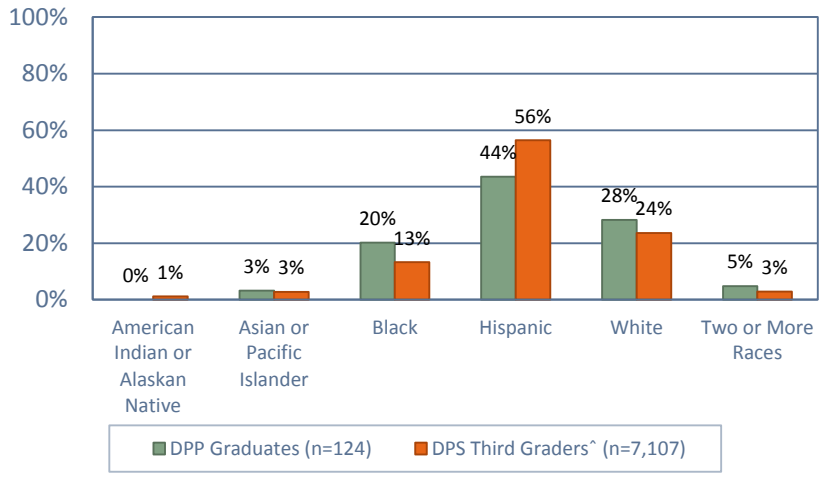
COHORT 2

[45] It does not make sense to conduct this set of analyses for assessments administered in Spanish, since there is not adequate variability in children's primary language among children assessed in Spanish.

[46] Available at: http://www.cde.state.co.us/cdereval/pupilcurrentschool

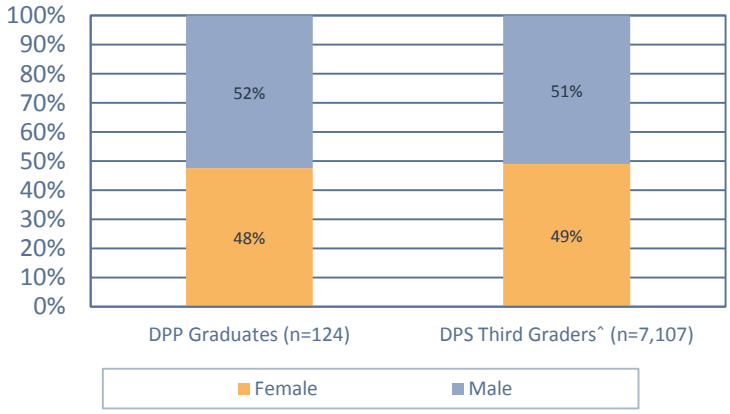
Figure 14 displays a comparison of the racial/ethnic composition of the sample of Cohort 2 DPP graduates with TCAP reading assessment data for school year 13-14 and the population of children enrolled in third grade as of fall 2013. The sample was fairly similar to the district as a whole in terms of race and ethnicity.

Figure 14: Race/Ethnicity of Cohort 2 DPP Graduates and All Third Graders Enrolled in DPS in Fall 2013*



*This figure includes only DPP Graduates for whom TCAP reading data were available for the 13-14 school year.
^The group of DPS Third Graders includes the DPP graduates. Figures may not sum to 100 because of rounding error.

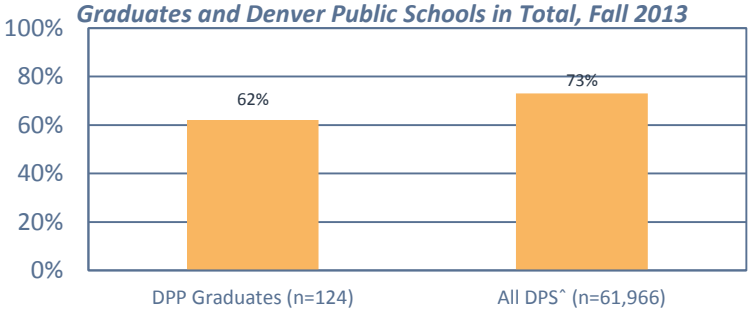
Figure 15: Gender of Cohort 2 DPP Graduates and All Third Graders Enrolled in DPS in Fall 2013*



*This chart includes only Cohort 2 DPP Graduates for whom TCAP reading assessment data were available for the 12-13 school year.
^The group of DPS Third Graders includes the DPP graduates.

Figure 15 presents the gender composition of Cohort 2 DPP graduates with reading assessment data and all third graders in DPS as of fall 2013. The district as a whole was split evenly between the genders. Cohort 2 DPP Graduates with spring 2014 reading assessment data were split a bit less evenly, with slightly more girls than boys, but still rather close to an even split.

Figure 16: Free/Reduced Lunch Status of Cohort 2 DPP Graduates and Denver Public Schools in Total, Fall 2013



ˆThe group All DPS includes the DPP graduates.

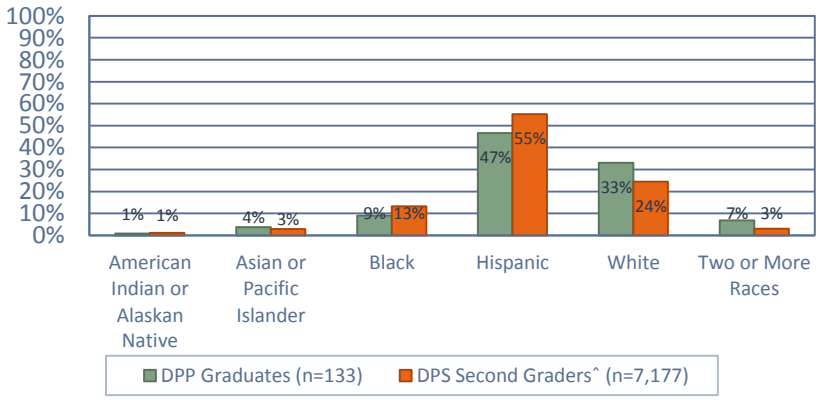
Finally, Figure 16 presents the proportion of children eligible for free or reduced lunch for the district as a whole and for the sample of DPP graduates. CDE does not provide free or reduced lunch data by grade level. As a result the comparison group in this figure is the entire district, from kindergarten through grade 12. Nearly three-quarters of the district as a whole qualified for free or reduced lunch. In contrast, 62% of Cohort 2 graduates with reading assessment data qualified, suggesting that the DPP Cohort 2 sample is slightly wealthier than the district as a whole.

In sum, Cohort 2 children with spring 2014 TCAP reading assessment data were slightly wealthier than the district as a whole, but were similar to the district in terms of their ethnic and gender composition.

COHORT 3

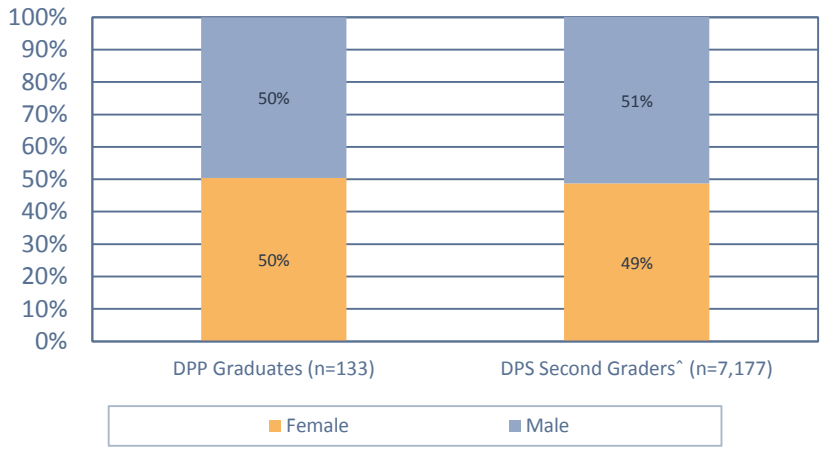
Figure 17 displays a comparison of the racial/ethnic composition of the sample of Cohort 3 DPP graduates with reading assessment data and the population of children enrolled in second grade in DPS as of fall 2013. The sample of DPP graduates includes slightly more white children, with fewer black children and Hispanic children than the district as a whole. Otherwise, the racial and ethnic compositions of the two groups are similar.

*Figure 17: Race/Ethnicity of Cohort 3 DPP Graduates and All Second Graders Enrolled in DPS in Fall 2013**



*This figure includes only Cohort3 DPP Graduates for whom reading assessment data were available for the 13-14 school year.
ˆThe group of DPS Second Graders includes the DPP graduates. Percentages may not sum to 100 because of rounding error.

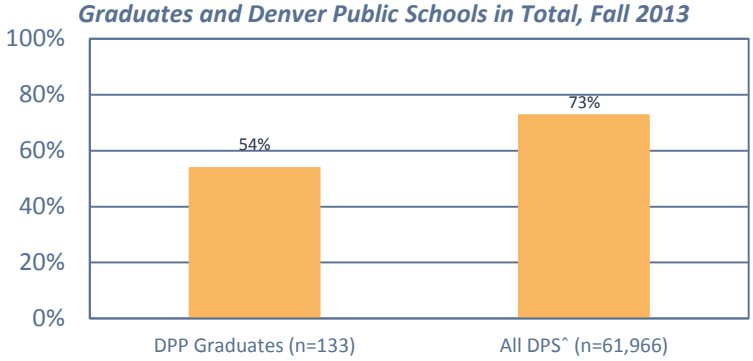
*Figure 18: Gender of Cohort 3 DPP Graduates and All Second Graders Enrolled in DPS in Fall 2013**



*This figure includes only Cohort 3 DPP Graduates for whom reading assessment data were available for the 13-14 school year.
^The group of DPS Second Graders includes the DPP graduates.

Figure 18 presents the gender composition of Cohort 3 DPP graduates with reading assessment data and all second graders in DPS as of fall 2013. The district as a whole and the DPP Cohort 3 sample with spring 2014 reading data were split evenly between the genders.

Figure 19: Free/Reduced Lunch Status of Cohort 3 DPP Graduates and Denver Public Schools in Total, Fall 2013



^The group All DPS includes the DPP graduates.

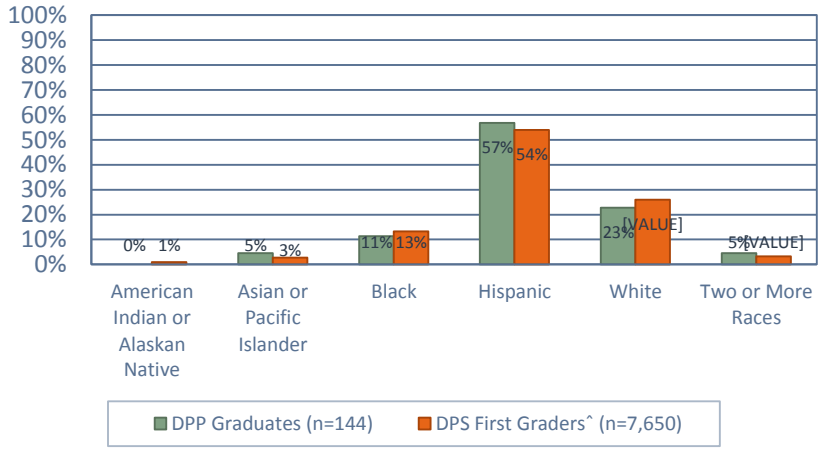
Finally, Figure 19 presents the proportion of children eligible for free or reduced lunch for the district as a whole and for the sample of DPP graduates. Nearly three-quarters of the district as a whole qualified for free or reduced lunch. In contrast, only 54% of Cohort 3 graduates with reading assessment data qualified, suggesting that the DPP Cohort 3 sample is wealthier than the district as a whole.

In sum, Cohort 3 children with spring 2014 reading assessment data differed somewhat from the district as a whole in terms of their racial and ethnic composition. A smaller proportion of children in Cohort 3 qualified for free or reduced lunch than in the district as whole, suggesting that this sample might be slightly wealthier. Any differences observed between DPP children and the district as a whole may due to the DPP program, factors related to these differences in income, or other unmeasured factors.

COHORT 4

Figure 20 displays a comparison of the racial/ethnic composition of the sample of Cohort 4 DPP graduates with reading assessment data and the population of children enrolled in first grade in DPS as of fall 2013. The racial and ethnic compositions of the two groups are remarkably similar.

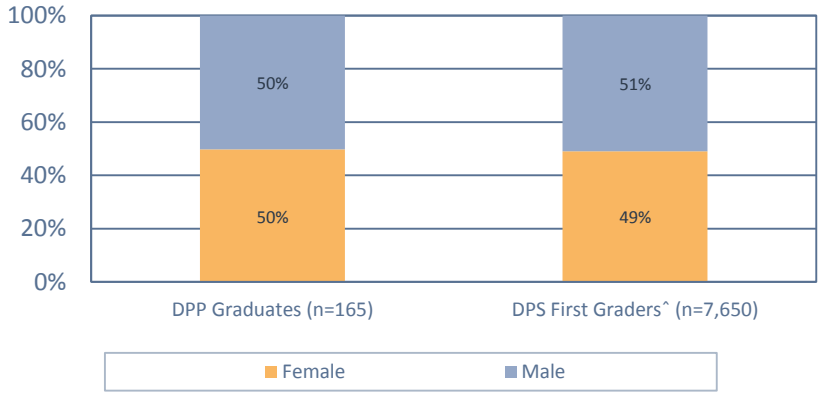
Figure 20: Race/Ethnicity of Cohort 4 DPP Graduates and All First Graders Enrolled in DPS in Fall 2013*



*This figure includes only Cohort 4 DPP Graduates for whom reading assessment data were available for the 13-14 school year. Percentages for DPP Graduates are weighted.
^The group of DPS First Graders includes the DPP graduates.

Figure 21 presents the gender composition of Cohort 4 DPP graduates with reading assessment data and all first graders in DPS as of fall 2013. The district as a whole was split very close to evenly between the genders, with slightly more males than females. Cohort 4 DPP Graduates with spring 2014 reading assessment data were also split quite evenly.

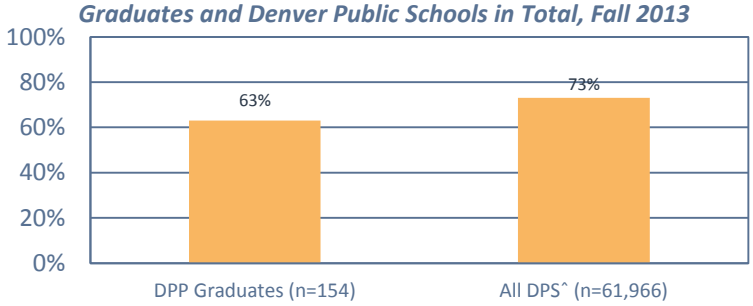
Figure 21: Gender of Cohort 4 DPP Graduates and All First Graders Enrolled in DPS in Fall 2013*



*This figure includes only Cohort 4 DPP Graduates for whom reading assessment data were available for the 13-14 school year.
^The group of DPS First Graders includes the DPP graduates.

Figure 22 presents the proportion of children eligible for free or reduced lunch for the district as a whole and for the sample of DPP graduates. Over two-thirds of Cohort 4 graduates with reading assessment data qualified for free or reduced lunch, which was less than the district as a whole, which approached three-fourths qualifying for free or reduced lunch.

Figure 22: Free/Reduced Lunch Status of Cohort 4 DPP Graduates and Denver Public Schools in Total, Fall 2013



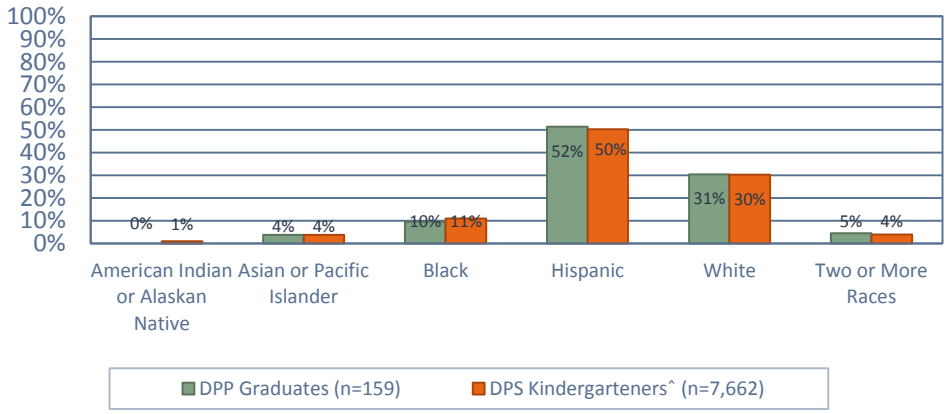
ˆThe group All DPS includes the DPP graduates.

In sum, Cohort 4 children with spring 2014 reading assessment data were similar to the district in terms of their ethnic and gender composition. A smaller proportion of children in Cohort 3 qualified for free or reduced lunch than for the district as whole, suggesting that this sample might be slightly wealthier. Any differences observed between DPP children and the district as a whole may due to the DPP program, factors related to these differences in income, or other unmeasured factors.

COHORT 5

Figure 23 displays a comparison of the racial/ethnic composition of the sample of Cohort 5 DPP graduates with reading assessment data and the population of children enrolled in kindergarten in DPS as of fall 2013. The group of Cohort 5 DPP graduates has a similar racial and ethnic composition as the group of all DPS kindergarteners.

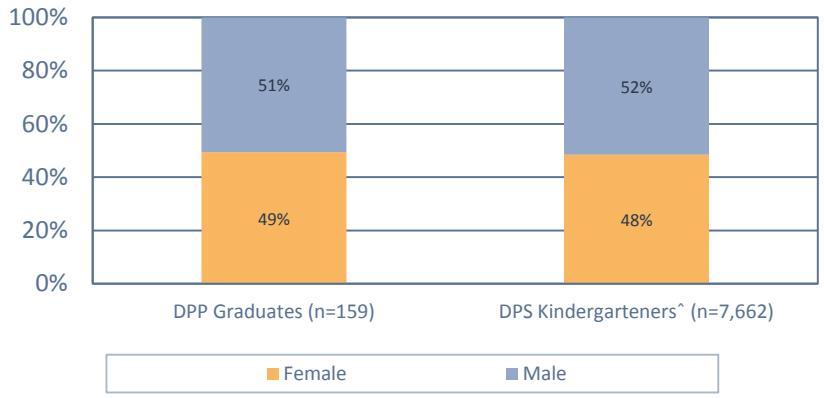
Figure 23: Race/Ethnicity of Cohort 5 DPP Graduates and All Kindergarteners Enrolled in DPS in Fall 2013*



*This figure includes only Cohort 5 DPP Graduates for whom reading assessment data were available for the 13-14 school year. Percentages for DPP Graduates are weighted. Percentages do not sum to 100 because of rounding error.
ˆThe group of DPS Kindergarteners includes the DPP graduates.

Figure 24 presents the gender composition of Cohort 5 DPP graduates with reading assessment data and all kindergarteners in DPS as of fall 2013. Both the group of Cohort 5 graduates with spring 2014 reading assessment data and the district kindergartners as a whole were split very close to evenly between the genders.

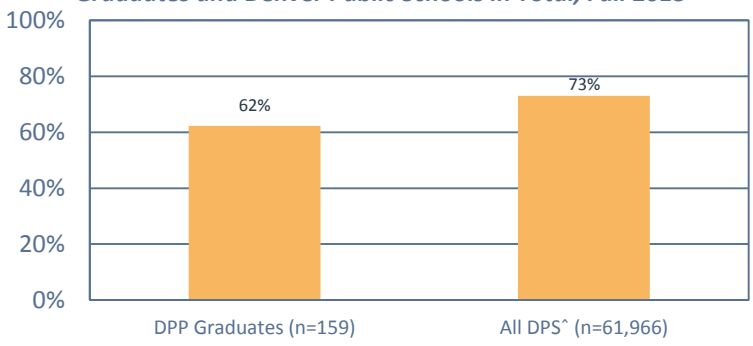
Figure 24: Gender of Cohort 5 DPP Graduates and All Kindergarteners Enrolled in DPS in Fall 2013*



*This figure includes only Cohort 5 DPP Graduates for whom reading assessment data were available for the 13-14 school year.
^The group of DPS Kindergarteners includes the DPP graduates.

Figure 25 presents the proportion of children eligible for free or reduced lunch for the district as a whole and for the sample of DPP graduates. Nearly two-thirds of Cohort 5 graduates with reading assessment data qualified for free or reduced lunch, which was lower than the district as a whole, which approached three-fourths qualifying for free or reduced lunch.

Figure 25: Free/Reduced Lunch Status of Cohort 5 DPP Graduates and Denver Public Schools in Total, Fall 2013



^The group All DPS includes the DPP graduates.

In sum, Cohort 5 children with spring 2014 reading assessment data were similar to the district in terms of their ethnic and gender composition. In terms of the proportion of children who qualified for free or reduced lunch, the DPP graduates in Cohort 5 had a slightly lower proportion than all DPS. Any differences between DPP children and the district as a whole may due to the DPP program or other unmeasured factors.

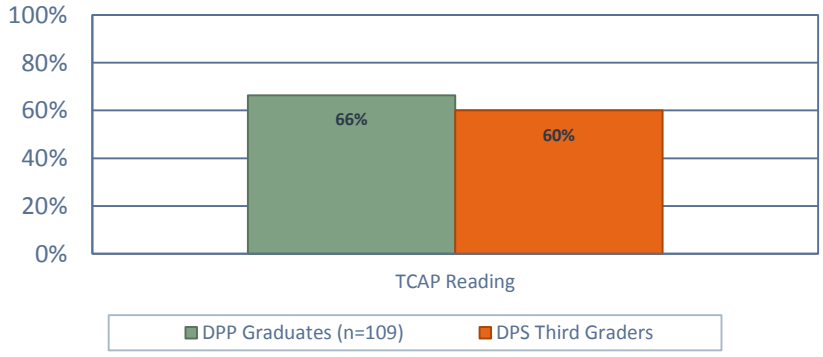
WHAT IS THE OVERALL READING PROFICIENCY OF DPP GRADUATES IN THE EARLY ELEMENTARY YEARS? HOW DOES THIS COMPARE TO THE DISTRICT AS A WHOLE?

To address this research question, we examined the proportion of DPP graduates who were reading on grade level as measured by the DRA2, EDL2, and TCAP alongside statistics for the district as a whole. We focused our analyses on children who were enrolled in the expected grades (i.e., third grade for Cohort 2, second grade for Cohort 3, first grade for Cohort 4, and kindergarten for Cohort 5). The sample of children in other grades was too small to permit analysis. It is important to keep in mind that the statistics for the district as a whole include the DPP graduates, as well as children who were enrolled in DPP but did not participate in the research study. The statistics for the district as a whole may also include children who may have been enrolled in DPP preschools but did not participate in DPP to receive tuition credits.

COHORT 2

Figure 26 displays the proportion of Cohort 2 DPP graduates who scored proficient or advanced on the third grade TCAP reading assessment. All children were assessed in English. Sixty-six percent of Cohort 2 DPP graduates scored proficient or advanced, compared with 61% of the district as a whole.

Figure 26: Proportion of Children Scoring Proficient or Advanced on Third Grade TCAP Reading, Cohort 2 DPP Graduates and All Third Graders Enrolled in DPS, Spring 2014

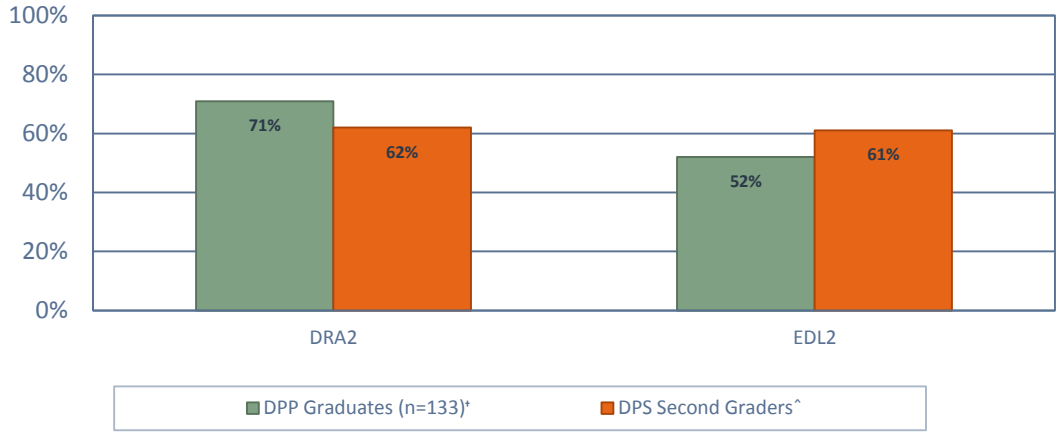


^The group of DPS third graders includes the DPP graduates.

COHORT 3

Figure 27 displays the proportion of Cohort 3 DPP graduates whose reading level was at or above grade level as assessed by the DRA2 and EDL2. This is presented alongside the reading levels for second graders in the district as a whole in spring 2013. Almost three-quarters of DPP graduates assessed in English with the DRA2 were reading at or above grade level at the end of second grade, compared with just 62% of second graders in the district as a whole. Only 20 DPP graduates were assessed using the EDL2. Of these, 52% were reading on grade level compared to 61% of the second graders assessed with EDL2 in the district as a whole.

*Figure 27: Proportion of Children Reading On Grade Level in the Spring of the Second Grade Year, Cohort 3 DPP Graduates and All Second Graders Enrolled in DPS, Spring 2014**

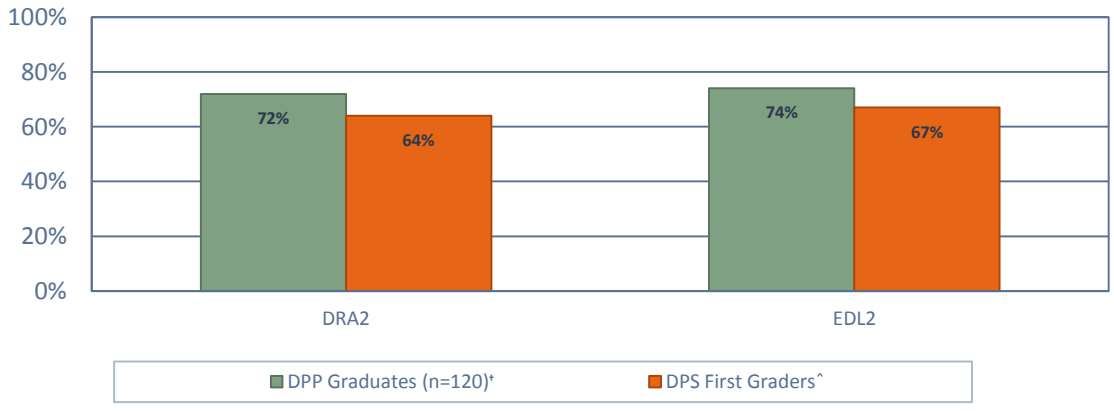


*A score of 28 is considered reading "on grade level" for the end of second grade.
†The group of DPP Graduates includes 113 children assessed with the DRA2 and 20 assessed with the EDL2.
^The group of DPS second graders includes the DPP graduates.

COHORT 4

Figure 28 displays the proportion of Cohort 4 DPP graduates whose reading level was at or above grade level as assessed by the DRA2 and EDL2. This is presented alongside the reading levels for first graders in the district as a whole in spring 2013. Nearly three-quarters of DPP graduates assessed in English with the DRA2 were reading at or above grade level at the end of first grade. This is greater than the proportion reading at or above grade level in the district as a whole (64%). Three-quarters of DPP graduates assessed in Spanish using the EDL2 were reading at or above grade level at the end of first grade. In contrast, two-thirds of children in the district as a whole were reading at or above grade level as assessed by the EDL2.

*Figure 28: Proportion of Children Reading On Grade Level in the Spring of First Grade Year, Cohort 4 DPP Graduates and All First Graders Enrolled in DPS, Spring 2014**

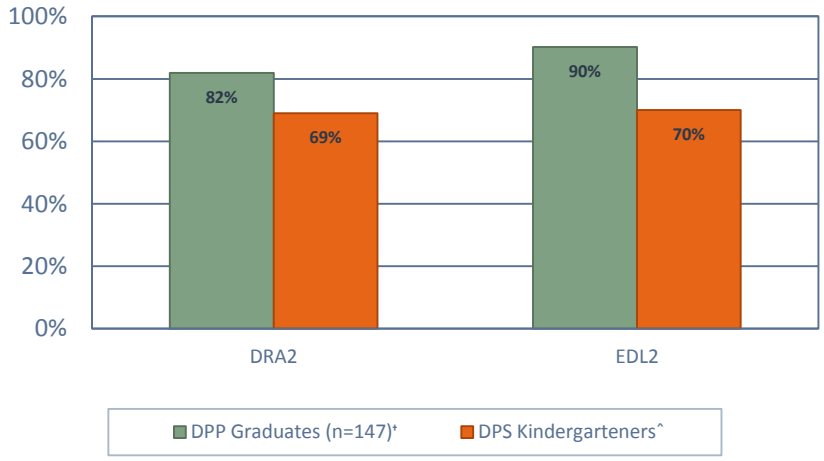


*A score of 16 is considered "on grade level" for the end of first grade.
†The group of DPP Graduates includes 120 children assessed with the DRA2 and 24 assessed with the EDL2. Analyses are weighted.
^The group of DPS First Graders includes the DPP graduates.

COHORT 5

Figure 29 displays the proportion of Cohort 5 DPP graduates whose reading level was at or above grade level as assessed by the DRA2 and EDL2. This is presented alongside the reading levels for kindergarteners in the district as a whole in spring 2014. Eight-two percent of DPP graduates assessed in English with the DRA2 were reading at or above grade level at the end of kindergarten. This exceeds the proportion reading at or above grade level in the in the district as a whole (69%). Similarly, 90% of DPP graduates assessed in Spanish using the EDL2 were reading at or above grade level at the end of kindergarten. In contrast, just 70% of children in the district as a whole were reading at or above grade level as assessed by the EDL2.

Figure 29: Proportion of Children Reading On Grade Level in the Spring of Kindergarten Year, Cohort 5 DPP Graduates and All Kindergarteners Enrolled in DPS, Spring 2014*



*A score of 4 is considered "on grade level" for the end of kindergarten.
ᵗThe group of DPP Graduates includes 119 children assessed with the DRA2 and 28 assessed with the EDL2. Analyses are weighted.
ˆThe group of DPS Kindergarteners includes the DPP graduates.

TO WHAT EXTENT IS PRESCHOOL READINESS ASSOCIATED WITH READING PROFICIENCY IN THE EARLY ELEMENTARY YEARS?

To address this question, we examined correlations between academic assessments administered in the spring of the preschool year and children's reading proficiency in the elementary school years as assessed by the DRA2 and EDL2. During the preschool year, all children were assessed in English. Those who were identified by parents and/or teachers as Spanish speakers were also assessed in Spanish. In the elementary school years, children were only assessed in one language.

COHORT 2

As with the previous question, our analysis focused only on the children who were in the expected grades. In spring 2014, 124 Cohort 2 children were enrolled in third grade and had TCAP reading assessment data. Of these, 82 were assessed only in English during the preschool year. All of these children were assessed with the English version of the TCAP at the end of the third grade year. Logistic regression models were run using preschool assessment scores to predict whether or not students scored proficient or advanced on the TCAP. PPVT-4 scores in

preschool significantly predicted third grade TCAP proficiency (χ_1^2=15.03, p<.001). A one point increase in PPVT-4 standard score in preschool is associated with an 11.2% increase in the odds of scoring proficient or advanced on third grade TCAP. Letter-Word Identification also significantly predicted third grade TCAP proficiency (χ_1^2=8.25, p=.004). For every one point increase in Letter-Word Identification standard score, there is an 8.1% increase in the odds of scoring proficient on the third grade TCAP. Finally, Applied Problems was also a significant predictor of third grade reading proficiency (χ_1^2=10.39, p=.001). For every one point increase in Applied Problems standard score, there was a 9.2% increase in the odds of scoring proficient on the third grade TCAP.

Forty-one children with third grade reading assessment data in spring 2014 had been assessed in both English and Spanish during their preschool year. All of these children were assessed with the English version of the TCAP at the end of third grade. For this group, no associations were found among the assessments conducted in English. The TVIP administered in Spanish in the spring before entering kindergarten, however, significantly predicted third grade TCAP proficiency (χ_1^2=8.30, p<.003). A one point increase in TVIP standard score in preschool is associated with a 6.6% increase in the odds of scoring proficient or advanced on third grade TCAP. In addition, the Woodcock-Muñoz Applied Problems subtask also significantly predicted third grade TCAP proficiency (χ_1^2=9.20, p<.002). A one point increase in the Applied Problems standard score in preschool is associated with a 20% increase in the odds of scoring proficient or advanced on third grade TCAP.

In sum, for children who spoke English in preschool, there was evidence for an association between kindergarten readiness and third grade reading proficiency. No strong association between kindergarten readiness, as assessed in English, and third grade reading proficiency for children who were identified as Spanish speakers in preschool. However, there was a strong association between kindergarten readiness assessed in Spanish and third grade reading proficiency among preschool Spanish speakers.

COHORT 3

In spring 2014, 148 children were enrolled in second grade and had reading assessment data. Seventy-five of these children were assessed only in English in preschool and were assessed with the DRA2 in the second grade year. Associations between the preschool assessments and second grade DRA2 scores were strong and significant. PPVT-4 scores in preschool were correlated with second grade DRA2 scores (*r*=.52, p<.0001). Letter-Word Identification was correlated with DRA2 (*r*= .51, p<.0001). Applied Problems was associated with DRA2 scores at (*r*=.32, p<.001).

Seventy-three Cohort 3 children with second grade reading assessment data in spring 2013 had been assessed in both English and Spanish during their preschool year. Forty-two of these children were assessed in English using the DRA2 at the end of second grade. Thirty-one were assessed in Spanish using the EDL2 at the end of second grade. Analyses were conducted to examine the associations of both Spanish and English preschool test scores with second grade assessment data, separately by language of assessment in second grade.

For children assessed in English in second grade, there was a significant association between DRA2 scores in second grade and the PPVT-4, WJ Letter-Word Identification, and the WJ Applied Problems administered in English. PPVT-4 was correlated .45 with DRA2 (*p*<.01), Letter-Word Identification was correlated .45 with DRA2 (*p*<.01), and Applied Problems was correlated at .46 (*p*<.01). DRA2 scores in second grade were also significantly associated with the WM Applied Problems assessment conducted in Spanish (*r*=.41, *p*=.02).

For children assessed in Spanish in second grade, EDL2 scores were not significantly correlated with Letter-Word Identification administered in either English or Spanish, nor were they correlated with the TVIP assessed in

Spanish. Significant associations were found, however, between EDL2 and the PPVT-4 (English: r=.57, p<.01). In addition, significant correlations were found for the Applied Problems assessed in both languages during preschool (English: r=62. P<.01; Spanish: r=48, P<.05).

In sum, for Cohort 3, there was a strong pattern of associations for children whose primary language was English. For these children, there is strong evidence that kindergarten readiness at the end of preschool was strongly associated with reading assessment data at the end of second grade. For developing Spanish-English bilingual children, the pattern of results showed several statistically significant associations. For children who were assessed in English at the end of the second grade year, presumably children who were judged by their teachers to have stronger English skills, there were significant associations between language, literacy, and early math skills assessed in English at the end of the preschool year and children's reading scores at the end of second grade. In addition, the children assessed in Spanish using the WM Applied Problems tended to also do well on the English DRA2. For children who were assessed in Spanish at the end of the second grade year, English vocabulary and early math skills assessed in both English and Spanish at the end of the preschool year were associated with reading skills assessed in Spanish at the end of the second grade year.

COHORT 4

In spring 2014, 142 children were enrolled in first grade and had reading assessment data. Ninety-two of these children were assessed only in English in preschool and were assessed with the DRA2 in the first grade year. Associations between the preschool assessments and first grade DRA2 scores were strong and significant. PPVT-4 scores in preschool were correlated with first grade DRA2 scores at .53 (p<.0001). Letter-Word Identification was correlated with DRA2 at .46 (p<.0001). Applied Problems was associated with DRA2 scores at .47 (p<.001).

Fifty Cohort 4 children with first grade reading assessment data in spring 2014 had been assessed in both English and Spanish during their preschool year. Twenty-six of the children were assessed in English using the DRA2 at the end of first grade. Twenty-four children were assessed in Spanish using the EDL2 at the end of first grade. Analyses were conducted to examine the associations of both Spanish and English preschool test scores with first grade assessment data, separately by language of assessment in first grade.

For children assessed in English in first grade, there was a significant association between DRA2 in first grade and one preschool assessment administered in English. DRA2 was correlated with English Applied Problems at (r=.41, p=.019). No pattern of association was observed between DRA2 and the Spanish preschool assessments.

For children assessed in Spanish at first grade, there were three significant associations between preschool assessments and EDL2. TVIP was correlated with EDL2 (Spanish; r=.37, p<.05), and so was WJ Letter Word Identification (English; r=.47, p<.01), and the WM Applied Problems (Spanish; r=.68, p<.001).

In sum, for Cohort 4, preschool readiness related to early math was positively associated with the DRA2 for children whose primary language was English, but a stronger pattern of associations was found for developing bilingual children who were assessed in Spanish, at the end of first grade. For these children, there is strong evidence that kindergarten readiness at the end of preschool is strongly associated with reading assessment data at the end of first grade.

COHORT 5

In spring 2014, 147 children were enrolled in kindergarten and had reading assessment data. Ninety-two of these children were assessed only in English in preschool and were assessed with the DRA2 in the kindergarten year. Associations between the preschool assessments and kindergarten DRA2 scores were strong and significant. PPVT-4 scores in preschool were correlated with kindergarten DRA2 scores at .45 ($\underline{p}$<.001). Letter-Word Identification was correlated with DRA2 at .49 ($\underline{p}$<.001). Applied Problems was associated with DRA2 scores at .36 ($\underline{p}$<.001).

Fifty-five Cohort 5 children with kindergarten reading assessment data in spring 2014 had been assessed in both English and Spanish during their preschool year. Of these, twenty-seven children were assessed in English using the DRA2 at the end of kindergarten. Twenty-eight were assessed in Spanish using the EDL2 at the end of kindergarten. Analyses were conducted to examine the associations of both Spanish and English preschool test scores with kindergarten assessment data, separately by language of assessment in kindergarten.

For children assessed in English in kindergarten, there was pattern of association between the DRA2 and preschool early math assessments in English or Spanish, but not for the language and literacy measures. Significant associations were found between DRA2 and both the WJ and WM Applied Problems (English: r=.67, p<.001, Spanish: r=.47, p<.01). For children assessed in Spanish in kindergarten, there was only one significant association. EDL2 was correlated with Spanish Applied Problems at .36 (p<.05).

In sum, for Cohort 5, there was a strong pattern of associations for children whose primary language was English. For these children, there is strong evidence that kindergarten readiness at the end of preschool is strongly associated with reading assessment data at the end of kindergarten. For developing Spanish-English bilingual children, there was a strong pattern of associations between kindergarten readiness related to the early math and reading skill at the end of kindergarten, but not for the language and literacy measures.

DO CHILDREN FROM DIFFERENT DEMOGRAPHIC SUBGROUPS DIFFER IN THEIR READING PROFICIENCY IN THE EARLY ELEMENTARY YEARS?

To address this question, we examined the associations between demographic characteristics and reading scores for each cohort.

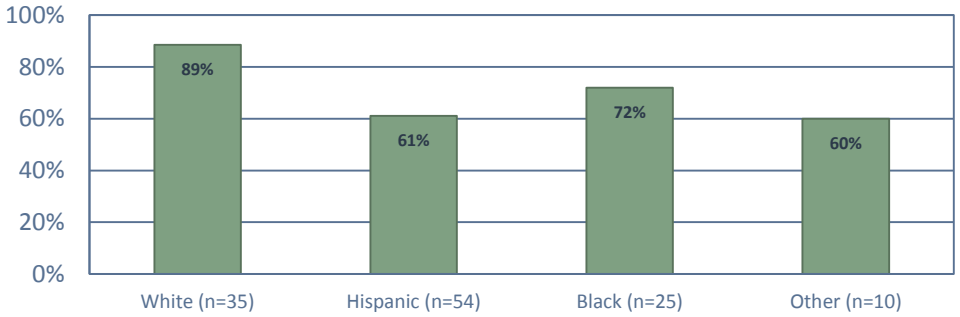
COHORT 2

TCAP

In spring 2014, among Cohort 2 DPP graduates, TCAP proficiency differed as a function of racial/ethnic group (see Figure 30).[47] The vast majority of the white children and three-fourths of those in the Black race/ethnicity category scored proficient or advanced on the 3rd grade reading TCAP. Hispanic and children categorized as "other" race/ethnicities were slightly less likely to score proficient or advanced with about two-thirds scoring proficient or advanced in reading.

[47] χ^2_3=16.70; $\underline{p}$<.01

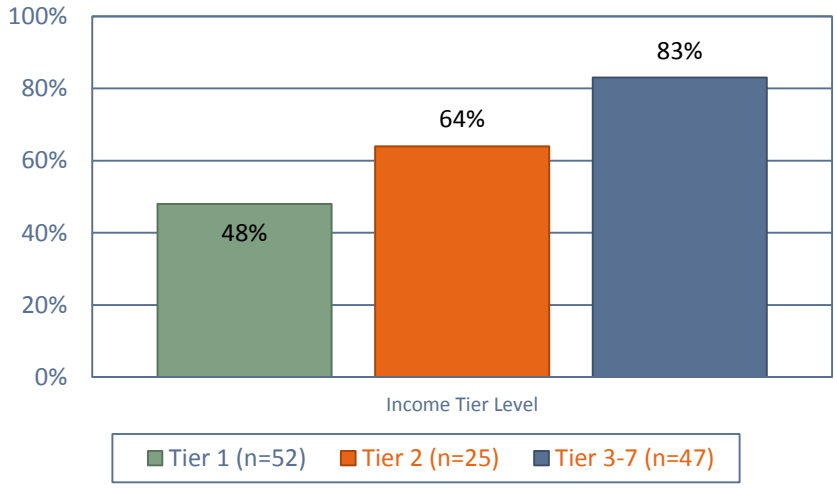
Figure 30: Percent of Children Scoring Proficient or Advanced on 3rd Grade Reading TCAP by Race/Ethncity, Cohort 2, Spring 2014



Income tier significantly associated with third grade reading proficiency (see Figure 31).[48] As income level increased, the proportion of children scoring proficient or advanced on TCAP reading. In Tier 1 about half the children score in the proficient range, while over 80% in Tiers 3-7 do so.

Home language was not significantly associated with third grade reading proficiency. About two-thirds of the children whose home language was English scored proficient or advanced on the TCAP reading. In contrast, about a third of children with another home language scored proficient or advanced. A similar pattern was observed for child primary language.[49] Child gender was also not associated with TCAP reading proficiency.[50]

Figure 31: Percent of Children Scoring Proficient or Advanced on Third Grade TCAP Reading by Income Tier Level, Cohort 2, Spring 2014



[48] χ^2_1=14.03, p<.01

[49] Home Language: χ^2_1=1.21, p=.27, Primary Language: χ^2_1=.48, p=.49

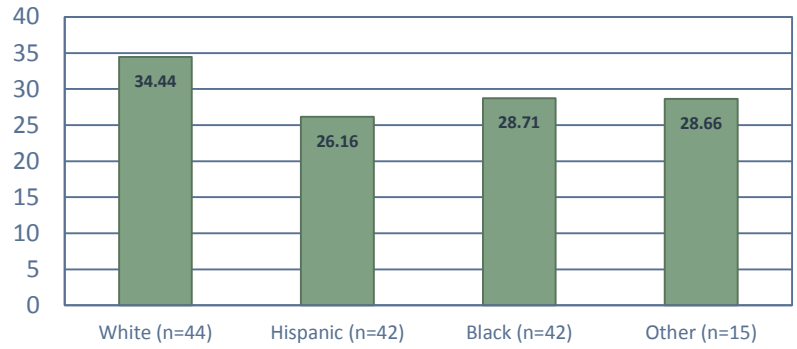
[50] χ^2_1=.03, p=.86

COHORT 3

ENGLISH READING ASSESSMENT (DRA2)

In spring 2014, there was not a significant difference between boys and girls on the DRA2. There was, however, a difference in DRA2 scores by race/ethnicity (see Figure 32).[51] Follow-up Tukey tests revealed that white children, whose scores exceeded grade level expectations on average, scored significantly higher than Hispanic children, whose average score fell short of grade level expectations.[52]

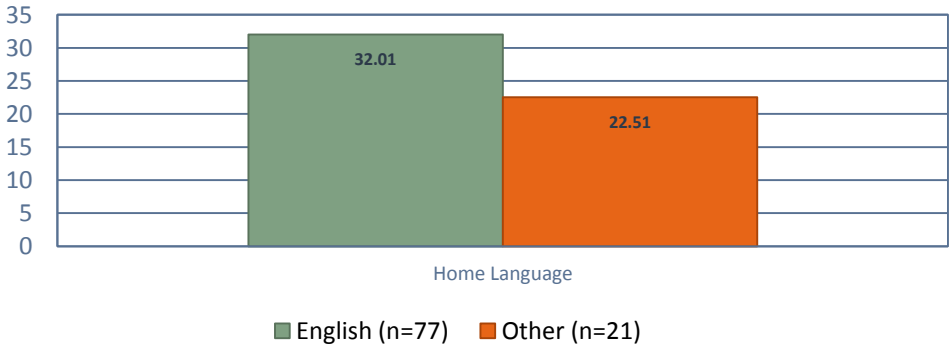
Figure 32: Second Grade Reading Level (DRA2) by Race/Ethnicity, Cohort 3, Spring 2014*



*A score of 28 is considered reading "on grade level" for the end of second grade.

DRA2 scores also differed by children's home language (see Figure 33).[53] Children whose home language was English scored significantly higher than children with another home language. Children's primary language was also significantly associated with DRA2 scores.

Figure 33: Second Grade Reading Level (DRA2) by Home Language, Cohort 3, Spring 2014*

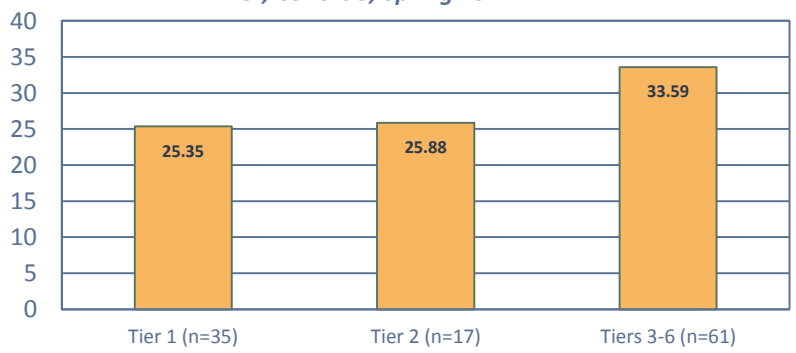


*A score of 28 is considered reading "on grade level" for the end of second grade.

[51] F(3,113)=5.66, p<.01
[52] A DRA2 score of 28 is considered "on grade level" for the end of second grade.
[53] Home Language: t=4.79, df=100, p<.001; Primary Language: t=3.41, df=105, p<.01;

Significant differences were found by income tier (see Figure 34).[54] As income tier increased, so did average reading levels. Follow-up Tukey tests revealed that the group of children in Tiers 3-6[55] had significantly higher reading levels than children in Tier 1 and Tier 2.

*Figure 34: Second Grade Reading Level (DRA2) by Income Tier, Cohort 3, Spring 2014**



*A score of 28 is considered reading "on grade level" for the end of second grade.

Finally, we examined whether DRA2 scores differed by the region of the city where children lived. This effect was non-significant, indicating that performance on the DRA2 did not systematically differ depending on where children lived.[56]

SPANISH READING ASSESSMENT (EDL2)

No significant differences in EDL2 scores were found by child gender. We were unable to test for differences by race/ethnicity because virtually all of the children assessed in Spanish using the EDL2 were Hispanic. Similarly, we were unable to test for differences by primary language and home language because, as expected, all children had primary and home languages other than English. The effect for income was non-significant, as was the effect for region of the city.

COHORT 4

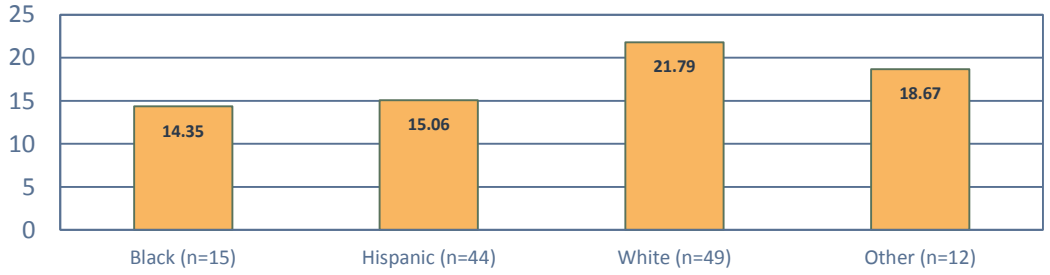
ENGLISH READING ASSESSMENT (DRA2)

For children assessed in English in first grade, there were significant associations between the DRA2 and four of the background characteristics: ethnicity, home and primary language, and income tier. Figure 35 displays the results for ethnicity. White children scored at statistically significantly higher proficiency levels than Hispanic and Black children who were below the first grade proficiency level.[57] Figure 36 displays the results for home language (primary language findings are similar). Children whose home language was English scored significantly higher on

[54] Because of small sample sizes in some of the tiers, a collapsed version of income tier with three levels was used for this analysis: tier 1, tier 2 and tiers 3-6. $F(2,114)=10.70$, $p<.001$.
[55] The DPP income tier structure changed from a 7-level to a 6-level tier after Cohort 2.
[56] $F(4,112)=1.61$, $p=.18$
[57] $F(2,120)=9.82$, $p<.001$

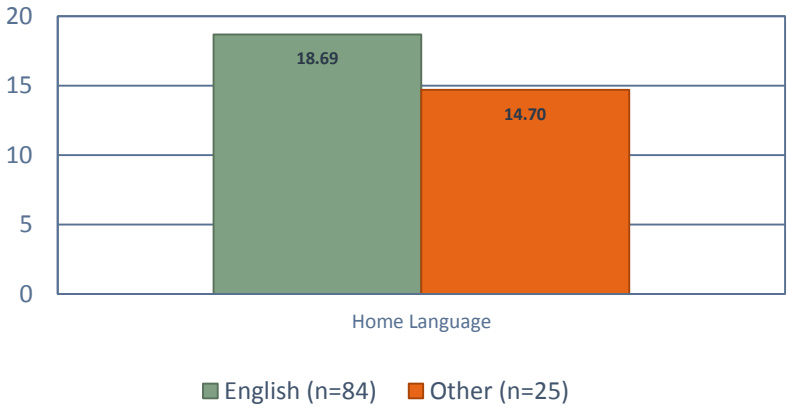
the DRA2 than children with another home language.[58] Figure 37 displays the results for income tier. As income tier increases, so do average DRA2 scores. Follow-up Tukey tests revealed that children in tier 2 and tiers 3-5 scored significantly higher than children in tier 1.[59]

*Figure 35: First Grade Reading Level (DRA2) by Ethnicity, Cohort 4, Spring 2014**



*A score of 16 is considered reading "on grade level" for the end of first grade.

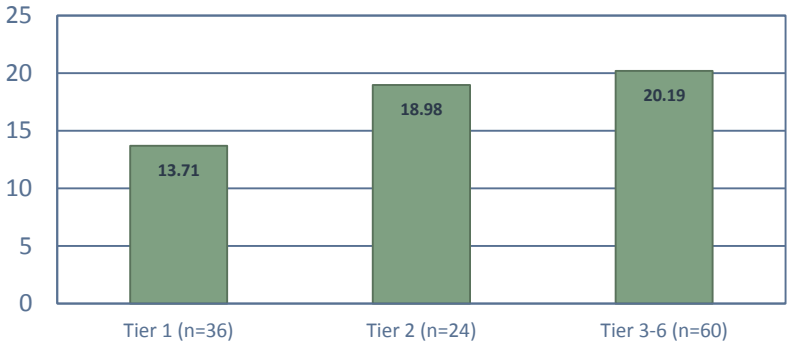
*Figure 36: First Grade Reading Level (DRA2) by Child Primary Language and Home Language, Cohort 4, Spring 2014**



*A score of 16 is considered "on grade level" for the end of first grade.
'86 children had English identified as their primary language, 24 had another language.
ˆ84 children had English identified as the home language, 25 had another language.

[58] t=2.70, df=111, p<.01
[59] F(2,121)=13.28, p<.001

*Figure 37: First Grade Reading Level (DRA2) by Income Tier, Cohort 4, Spring 2014**



*A score of 16 is considered "on grade level" for the end of first grade.

SPANISH READING ASSESSMENT (EDL2)

No significant associations were found between first grade EDL2 and any of the demographic characteristics. We were unable to test for differences by race/ethnicity because virtually all of the children assessed in Spanish using the EDL2 were Hispanic. Similarly, we were unable to test for differences by primary language and home language because, as expected, nearly all children had primary and home languages other than English.

COHORT 5

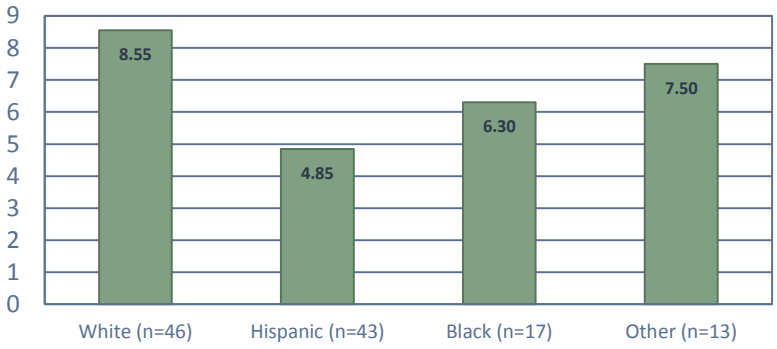
ENGLISH READING ASSESSMENT (DRA2)

Kindergarten DRA2 scores varied significantly by race/ethnicity and tier level. We were unable to test for differences by primary language and home language because, as expected, all children had primary and home languages of English. No differences were found by gender or region of the city. The results for race/ethnicity are presented in Figure 38. Follow-up Tukey tests revealed that Hispanic children scored significantly lower than white children on the DRA2.[60] . There was also a significant difference by income tier (see Figure 39).[61] While all income tier groups were reading, on average, at or above grade level, there was clearly an association between income tier and reading level. As income tier increased, so did average reading levels. Follow-up Tukey tests revealed that the group of children in Tiers 3-6 had significantly higher reading levels than children in Tier 1. The other pairwise comparisons were not statistically significant.

[60] $F(3,113)=4.01$, $p<.01$

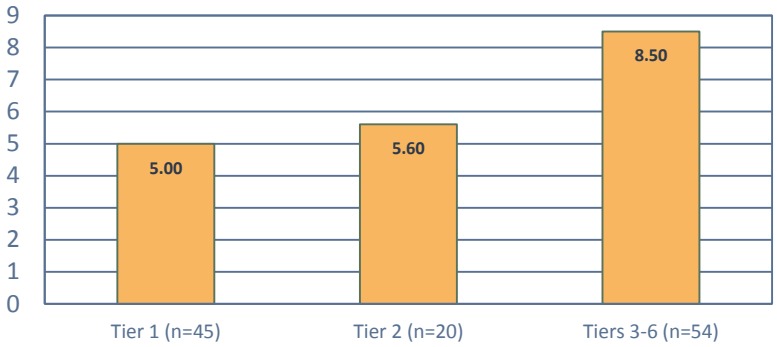
[61] Because of small sample sizes in some of the tiers, a collapsed version of income tier with three levels was used for this analysis: tier 1, tier 2, tiers 3-6). $F(2,122)=6.13$, $p<.01$

*Figure 38: Kindergarten Reading Level (DRA2) by Race/Ethnicity, Cohort 5, Spring 2014**



*A score of 4 is considered reading "on grade level" for the end of kindergarten.

*Figure 39: Kindergarten Reading Level (DRA2) by Tier Level, Cohort 5, Spring 2014**



*A score of 4 is considered "on grade level" for the end of kindergarten.

SPANISH READING ASSESSMENT (EDL2)

There was not a significant difference in EDL2 scores by child gender or income tier. We were unable to test for differences by race/ethnicity because virtually all of the children assessed in Spanish using the EDL2 were Hispanic. Similarly, we were unable to test for differences by primary language and home language because, as expected, nearly all children had primary and home languages other than English.

DO CHILDREN FROM DIFFERENT DPP PROVIDER TYPES (DPS VS. COMMUNITY SITES) DIFFER IN THEIR READING PROFICIENCY IN THE EARLY ELEMENTARY YEARS?

COHORT 2

We compared Cohort 2 children who had been enrolled in community preschools (n=80) with children who had been enrolled in DPS preschools (n=44) on third grade TCAP proficiency in reading administered during the spring of 2013. There was not a significant difference between these two groups.[62]

COHORT 3

We compared Cohort 3 children who had been enrolled in community preschools (n=53) with children who had been enrolled in DPS preschools (n=60) on second grade reading assessments in English (DRA2) administered during the spring of 2014. No significant difference between these two groups.[63] [64] No comparison was made by provider type for EDL2 scores because only one child in the community preschool site sample who received the EDL2.

COHORT 4

We compared Cohort 4 children who had been enrolled in community preschools (n=57) with children who had been enrolled in DPS preschools (n=63) on first grade reading assessments in English (DRA2) administered during the spring of 2013. A significant difference was found between these two groups.[65] We were limited in our ability to test for a provider type difference in EDL2 by the fact that only one of the children who had been enrolled in community sites were assessed with the EDL2.

COHORT 5

Among children assessed in English at the end of kindergarten with the DRA2, there was no significant difference between the 50 children who had attended community sites and the 69 children who had attended DPS preschools.[66] Among children assessed in Spanish at the end of kindergarten with the EDL2, there was no significant difference between the 5 children who attended community sites and the 23 children who attended DPS.

IS THE QUALITY OF THE PRESCHOOL PROGRAM ATTENDED ASSOCIATED WITH READING PROFICIENCY IN THE EARLY ELEMENTARY YEARS?

As described in previous annual evaluation reports,[67] there has been very limited variability in the star rating of the preschools attended by DPP children. Very few children were enrolled in programs with less than a star 3 rating.

---

[62] χ^2_1=.41, n.s.

[63] $\underline{t}$=1.14, df=111, n.s.

[64] Community preschools: mean=31.0, sd=8.99; DPS preschools: mean=29.0, sd=10.41; 28 is considered "on grade level" for second grade.

[65] $\underline{t}$=2.09, df=118, p<.05

[66] $\underline{t}$=2.94, df=74.7, $\underline{p}$<.01

[67] Klute, M. M. (2009). *Denver Preschool Program: Report on Child Outcomes, 2008-09 School Year.* Unpublished Report, October. Denver: Clayton Early Learning Institute.

Because of this, we also examined total Qualistar rating points, number of points earned for training and education, and mean ECERS-R score for the DPP classrooms at the site. These variables had a bit more variability, but were still quite restricted in range. Beginning with Cohort 3, an additional measure of program quality, the CLASS observation (described above) was added to our evaluation design to provide a measure that would be more sensitive to the variability in the quality of DPP preschools. For Cohorts 3, 4, and 5, we examine the CLASS in lieu of the Qualistar rating.

COHORT 2

When we examined the distribution of star rating for children who had third grade TCAP data in spring 2014, we found that only 8.7% had been enrolled in preschools with less than a star 3 rating. Sixty-four percent of children had been enrolled in star 3 preschools and 27% had been enrolled in star 4 preschools. There was not a significant association between star level and TCAP reading proficiency.

Correlations were computed between measures of quality (total rating points earned, number of training and education points earned, and mean ECERS-R score for DPP classrooms at the site) with TCAP proficiency. All of these correlations were non-significant.

COHORT 3

Cohort 3 was the first cohort for which we have CLASS observation data available. There was very little variability in CLASS Emotional Support, so we restricted our analysis to Classroom Organization and Instructional Support. We computed correlations between these two CLASS subscales and the DRA2 and EDL2. None of these four correlations was significant.

COHORT 4 AND COHORT 5

Cohort 4 and 5 matched findings from Cohort 3 on this examination. There was very little variability in CLASS Emotional Support, so we restricted our analysis to Classroom Organization and Instructional Support. We computed correlations between these two CLASS subscales and the DRA2 and EDL2. None of these four correlations was significant.

CONCLUSIONS

Our evaluation of the Denver Preschool Program focused on five descriptive questions about the progress DPP participants make during their preschool year and beyond:

1. Do children make progress in their development while in DPP early childhood environments (i.e., language, literacy, mathematics, and social-emotional development)?

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

Klute, M. M. (2010). *Denver Preschool Program: Report on Child Outcomes, 2009-10 School Year.* Unpublished Report, November. Denver: Clayton Early Learning Institute.
Klute, M. M. & Ponce, C. (2011). *Denver Preschool Program: Report on Child Outcomes, 2010-11 School Year.* Unpublished Report, November. Denver: Clayton Early Learning Institute.

3. Do children from different income levels and with different primary languages make similar progress in their development while in DPP early childhood environments?

4. Do children who received DPP tuition credits compare favorably with the district as a whole on assessments administered by Denver Public Schools in kindergarten and beyond?

5. Is attendance at higher-rated preschool programs associated with greater kindergarten readiness and long-term academic success (as measured by TCAP)?

The 2013-14 school year was the seventh year of the DPP program. Cohort 1 from the second year of DPP's operation was the first cohort for which we were able to fully implement our evaluation design. Cohort 2 was expected to be in the third grade during the 2013-14 school year, the first grade in which students take the TCAP. As such, this year's annual report represents the second time that we were able to address question 5.

QUESTION 1: DO CHILDREN MAKE PROGRESS IN THEIR DEVELOPMENT WHILE IN DPP EARLY CHILDHOOD ENVIRONMENTS?

Children did make significant progress in their academic and social-emotional development during their preschool year. With respect to academic skills, assessments of all children in English demonstrated that children made progress in the areas of vocabulary, literacy, and math skills. Spanish-speaking children also made progress in their vocabulary, literacy, and math skills assessed in Spanish over the course of their preschool year. The gains observed were above and beyond what would be expected based on normal development. Progress was observed in social-emotional development as well. Over the course of the preschool year, teachers reported that children demonstrated significantly more protective factors and significantly fewer behavioral concerns.

QUESTION 2: TO WHAT EXTENT AND IN WHAT AREAS ARE CHILDREN ENROLLED IN DPP READY FOR KINDERGARTEN?

Results of the evaluation suggest that the vast majority of children are ready for school, both academically and social-emotionally. When considering both languages of assessment, we concluded that relatively few children had scores in the risk range (below 85) on assessments of their vocabulary, literacy and math skills. These standardized assessments are scaled such that 84% of the general population would be expected to score above the at-risk range (a score of 85 or above). Scores for literacy and math in this sample clearly exceed that threshold. Vocabulary scores in this sample did not quite reach this threshold. We also considered a more stringent criterion to examine readiness, namely scores that met or exceeded the population average (a score of 100). The assessments are scaled such that half of children in the general population would be expected to meet or exceed this threshold. When both languages of assessment were considered, more children than would be expected (i.e., more than half) met this more stringent criterion: 60% for vocabulary and about three-quarters for literacy and math. When teachers rated children's behaviors, their ratings of protective factors were high for most children. Protective factors were rated as an area of concern by teachers for 14% of children in this sample. Teachers' ratings of behavioral concerns were low on average. Teachers identified behavioral concerns as an area of concern for about 9% of children. Parents identified protective factors as an area of concern for about 20% of children and behavioral concerns as an area of concern for about a third of children. The DECA, the social-emotional assessment we used, provides T-scores, which are scaled such that nearly 16% of the general population would be expected to be identified as having a concern. All of the teachers' ratings are within that threshold. Parents' ratings of protective factors, however, exceeded that threshold. Parents identified below threshold protective

factors. They additionally reported above-threshold behavioral concerns for about twice as many children as would be expected based on the way in which the assessment is scored.

It is interesting that parents' identify behavioral concerns more frequently than do teachers. About 37% of parents rated their child significantly higher on this area than teachers, with higher scores indicating greater concerns. The DECA uses different norms to take into account systematic differences between parents' and teachers' points of view in the general population. As a result of these different norms, one should interpret these differences as real differences between parents and teachers and not simply an artifact of a difference in the way that parents and teachers generally view behavior.

QUESTION 3: DO CHILDREN FROM DIFFERENT INCOME LEVELS AND WITH DIFFERENT PRIMARY LANGUAGES MAKE SIMILAR PROGRESS IN THEIR DEVELOPMENT WHILE IN DPP EARLY CHILDHOOD ENVIRONMENTS?

Our ability to address this question is limited by a strong association between income and children's primary language. In this year's sample, nearly all children whose primary language was not English were from the lowest two income tiers as compared with about 50% of children whose primary language is English. As a result, it is impossible to disentangle the effects of income and primary language. Any associations that are observed are likely associated with the co-occurrence of these two factors.

Results of this year's study revealed a consistent pattern of results for income tier (defined by income adjusted for family size). In all cases, the income tier by time interaction was non-significant, indicating that children progressed similarly in these areas over the course of their preschool year, regardless of their income tier. Child of lower income tiers started and ended the year with lower scores on the assessments than their more wealthy counterparts. Likewise, analyses of primary language groups revealed that children whose primary language is not English started the year lower than their primarily English-speaking counterparts on English vocabulary and the social-emotional assessments. Unlike last year, the children whose primary language was not English did not increase at a greater rate than the primarily English-speakers; in fact, the persistence of low vocabulary scores was of concern.

QUESTION 4: DO CHILDREN WHO RECEIVED DPP TUITION CREDITS COMPARE FAVORABLY WITH THE DISTRICT AS A WHOLE ON ASSESSMENTS ADMINISTERED BY DENVER PUBLIC SCHOOLS IN KINDERGARTEN AND BEYOND?

Cohorts 2, 3, 4, and 5 were demographically similar to the populations of children in third grade, second grade, first grade and kindergarten, respectively, in terms of their gender and ethnic backgrounds. A smaller proportion of children from each cohort qualified for free or reduced lunch than in the district as a whole, but this was most pronounced for Cohort 3.

Cohort 2 children were compared to the population of third graders in DPS. DPP graduates scored proficient or advanced on the reading TCAP compared at a slightly higher rate (6%) than the district as a whole.

Cohort 3 children were compared to all second graders in DPS. Among children whose reading was assessed in English in second grade, DPP graduates were more likely to be reading at or above grade level than children in the district as a whole. Among children assessed in Spanish, DPP graduates were slightly less likely to be reading at grade level than the district as a whole.

Cohort 4 children were compared to the population of first graders in DPS. Among children whose reading was assessed in English, the proportion of children in Cohort 3 who were reading at or above grade level exceeded the proportion in the district as a whole. Among children assessed in Spanish at the end of the first grade year, the proportion of DPP graduates reading at or above grade level was also greater than the proportion in the district as a whole.

Cohort 5 children were compared to the population of kindergarteners in DPS. For children assessed in both English and Spanish at the end of kindergarten, DPP graduates were much more likely to be reading at or above grade level than in the district as a whole.

QUESTION 5: IS ATTENDANCE AT HIGHER-RATED PRESCHOOL PROGRAMS ASSOCIATED WITH GREATER KINDERGARTEN READINESS AND LATER ACADEMIC SUCCESS?

Our ability to examine quality in conjunction with later academic success for cohorts 1 and 2 was limited by the lack of variability in the Qualistar rating. Very few children had been enrolled in preschools with less than a star 3 rating. In our analyses, we did not find a strong pattern of association between preschool quality and reading skill in second or third grade.

Starting with Cohort 3, with hopes for increased variability, we examined the association between CLASS observation scores and child outcomes. For Cohort 6 (in preschool in 13-14), there were no significant correlations between the English academic assessments (vocabulary, literacy, math, and sustained attention) and classroom quality. There were also no significant correlations for Spanish assessments. Similarly, there were no associations observed for the spring teacher DECA ratings and any of the CLASS domains after controlling for fall ratings.

Children in Cohort 5 who were enrolled in classrooms scoring higher on Classroom Organization and Instructional Support made greater gains over the preschool year in their literacy skills assessed in English. Children in classrooms that were rated higher on Instructional Support also demonstrated greater gains in Protective Factors. For Cohorts 3 and 4, there were no associations between CLASS observation scores and reading assessment scores in either language.

SUMMARY AND FUTURE DIRECTIONS

This evaluation described children's progress during the course of their DPP preschool year. In general, children progressed in their vocabulary, literacy, and math skills as assessed in both English and Spanish at a rate which exceeded what would be expected simply because of maturation, with the exception of Spanish-speakers English receptive vocabulary. Children demonstrated positive changes in their social-emotional functioning over time; teachers reported that children demonstrated more positive behaviors and fewer negative behaviors at the end of the school year than at the beginning.

With the first two cohorts of children we studied, we were limited in our ability to examine preschool quality in conjunction with child outcomes because we had relied on Qualistar data as our measure of quality. There was very little variability in Qualistar ratings; over 90% of children in these cohorts attended star 3 or 4 preschools. Nonetheless, we attempted to examine the association between quality and first and second grade reading skills for these cohorts of children. We did not find a strong pattern of associations.

In an attempt to address this restriction of range problem, starting with the 2010-11 school year, we directly observed classrooms with an observational measure focused on teacher-child interactions. We did see greater

variability among classrooms on 2 of the 3 domains assessed by this measure (Classroom Organization and Instructional Support), but we did not find a strong pattern of associations between this measure of quality and child outcomes in kindergarten or first grade. In Cohort 5, the CLASS was related to literacy skills and protective factors. However, Cohort 6 again showed no association. One possibility will be to examine scores at the dimension level in subsequent years.

Overall, children in this study were enrolled in DPP preschools that were of relatively high quality and the children made excellent progress over the course of their preschool year, on average. The results of this study also suggest that DPP graduates tend to demonstrate similar or greater reading proficiency in kindergarten, first grade, and second grade than the district as a whole. The only exception to this was a group of Cohort 3 children assessed in Spanish in second grade (which matches the previous year's findings from examining this cohort). Results from future years of this annual evaluation will provide the opportunity to replicate these findings as well as to continue to follow these cohorts of children as they move through elementary school.

APPENDIX

Table A1: Sample Characteristics—Spring 2014

	Entire Sample, weighted[1]	By Provider Type, Unweighted		
Characteristic		Community	DPS	Significance of Difference by Provider Type
Sex				χ^2_1=1.00; p<.01
Female	48.5%	48.5%	48.5%	
Male	51.5%	51.5%	51.5%	
Ethnicity				χ^2_5=17.5; p<.01
Hispanic	44.8%	26.3%	53.5%	
White (not of Hispanic origin)	35.7%	46.5%	30.7%	
African-American (not of Hispanic origin)	9.9%	12.1%	8.9%	
Multi-Racial	4.3%	9.1%	2.0%	
Asian/Pacific Islander	4.0%	4.0%	4.0%	
Native American	1.0%	1.0%	1.0%	
Child's Primary Language				χ^2_2=15.4; p<.01
English	67.2%	83.8%	59.4%	
Another Language	29.1%	13.1%	36.6%	
Not Reported	2.7%	2.0%	3.0%	
Home Language				χ^2_3=6.24; p<.01
English	53.6%	72.7%	44.6%	
Another Language	17.0%	13.1%	18.8%	
Not Reported	3.3%	6.1%	2.0%	
DPP Income Tier[2]				χ^2_5=3.59; p<.05
Tier 1	36.9%	35.4%	37.6%	
Tier 2	18.7%	14.1%	20.8%	
Tier 3	2.6%	4.0%	2.0%	
Tier 4	6.0%	8.1%	5.0%	
Tier 5	19.6%	23.2%	17.8%	
Tier 6—Income Not Reported	16.3%	15.2%	16.8%	
Star Level of Preschool				χ^2_3=10.91; p<.01
Not Yet Rated	0.0%	0.0%	0.0%	
Provisional	1.0%	3.0%	0.0%	
Star 2	6.3%	7.1%	5.9%	
Star 3	58.7%	44.4%	65.3%	
Star 4	34.1%	45.5%	28.7%	
Region of the City				χ^2_4=11.20; p<.05
Central	12.5%	18.2%	9.9%	
Northeast	36.1%	22.2%	42.6%	
Northwest	17.9%	24.2%	14.9%	
Southeast	8.0%	8.1%	7.9%	
Southwest	25.6%	27.3%	24.8%	

[1]The weighted sample results are representative of the population of children enrolled in DPP in Fall 2013.

[2]DPP Income Tiers are determined using family income and family size. Complete information about how DPP Income Tiers are calculated is included in the Appendix Table A2.

Table A2: DPP Income Tiers

Income Tier 1	
Household Size	*Annual Income Equal to or Less Than*
2	$14,710
3	$18,530
4	$22,350
5	$26,170
6	$29,990
7	$33,810
8	$37,630
9	$41,450
If more than 9 family members	Add $3,820 for each additional family member

Income Tier 2		
Household Size	*Annual Income*	
	More Than	*Equal to or Less Than*
2	$14,711	$27,214
3	$18,531	$34,281
4	$22,351	$41,348
5	$26,171	$48,415
6	$29,991	$55,482
7	$33,811	$62,549
8	$37,631	$69,616
9	$41,451	$76,683
If more than 9 family members	Add $7,067 for each additional family member	

Income Tier 3		
Household Size	*Annual Income*	
	More Than	*Equal to or Less Than*
2	$27,215	$33,098
3	$34,282	$41,693
4	$41,349	$50,288
5	$48,416	$58,883
6	$55,483	$67,478
7	$62,550	$76,073
8	$69,617	$84,668
9	$76,684	$93,263
If more than 9 family members	Add $8,595 for each additional family member	

Income Tier 4		
Household Size	*Annual Income*	
	More Than	*Equal to or Less Than*
2	$33,099	$44,130
3	$41,694	$55,590
4	$50,289	$67,050
5	$58,884	$78,510
6	$67,479	$89,970
7	$76,074	$101,430
8	$84,669	$112,890
9	$93,264	$124,350
If more than 9 family members	Add $10,028 for each additional family member	

Income Tier 5	
Household Size	*Annual Income*
	More Than
2	$44,130
3	$55,590
4	$67,050
5	$78,510
6	$89,970
7	$101,430
8	$112,890
9	$124,350
If more than 9 family members	Add $11,460 for each additional family member