



Denver Preschool Program: Report on Child Outcomes

2012-13 School Year

Prepared for the Denver Preschool Program

**By Mary Maguire Klute, Ph.D., Buechner Institute for Governance, School of
Public Affairs, University of Colorado, Denver**

In collaboration with Clayton Early Learning Institute

October 20, 2013

CONTENTS

Executive Summary	4
Do children make progress in their development while in DPP early childhood environments?	4
To what extent and in what areas are children enrolled in DPP ready for kindergarten?	4
Do children from different income levels and with different primary languages make similar progress in their development while in DPP early childhood environments?	5
Do children who received DPP tuition credits compare favorably with the district as a whole on assessments administered by Denver Public Schools?	5
Is attendance at higher-rated preschool programs associated with greater kindergarten readiness and later academic success?	5
Summary and Future Directions	6
Introduction	7
Methods	8
Sample	8
Procedures	15
Measures	16
Results: Preschool Year	18
Preliminary Analyses	18
Kindergarten Readiness	24
Change in Assessments over the Preschool Year	29
Preschool Quality and Child Outcomes	35
Results: Elementary School	36
How similar are DPP graduates to the population of children in the district as a whole?	36
What is the overall reading proficiency of DPP graduates in the early elementary years? How does this compare to the district as a whole?	42
To what extent is preschool readiness associated with reading proficiency in the early elementary years?	45
Do children from different demographic subgroups differ in their reading proficiency in the early elementary years?	48

Do children from different DPP provider types (DPS vs. community sites) differ in their reading proficiency in the early elementary years?	56
Is the quality of the preschool program attended associated with reading proficiency in the early elementary years?	57
Conclusions	58
Question 1: Do children make progress in their development while in DPP early childhood environments?	59
Question 2: To what extent and in what areas are children enrolled in DPP ready for kindergarten?	59
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?	60
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?	60
Question 5: Is attendance at higher-rated preschool programs associated with greater kindergarten readiness and later academic success?	61
Summary and Future Directions	61
Appendix.....	62

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 Clayton Early Learning Institute, which is focused on questions related to the development of children enrolled in DPP both during their preschool year and beyond.

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.

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 skills assessed in English and Spanish, where appropriate, we concluded that relatively few children had scores in the risk range on assessments of their vocabulary, literacy and math skills. Further, more children than would be expected scored at or above the average on these assessments. Results were particularly striking for the literacy and math assessments, where about three-quarters of children scored at the average or above. Based on the way the assessments are scaled, one would only expect about half of children in the general population to score in this range.

Teachers' ratings of children's positive behaviors, called *protective factors* (attachment, initiative, and self-control) were high for most children. These protective factors were rated as an area of concern for fewer than 5% of children. In addition, teachers' ratings of behavioral concerns were rather low on average. Teachers identified behavioral concerns as an area of concern for only about 11% of children. Based on the way this assessment is scaled, one would expect about 16% of children to be classified in the concern range.

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 somewhat by a strong association between income and children's primary language. In the sample of children enrolled in DPP during the 2012-13 school year, 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.

Children whose primary language was not English tended to start the year lower than children with another primary language on English vocabulary and the social-emotional assessments. However, there was a somewhat consistent pattern of effects demonstrating that these children increased at a more rapid pace over the course of the year. Results for income tier were inconsistent.

DO CHILDREN WHO RECEIVED DPP TUITION CREDITS COMPARE FAVORABLY WITH THE DISTRICT AS A WHOLE ON ASSESSMENTS ADMINISTERED BY DENVER PUBLIC SCHOOLS?

We followed four cohorts of DPP graduates who were enrolled in kindergarten, first, second, and third grade during the 2012-13 school year. DPP graduates who took the third grade TCAP scored proficient or advanced at a rate similar to the district overall. DPP graduates whose reading ability was assessed in English at the end of second grade were slightly more likely to be reading at or above grade level than children in the district as a whole. The small group of DPP graduates who were assessed in Spanish at the end of the second grade year was more likely to be reading on grade level than the district as a whole.

Among children whose reading was assessed in English in first grade, the proportion of DPP graduates who were reading at or above grade level was greater than the proportion in the district as a whole. Among children assessed in Spanish, the proportion of DPP graduates reading at or above grade level at the end of first grade was similar to the district as a whole.

Among kindergarteners, the proportion of DPP graduates who were reading at or above grade level exceeded the proportion of children in the district as a whole who were reading at or above grade level for both languages of assessment.

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

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). There were some associations between CLASS observation scores and children's gains on assessments over the course of the preschool year. Specifically, Classroom Organization and Instructional Support were positively correlated with gains in literacy skills assessed in English. Instructional Support was also positively associated with gains in Protective Factors. We did not find any associations between quality and reading skill at the end of kindergarten or first grade.

SUMMARY AND FUTURE DIRECTIONS

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 children assessed in Spanish in first grade. 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.

Clayton Early Learning Institute and the Buechner Institute for Governance collaborate with Augenblick, Palaich and Associates to complete an annual evaluation of DPP. This report details the work completed by Clayton Early Learning Institute and the Buechner Institute for Governance, which is focused on questions related to the development of children enrolled in DPP both during their preschool year and beyond¹. 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 2012-13 school year was the sixth 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 fully 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 first time that we were able to address question 5.

¹ Augenblick, Palaich and Associates has prepared 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 5 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
Cohort 1	Preschool	Kindergarten	1 st Grade	2 nd Grade	3 rd Grade
Cohort 2		Preschool	Kindergarten	1 st Grade	2 nd Grade
Cohort 3			Preschool	Kindergarten	1 st Grade
Cohort 4				Preschool	Kindergarten
Cohort 5					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.² 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 third grade during the 12-13 school year (see Table 1). We obtained spring reading assessment data for 155 children (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. In order 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.

² For more information about this sample and results from the preschool year, readers are referred to the Annual Evaluation Report. Klute, 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 second grade during the 12-13 school year (see Table 1). We obtained reading data for 136 children (68% of the sample; 71% of the sample when sampling weights were applied). Ninety-seven percent of the children were in second 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 first grade during the 12-13 school year (see Table 1). We obtained reading data for 137 children (67% of the sample; 69% of children for whom we were able to obtain DPSIDs; 75% of the total sample when sampling weights were applied). All of the children were in first 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 kindergarten during the 12-13 school year (see Table 1). We obtained reading data for 146 children (72% of the sample; 83% of the sample when sampling weights were applied). All of the children were in kindergarten as expected.

COHORT 5

SAMPLING PLAN

As with previous years, we stratified our sample for Cohort 5 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.³ In August 2012, 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 2012, a sample of 100 children

³ Information about the evaluation was provided on the DPP application, which was available in both English and Spanish.

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 722 volunteers and the 385 non-volunteers for two variables. First, there was a significant difference between volunteers and non-volunteers on ethnicity.⁴ Follow-up analyses revealed that this was due to differences in the rates of volunteering for three ethnic groups: Asians, blacks, and whites. Families with children identified as Asian or black were less likely to volunteer than families with children belonging to other ethnic groups. Fifty percent of families with children identified as Asian volunteered compared with 66% of families with children belonging to other ethnic groups. Fifty-seven percent of families with black children volunteered compared with 67% of families with children belonging to other ethnic groups. Families with white children were more likely to volunteer than families with children belonging to other ethnic groups. Seventy percent of families with children identified as white volunteered compared with 63% of families with children belonging to other ethnic groups. Second, there was a significant association between tier level and volunteer status.⁵ Follow-up analyses revealed that this was primarily due to tiers 1 and 5. Families in tier 1 were less likely to volunteer than families in other tiers. Sixty-two percent of families living in tier 1 volunteered to be contacted, compared with 68% of families from other tiers. In contrast, families in tier 5 were more likely to volunteer than families from other income tiers. Seventy-two percent of families from tier 5 volunteered to be contacted compared with 64% of families from other tiers. To adjust for these differences, the sampling frame was stratified by ethnicity (Asian, black, white, other) and tier (1, 5, 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 2012 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 1346 volunteers and 875 non-volunteers on five 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 three ethnic groups: Asian, white, and black.

⁴ $\chi^2=17.49$, $p<.01$

⁵ $\chi^2=11.89$, $p<.05$

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 $\chi^2=42.61$, $p<.0001$

Parents of Asian and black children were significantly less likely to volunteer than other parents. Thirty-one percent of parents of Asian children volunteered compared with 62% of parents of children from other ethnic groups. Fifty-one percent of parents of black children volunteered compared with 61% of parents of children from other ethnic groups. Parents of white children were significantly more likely to volunteer than parents of children from other ethnic groups (percent who volunteered: 66% white, 59% non-white).

A significant difference in rates of volunteering was also detected by region of the city.⁹ Follow up analyses revealed that this difference was due to parents of children residing in the southeast region of the city, who were less likely to volunteer than parents from other regions of the city. Fifty-two percent of parents of children from the southeast region volunteered compared with 62% of parents from other regions.

Rates of volunteering also varied significantly by the Qualistar rating of the preschool attended by the child.¹⁰ Follow-up analyses revealed that this difference was due to differences in the rates of volunteering among parents of children whose children attended Star 3 and Star 4 preschools. Sixty-three percent of parents of children attending Star 3 preschools volunteered as compared with just 58% of parents of children attending other preschools. In contrast, 57% of parents of children attending Star 4 preschools volunteered compared with 63% of parents of children attending other preschools.

Finally, the likelihood of volunteering varied significantly by home language and child primary language.¹¹ Parents of children with primary languages and home languages that were not English were less likely to volunteer than parents of children with English primary languages and home languages. For both variables, 58% of parents of children with home or primary languages other than English volunteered compared with 69% of parents of children with English home or primary languages.

To adjust for these differences, the sampling frame was stratified by ethnicity (Asian, black, white, other), region of the city (southeast vs. other), Qualistar rating (Star 4, Star 3, Star 2) and home language (English vs. other). Since all but two children had the same home language and primary language, we only used one of these variables in our stratification. 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 2012 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.¹² 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

⁹ $\chi^2_4=20.56$, $p<.001$

¹⁰ $\chi^2_2=9.71$, $p<.01$

¹¹ Home language: $\chi^2_2=24.09$, $p<.0001$; child primary language: $\chi^2_2=24.54$, $p<.0001$

¹² 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.

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 2.¹³ The sample was approximately equally split between boys and girls. Hispanics represented about half of the sample; the next most common ethnic group was whites. African-Americans made up slightly more than a tenth of the sample. About three-quarters of children spoke English as their primary language and in close to three-quarters 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 eight percent of families were assigned to the highest tier, Tier 6, because they opted out of the requirement to provide their income.

Nearly all (96%) 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. About a quarter of children lived in the southwest, northwest and northeast regions of the city. The smallest proportion of children lived in southeast Denver. The right hand side of Table 2 presents demographic characteristics by provider type. While there appear to be some slight differences between the demographic compositions of the two samples, none of the differences were statistically significant.

In the spring 2013, there were eleven children that were lost to follow-up for the following reasons: two children moved out of Denver before the spring round and nine children withdrew from their preschool programs and did not enroll in another preschool. An alternate from the same stratum was selected for each of these children and assessed during the spring round.¹⁴ As a result, the total sample size for the 2012-13 school year is 211.

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.

¹³ Sample characteristics for the spring sample, which were nearly identical, are presented in the appendix.

¹⁴ 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 4 Sample Characteristics Fall 2012¹

	Entire Sample, weighted ²	By Provider Type, Unweighted		
Characteristic		Community	DPS	Significance of Difference by Provider Type
Sex				$\chi^2_1=.20$; ns.
Female	49.0%	55.0%	46.0%	
Male	51.0%	45.0%	54.0%	
Ethnicity				$\chi^2_5=8.00$; n.s.
Hispanic	48.7%	38.0%	54.0%	
White (not of Hispanic origin)	30.3%	33.0%	29.0%	
African-American (not of Hispanic origin)	12.0%	18.0%	9.0%	
Multi-Racial	5.0%	7.0%	4.0%	
Asian/Pacific Islander	3.3%	4.0%	3.0%	
Native American	0.7%	0.0%	1.0%	
Child's Primary Language				$\chi^2_1=2.10$; n.s.
English	75.3%	80.0%	73.0%	
Another Language	24.0%	18.0%	27.0%	
Not Reported	0.7%	2.0%	0.0%	
Home Language				$\chi^2_1=3.38$; n.s.
English	73.3%	74.0%	73.0%	
Another Language	22.7%	14.0%	27.0%	
Not Reported	4.0%	12.0%	0.0%	
DPP Income Tier ³				$\chi^2_5=5.35$; n.s.
Tier 1	43.0%	47.0%	41.0%	
Tier 2	17.7%	15.0%	19.0%	
Tier 3	5.7%	5.0%	6.0%	
Tier 4	6.3%	3.0%	8.0%	
Tier 5	19.3%	18.0%	20.0%	
Tier 6—Income Not Reported	8.0%	12.0%	6.0%	
Star Level of Preschool				$\chi^2_2=3.89$; n.s.
Not Yet Rated	0.7%	2.0%	0.0%	
Provisional	0.3%	1.0%	0.0%	
Star 1	0.0%	0.0%	0.0%	
Star 2	3.0%	3.0%	3.0%	
Star 3	56.3%	51.0%	59.0%	
Star 4	39.7%	43.0%	38.0%	
Region of the City				$\chi^2_4=4.34$; n.s.
Central	15.0%	19.0%	13.0%	
Northeast	23.7%	25.0%	23.0%	
Northwest	21.3%	24.0%	20.0%	
Southeast	12.0%	12.0%	12.0%	
Southwest	28.0%	20.0%	32.0%	

¹Some percentages do not sum to 100 because of rounding error.

²The weighted sample results are representative of the population of children enrolled in DPP in Fall 2011.

³DPP 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.

COMMUNITY SAMPLE

Children who were included in the community sample were compared to 1007 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. All of these tests 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 in September 2012.

DPS SAMPLE

Children who were included in the DPS sample were compared to 2121 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. 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 children enrolled in DPS sites in August 2012.

SUMMER 2013

COMMUNITY SAMPLE

Children who were included in the community sample were compared to 1729 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.¹⁷ 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 3495 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.¹⁸ That is, the DPS sample was representative of the population of enrolled children in DPS sites at the end of the school year.

¹⁵ Gender: $\chi^2_1=2.12$, n.s.; ethnicity: $\chi^2_6=3.12$, n.s.; home language: $\chi^2_1=0.30$, n.s.; child primary language: $\chi^2_1=0.02$, n.s.; income tier: $\chi^2_5=3.86$, n.s.; Qualistar rating: $\chi^2_4=5.57$, n.s.; region of the city: $\chi^2_4=3.55$, n.s.

¹⁶ Gender: $\chi^2_1=0.55$, n.s.; ethnicity: $\chi^2_6=0.50$, n.s.; income tier: $\chi^2_6=2.02$, n.s.; Qualistar rating: $\chi^2_2=0.06$, n.s.; home language: $\chi^2_1=0.02$, n.s.; child primary language: $\chi^2_1=0.02$, n.s.; region of the city: $\chi^2_4=1.70$, n.s.

¹⁷ Gender: $\chi^2_1=1.76$, n.s.; ethnicity: $\chi^2_5=4.11$, n.s.; income tier: $\chi^2_6=6.08$, n.s.; Qualistar rating: $\chi^2_4=4.98$, n.s.; home language: $\chi^2_1=0.13$, n.s.; child primary language: $\chi^2_1=0.00$, n.s.; region of the city: $\chi^2_4=6.14$, n.s.

¹⁸ Gender: $\chi^2_1=0.13$, n.s.; Ethnicity: $\chi^2_5=2.28$, n.s.; income tier: $\chi^2_6=6.77$, n.s.; Qualistar rating: $\chi^2_2=1.03$, n.s.; home language: $\chi^2_1=2.64$, n.s.; region of the city: $\chi^2_4=1.49$, n.s.

PROCEDURES

Once parents or guardians of children selected for the study provided informed consent, children were assessed using standardized assessments at their preschool 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.

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 2013. 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 a third 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.

Table 3: Sample sizes by data collection type, Fall 2012 and Spring 2013

Data Collection Activity	Fall 2012	Spring 2013
Standardized Assessments—English	200	200
Standardized Assessments—Spanish	63	66
DECA—Teacher Report	196 (98%)	193 (96.5%)
DECA—Parent Report	198 (94% of the full sample n=211)	
Classroom Observations ¹	163 (77% of the full sample n=211)	

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

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.²⁵

Table 4: Areas of Child Development Assessed

Area Assessed	Name of Assessment	Acronym	Language of Assessment
Receptive Vocabulary	Peabody Picture Vocabulary Test-4 ²⁰	PPVT	English
	Test de Vocabulario en Imágenes Peabody ²¹	TVIP	Spanish
Literacy Skills	Woodcock-Johnson III Achievement Battery, Letter-Word Identification Subtest ²²	WJ LWI	English
	Batería III Woodcock-Muñoz, Letter-Word Identification Subtest ²³	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 ²⁴	DECA	English or Spanish

¹⁹ 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.

²⁰ Dunn, L. M., & Dunn, D. M. (2007). *Peabody Picture Vocabulary Test, Fourth Edition*. Minneapolis: Pearson Assessments.

²¹ Dunn, L. M., Lugo, D. E., Padilla, E. R., & Dunn, L. M. (1986). *Test de Vocabulario en Imágenes Peabody (TVIP)*. Minneapolis: Pearson Assessments.

²² 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.

²³ Muñoz Sandoval, A. F., Woodcock, R. W., McGrew, K. S., & Mather, N. (2005). *Batería III Woodcock-Muñoz*. Rolling Meadows, IL: Riverside Publishing.

²⁴ LeBuffe, P. A., & Naglieri, J. A. (1999). *Devereux Early Childhood Assessment, User's Guide*. Lewisville, NC: Kaplan.

²⁵ 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.

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.²⁶ The 2012-13 school year was the third year that we observed classrooms using the CLASS. Observers visit the classroom and observe for up to 6 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 teachers structure 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.

ELEMENTARY SCHOOL

Children's reading proficiency was measured using the Developmental Reading Assessment (DRA2)²⁷ and its Spanish language counterpart Evaluación del Desarrollo de la Lectura (EDL2).²⁸ 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

²⁶ Pianta, R. C., LaParo, K. M., & Hamre, B. K. (2008). Classroom Assessment Scoring System (CLASS) Manual, Pre-K. Baltimore, MD: Brookes.

²⁷ Beaver, J. M., & Carter, M. A. (2006). *The Developmental Reading Assessment, Second Edition (DRA2)*. Upper Saddle River, NJ: Pearson.

²⁸ Ruiz, O.A. & Cuesta, V. M. (2007). *Evaluación del desarrollo de la lectura*. Parsippany, NJ: Pearson Education, Inc.

children's reading level and design differentiated instruction to meet the needs of all children in their classroom.²⁹ The assessment yields a reading level for each child. In kindergarten, a reading level of 4 is considered reading on grade level.³⁰ 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.³¹ 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 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.

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. T-tests were performed to test for differences in PPVT, LWI and AP 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 rather large difference in the scores on the PPVT by primary language. Children whose primary language was English scored over 2 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. A similar pattern of findings was observed in the spring round (Table 7). For this round, once again, all three differences were statistically significant. Similar to the fall, the largest difference between the primary language groups was observed for the PPVT, almost two standard deviations in magnitude. Differences between

²⁹ *K-8 Technical Manual, Developmental Reading Assessment, Second Edition* (2009). Upper Saddle River, NJ: Pearson Education, Inc.

³⁰ Prior to the 2010-11 school year, a reading level of 3 was considered on grade level for kindergarten.

³¹ For more information about TCAP, visit <http://www.cde.state.co.us/assessment/GeneralInfo.asp>

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 two thirds of a standard deviation in magnitude.

Table 5: Weighted Descriptive Statistics for Child Outcome Measures

Variable	Fall 2012				Spring 2013			
	N	Mean	Standard Deviation	Range	N	Mean	Standard Deviation	Range
All Children								
<i>Standardized Assessments</i>								
PPVT Standard Score	200	94.09	24.42	28-145	200	99.00	21.27	27-146
WJ LWI Standard Score	200	99.60	15.20	65-197	200	104.83	13.79	59-185
WJ AP Standard Score	200	104.56	15.30	71-142	200	106.28	11.84	67-141
<i>Teacher-Rated DECA</i>								
Initiative T-Score	194 ¹	51.80	8.62	31-72	193	56.74	8.77	30-72
Self-Control T-Score	195	56.74	8.86	28-72	193	59.30	8.44	28-72
Attachment T-Score	195	49.64	8.29	28-72	193	53.31	8.73	33-72
Total Protective Factors T-Score	196	52.01	8.83	28-72	193	56.58	8.60	30-72
Behavioral Concerns T-Score	190	48.70	9.01	31-72	187	47.42	9.15	31-72
<i>Parent-Rated DECA</i>								
Initiative T-Score	--	--	--	--	198	52.27	9.08	30-72
Self-Control T-Score	--	--	--	--	198	55.48	8.83	30-72
Attachment T-Score	--	--	--	--	198	48.85	11.98	28-72
Total Protective Factors T-Score	--	--	--	--	198	51.78	9.50	28-72
Behavioral Concerns T-Score	--	--	--	--	198	55.16	9.50	28-72
Spanish-Speaking Children Only								
<i>Standardized Assessments</i>								
TVIP Standard Score	63	80.45	15.64	54-123	62	88.68	21.28	55-131
WM LWI Standard Score	61	95.23	11.96	73-118	66	100.51	12.37	72-125
WM AP Standard Score	63	89.61	16.23	52-121	66	93.19	14.10	49-122

¹Some 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 Round¹

Assessment		Primary Language					<u>I</u>	
		English			Another Language			
		N	Mean	SD	N	Mean	SD	
PPVT	Standard Score	153	102.20	20.27	45	68.95	18.79	10.48***
WJ LWI	Standard Score	153	103.00	14.26	45	88.97	13.09	6.34***
WJ AP	Standard Score	153	107.9	14.51	45	94.04	12.64	6.37***

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

¹Information about the child's primary language was missing or ambiguous for 17 children in the sample.

Table 7: Weighted English Assessment Scores by Child's Primary Language, Spring Round¹

Assessment	Primary Language						t
	English			Another Language			
	N	Mean	SD	N	Mean	SD	
PPVT Standard Score	149	106.90	16.82	49	80.59	20.60	8.09***
WJ LWI Standard Score	149	107.50	14.52	49	95.90	10.55	6.05***
WJ AP Standard Score	149	109.60	11.93	49	98.51	9.19	6.76***

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

¹Information about the child's primary language was missing or ambiguous for 17 children in the sample.

PRESCHOOL QUALITY

The 203 children in the sample were enrolled in 107 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.³²

QUALISTAR RATING AND ACCREDITATION

103 of 107 preschools were Qualistar rated. Detailed information about the quality of these preschools was provided to Clayton Early Learning from Qualistar. Two sites became eligible for DPP because they had obtained Accreditation from the National Association for the Education of Young Children (NAEYC). For these sites, the only quality information that is available is the number of stars.³³ Figure 1 presents the breakdown of programs by star level. Nearly two-thirds of the programs had earned 3 stars. Approximately one-third of programs had 4 stars. Only one preschool had 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.

³² 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 sample were enrolled. For information on the quality of all preschool programs participating in DPP during the 12-13 school year, readers are referred to the annual evaluation report prepared by Augenblick, Palaich and Associates.

³³ Providers who were accredited by NAEYC after October 2006 received a DPP Quality Rating of 4 stars. Providers accredited by the National Association of Family Child Care receive a DPP Quality Rating of 3 stars.

Figure 1: Star Level of Programs Attended by Children in the Sample (n=107 programs)

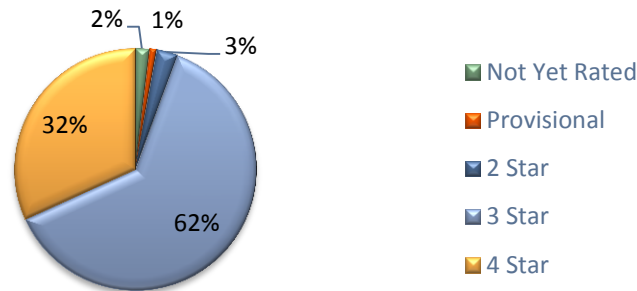


Table 8 presents descriptive statistics for the five component areas of the Qualistar rating for the 103 sites with a Qualistar rating.³⁴ Sites were strongest, on average, in the areas of Family Partnerships and Adult-to-Child Ratios and Group Size. Family Partnerships was a particularly strong area, with programs earning, on average, over 95% of the possible points for this area. While scores in this area covered a wide range (0-10), very few programs

Table 8: Qualistar Rating Components for Programs Attended by Children in the Sample (n=96 programs)

Component	Possible Range	Mean	SD	Range
Learning Environment	0-10	6.76	1.63	0-10
Family Partnerships	0-10	9.54	1.31	0-10
Training and Education	0-10	6.09	1.79	2-10
Adult-to-Child Ratios and Group Size	0-10	8.93	1.36	5-10
Accreditation	0-2	.06	0.34	0-2

earned very low scores on this component. One program earned no points for this area, one earned four 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 5 and 10. On average, programs earned slightly more than half of the possible points for Training and Education. There was considerable variability around this mean with some programs earning very few points and some earning all the points possible. Programs earned, on average, about two-thirds of the possible points for Learning Environment. Scores in this area covered quite a range as well, with some programs earning no points and others earning all of the possible points. A very small proportion of programs earned the 2 points for having earned an accreditation.

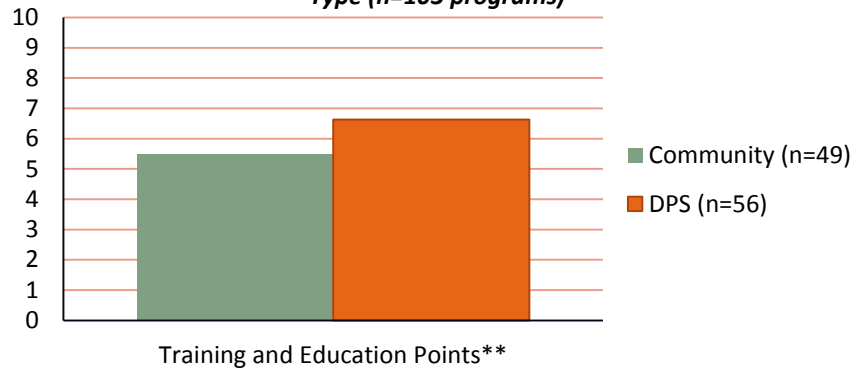
Analyses were conducted to test whether the type of provider (DPS vs. Community) was associated with the components of the Qualistar rating. The two types of programs only differed significantly in one area: Training and Education points (see Figure 2).³⁵ On average, DPS programs earned significantly more points than community preschools in this area.

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.

³⁴ More information about the five component areas of the Qualistar rating is available at: <http://www.qualistar.org/qualistar-rating-components.html>

³⁵ $t=3.29$, $df=76.78$, $p<.01$

Figure 2: Qualistar Rating Training and Education Points for Programs Attended by Children in the Sample, by Provider Type (n=103 programs)¹



**p<.01

¹Standard deviations: Community=2.09, DPS=1.25

CLASS OBSERVATIONS

Figure 3 displays the mean scores for the 125 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.

Figures 4, 5 and 6 provide some 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, over three-quarters of classrooms scored in the high range, a very small proportion of classrooms scored in the low range (below 3), and the remainder scored in the middle-range. For Instructional Support, over two-thirds of classrooms scored in the low range, nearly a third scored in the middle range, and very few classrooms scored in the high range.

Figure 3: Average CLASS Domain Scores (n=125 Classrooms)

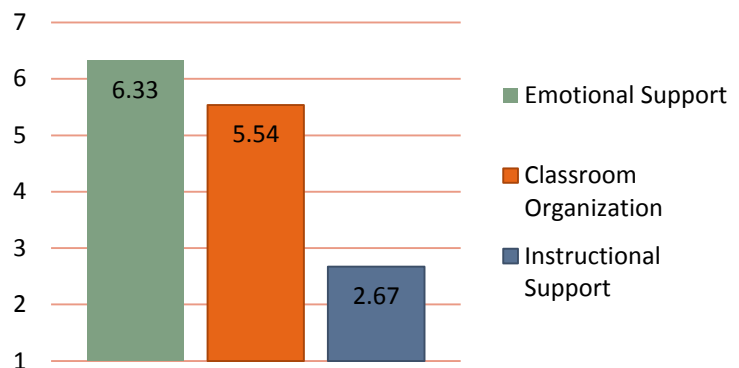


Figure 4: Distribution of Scores for Emotional Support (n=125 Classrooms)

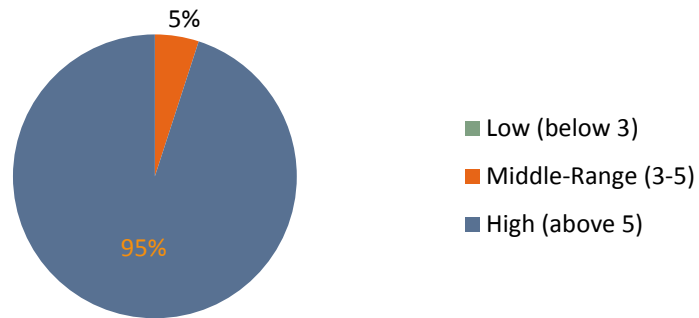


Figure 5: Distribution of Scores for Classroom Organization (n=125 Classrooms)

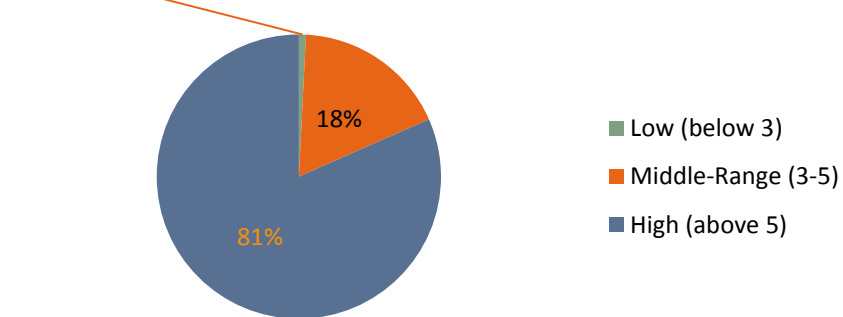
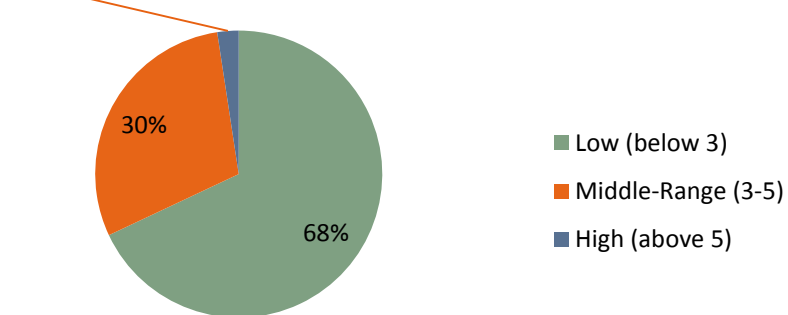


Figure 6: Distribution of Scores for Instructional Support (n=125 Classrooms)



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 significantly higher, on average, in DPS classrooms than in community-based preschool classrooms.³⁶ Differences between the two provider types

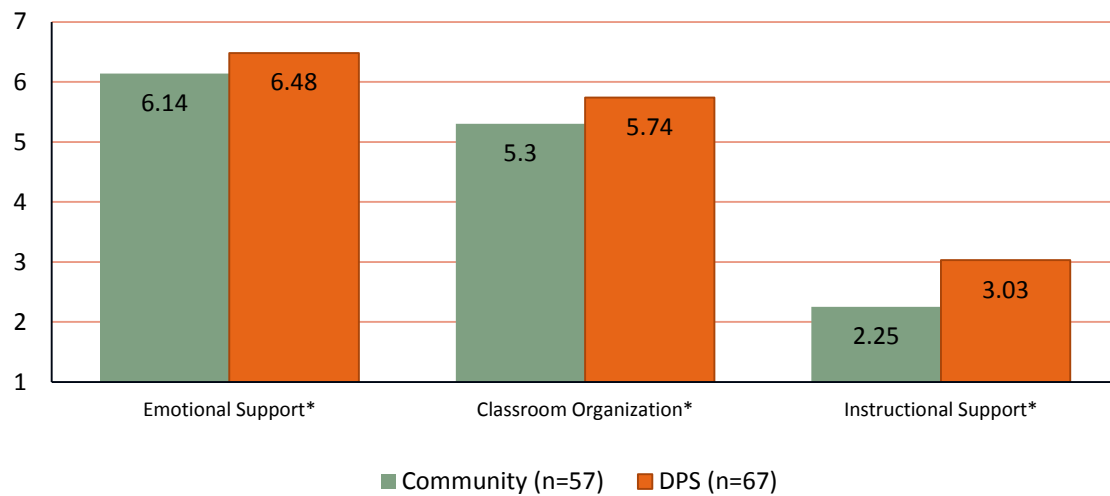
³⁶ Emotional Support— $t=3.24$, $df=93.64$, $p<.01$; Classroom Organization— $t=2.87$, $df=100.88$, $p<.01$; Instructional Support— $t=4.93$, $df=122$, $p<.0001$.

ranged from about one-third of a point in magnitude, on average, for Emotional Support to over three-quarters of a point, on average, for Instructional Support.

The ratio of children to adults during our observations varied widely. On average, classrooms had 6.8 children for every adult in the classroom. The smallest observed ratio was 1.9 children for every adult and the largest ratio was 17 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.³⁷

As with the Qualistar Rating, we conducted analyses to test for associations between CLASS domain scores and child and family background characteristics. Only one significant association was observed: between income tier and Classroom Organization.³⁸ Follow-up Tukey tests revealed that the children in Tiers 3-5 tended to be enrolled in classrooms with higher levels of Classroom Organization than children whose parents opted out of reporting their income. The difference between these groups was, on average, close to two-thirds of a point.³⁹

Figure 7: CLASS Domain Scores by Provider Type



* $p < .05$

^Standard deviations: Emotional Support—Community=0.69, DPS=0.45; Classroom Organization—Community=0.95, DPS=0.70; Instructional Support—Community=0.78, DPS=0.95

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

³⁷ Ratio with Emotional Support, $r = -.004$; ratio with Classroom Organization, $r = -.09$; ratio with Instructional Support, $r = .03$; all non-significant.

³⁸ $F(3,156) = 2.98$, $p < .05$

³⁹ Means for Classroom Organization: Tier 1, mean=5.47, sd=.80; Tier 2, mean=5.73, sd=.85; Tiers 3-5, mean=5.81, sd=.82; Tier 6 (income not reported), mean=5.18, sd=.85

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.

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.

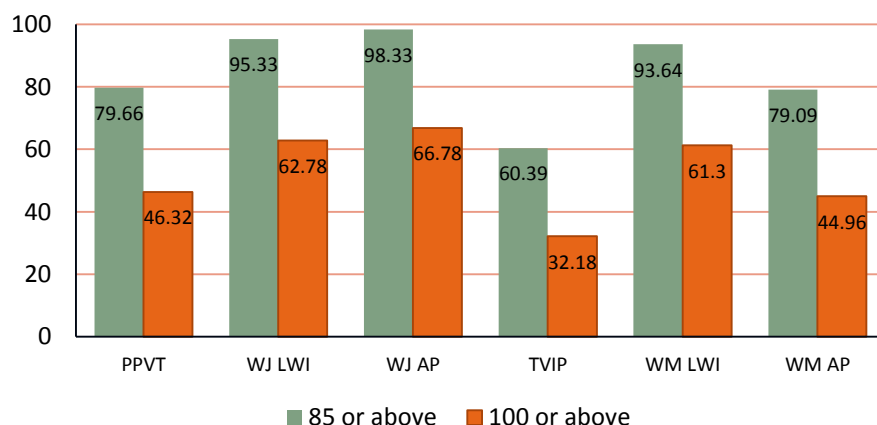
Not surprisingly, follow-up analyses revealed that the likelihood of scoring 85 or above on these assessments was strongly associated with children's primary language. About 89% of children whose primary language was English scored 85 or above on the PPVT as compared with about half of children whose primary language was not English (51%).⁴⁰ A similar, but less pronounced 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 87% of children whose primary language was not English.⁴¹ Nearly all children whose primary language was English (99.6%) scored 85 or above on WJ AP compared with 95% of children with another primary language.⁴²

A more pronounced pattern of results emerged when a score of 100 was used as the cutoff. For PPVT, nearly two-thirds of children whose primary language was English (63%) earned a score of 100 or greater as compared with just 12% of children with another primary language.⁴³ For WJ LWI, 77% of children whose primary language was English scored 100 or greater as compared with 39% of children with another primary language.⁴⁴

Finally, for WJ Applied Problems, 77% of children whose primary language was English earned scores of 100 or above compared with 47% of children whose primary language was something other than English.⁴⁵

For assessments administered in Spanish,

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



⁴⁰ $\chi^2_1=33.25, p<.0001$

⁴¹ $\chi^2_1=10.89, p<.001$

⁴² $\chi^2_1=5.30, p<.05$

⁴³ $\chi^2_1=38.94, p<.0001$

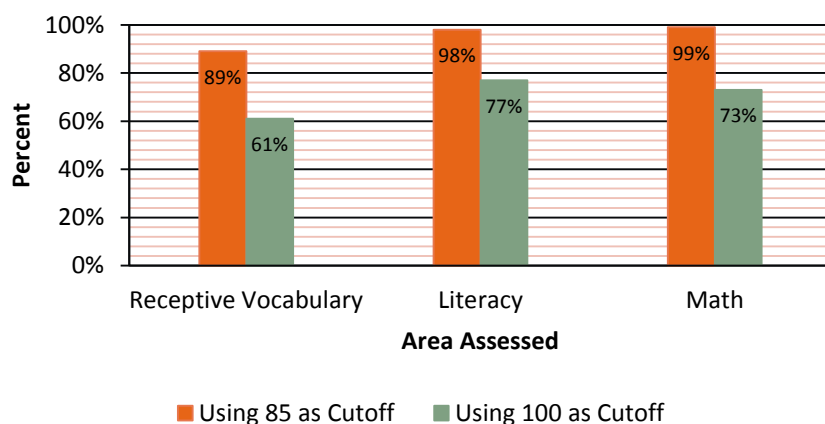
⁴⁴ $\chi^2_1=24.99, p<.0001$

⁴⁵ $\chi^2_1=15.09, p<.0001$

scores were once again stronger for LWI and Applied Problems than for vocabulary (TVIP). Over nine-tenths of children scored 85 or above on WM LWI and over three quarters of children scored 85 or above on Applied Problems. In contrast, just 60% of children scored 85 or above on the TVIP (see Figure 8). About a third of children scored 100 or above on the TVIP, about 60% scored 100 or above on the WM LWI, and close to half 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 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. Close to 90% 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, over half of children met or exceeded this benchmark for vocabulary; approximately three-quarters met or exceeded this benchmark for both literacy and math.

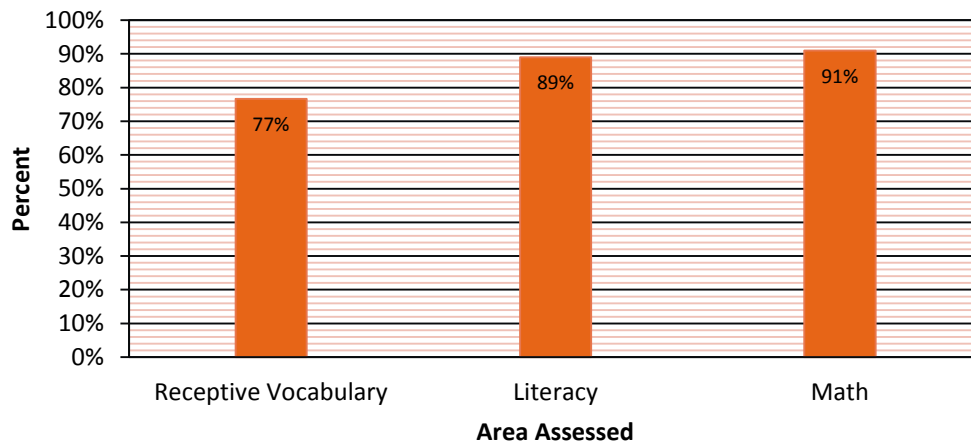
Figure 9: Weighted Percent of Children Scoring in the Average Range or Above on Spring Standardized Assessments in Spanish or English



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’d expect only half of children to score above this cutoff. The leadership of DPP and the author of this report struggled with the use of both of these cutoffs as indicators of whether children were ready for school. We felt that the cutoff of 85 was 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, over 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.

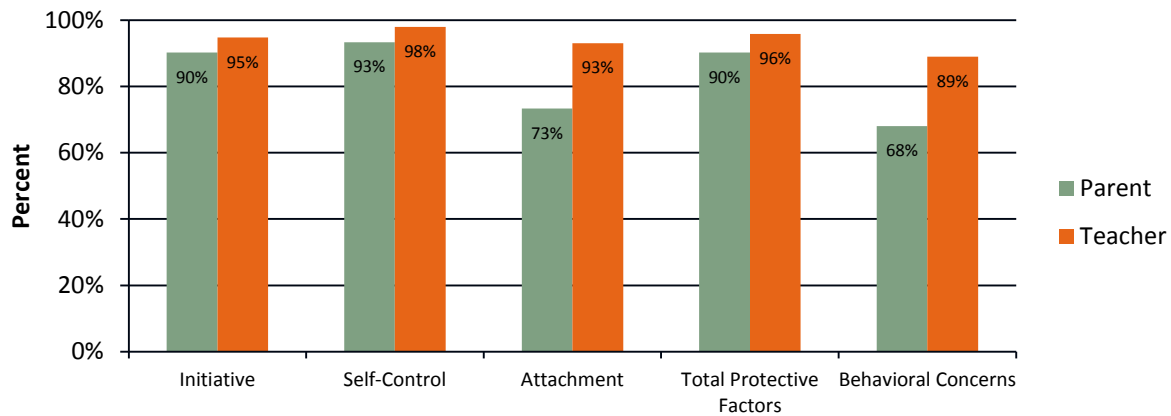
Figure 10: Weighted Percent of Children Scoring 92.5 or Above on Spring Standardized Assessments in Spanish or English



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, the vast 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 about two-thirds in the typical range for Behavioral Concerns. Teachers rated over 90% of children in the typical or strength range for Initiative, Self-Control, Attachment and Total Protective Factors and close to 90% in the typical range on Behavioral Concerns.

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



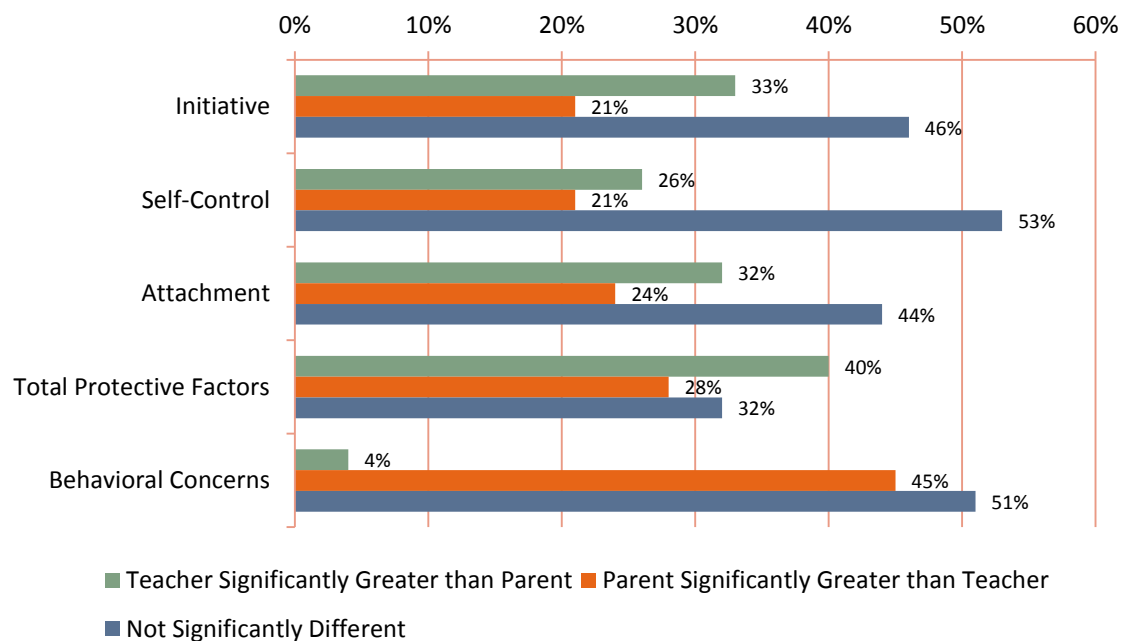
We examined the differences between teachers' and parents' ratings using guidelines from the authors of the DECA. The authors developed these guidelines to help users distinguish between differences in scores due to measurement error and differences that are likely due to a meaningful difference between 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 teachers' and parents' reports, 4.40 ($sd=11.3$), did not exceed this threshold, indicating that, on average, teachers' and parents' ratings did not differ. As displayed in Figure 12, for almost half of children, teachers' and parents' reports did not significantly differ. For about a third of the sample, the teacher's rating was significantly greater than the parent's rating. For about a fifth of the sample, the parent's rating was significantly greater than teacher's 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, 3.7 ($sd=10.9$), did not exceed this threshold, indicating that, on average, teachers' and parents' ratings did not differ. As displayed in Figure 12, for over half of children, teachers' and parents' reports did not significantly differ. For about a quarter of children, teachers rated children significantly higher than did parents. About a fifth of children 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, 4.4 ($sd=14.7$), did not exceed this threshold. Once again, for nearly half of the children, the parent's score and teacher's score 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 4.7 ($sd=12.0$), which was below that threshold. However, for 40% of children, the teacher's rating was significantly greater than the parent's rating (see Figure 12). For close to a third of children, there was not a significant difference between raters. For slightly more than a quarter 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 7.8 ($sd=12.3$), which did not reach this threshold. For about half 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 far more common for the parent to report significantly more Behavioral Concerns than vice versa.

In sum, for Initiative, Self-Control and Attachment, teachers and parents made similar ratings of about 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, the most common pattern was for parents' and teachers' ratings to be similar. When the ratings were different, it was most often because the parent rated the child significantly higher than the teacher.

CHANGE IN ASSESSMENTS 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, and rather large, increases in PPVT and WJ LWI. On average, scores on both of these assessments increased about one-third of a standard deviation. For Applied Problems, the increase was significant but small in magnitude, slightly over a tenth of a standard deviation. 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. For assessments administered in Spanish, there was a large increase in TVIP scores over the course of the school year, about half of standard deviation. A large increase, on average, was also observed for WM LWI, about a third of a standard deviation. Finally, on average, scores for WM AP increased about one-fifth of a standard deviation.

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	189	94.36 (24.64)	99.00 (21.55)	5.36***
WJ-LWI	189	99.81 (15.31)	104.98 (14.08)	7.53***
WJ-AP	189	104.59 (15.43)	106.35 (12.07)	2.27*
Standardized Assessments—Spanish				
TVIP	57	81.36 (15.76)	88.89 (21.93)	4.14***
WM-LWI	58	95.52 (12.08)	101.09 (12.56)	3.56**
WM-AP	60	89.97 (16.13)	93.18 (14.63)	2.07*
Teacher Survey				
Initiative T-Score ¹	180	51.58 (8.70)	56.65 (8.92)	7.52***
Self-Control T-Score	181	56.72 (8.86)	59.23 (8.56)	3.92**
Attachment T-Score	181	49.63 (8.36)	53.28 (8.80)	4.79***
Total Protective Factors T-Score	182	51.88 (8.93)	56.48 (8.68)	6.59***
Behavioral Concerns T-Score	172	49.06 (8.97)	47.82 (9.24)	2.40*

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

¹Some 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.

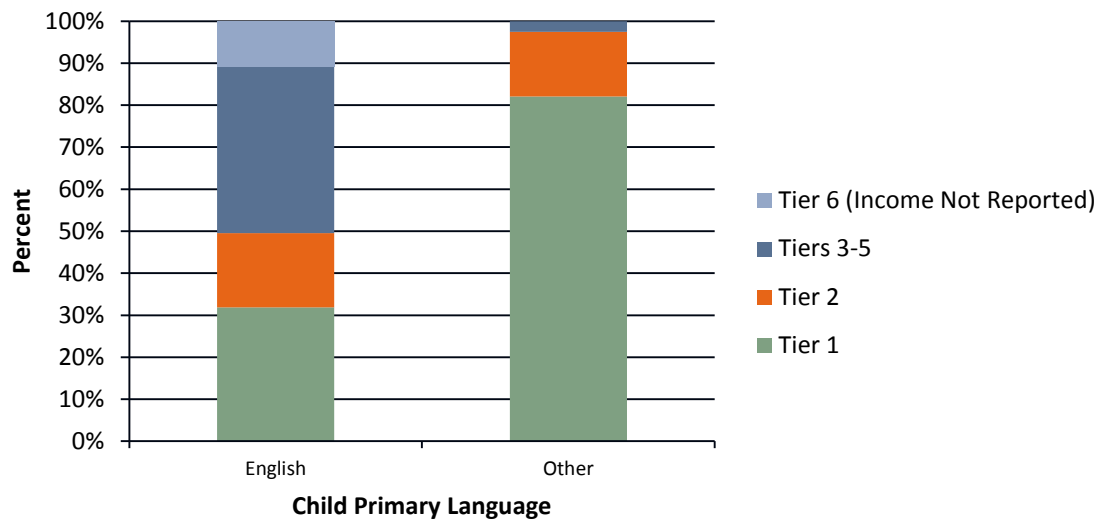
Significant improvements were also observed in all of the teachers' ratings on the DECA over the course of the school year. Change over time was significant and positive for all of the Protective Factors (Initiative, Self-Control, and Attachment, as well as Total Protective Factors). Teachers also reported significant decreases in Behavioral Concerns over the course of the school year. Of particular note are the increases in Initiative and Total Protective Factors (close to half of a standard deviation).

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).⁴⁶ It is important to note that these two background characteristics, income tier and child's primary language, are strongly associated (see Figure 13).⁴⁷ Nearly all children whose primary language is not English are from Tiers 1 or 2 whereas only about 50% 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 ANOVAs⁴⁸ was conducted with income tier group predicting scores over time on assessments administered in English and Spanish as well as teacher-rated DECA. There were significant interactions between income tier group and time for PPVT,⁴⁹ WM LWI,⁵⁰ and WM Applied Problems.⁵¹ Results of these analyses are depicted in Figures 14-16. For WJ LWI, WJ Applied Problems, TVIP, and all of the teacher-rated DECA subscales, 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.

Figure 14 shows average PPVT scores over time, by income tier group. The difference in average scores by tier groups is striking, with children in the lower income tiers scoring much lower on average than children in

⁴⁶ 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 no children assessed in Spanish fell into tier 6.

⁴⁷ $\chi^2_3=45.37, p<.0001$

⁴⁸ 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.

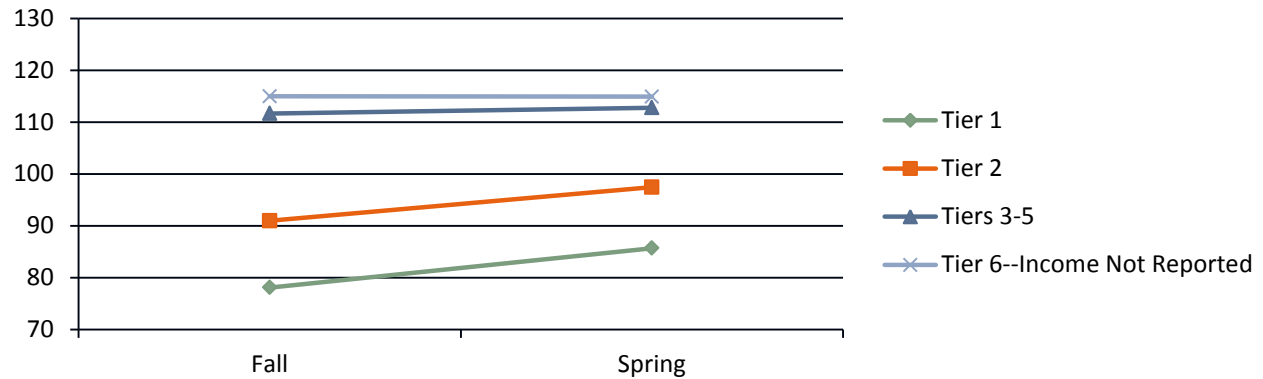
⁴⁹ $F(3,185)=4.76, p<.01$

⁵⁰ $F(1,46)=4.45, p<.05$

⁵¹ $F(1,48)=11.50, p<.01$

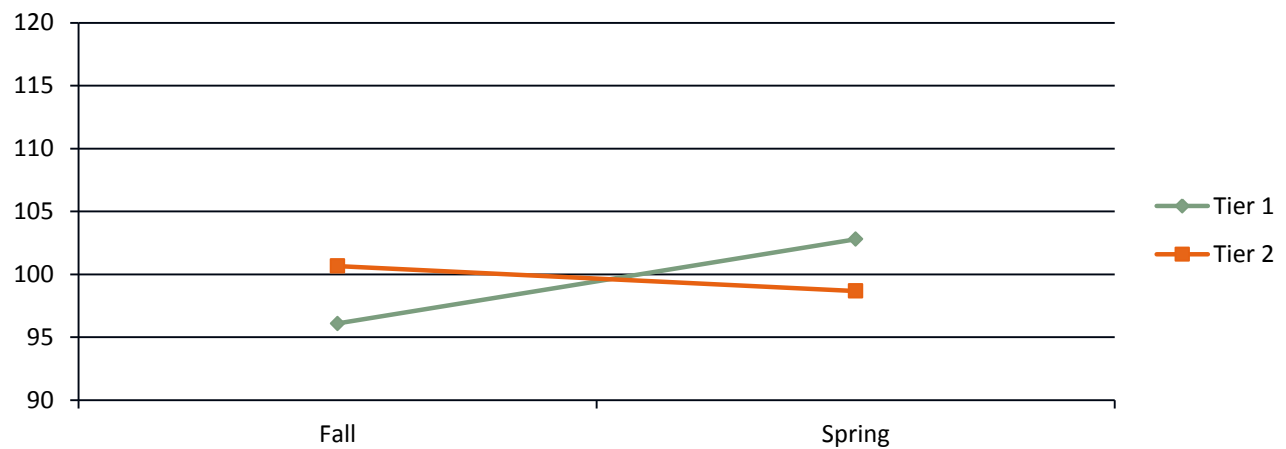
the higher tier groups. Follow-up Tukey⁵² tests revealed that children from Tier 1 increased significantly more over time than children in Tiers 3-5. Pairwise comparisons of the other groups were not significant.

Figure 14: Weighted PPVT Standard Scores over Time, by Income Tier Group¹



¹Standard Deviations: Tier 1: Fall=20.07, Spring=18.62; Tier 2: Fall=19.34, Spring=20.70; Tiers 3-5: Fall=18.73, Spring=15.38; Tier 6: Fall=13.90, Spring=12.98.

Figure 15: Weighted WM Letter-Word Identification Standard Scores over Time, by Income Tier Group¹

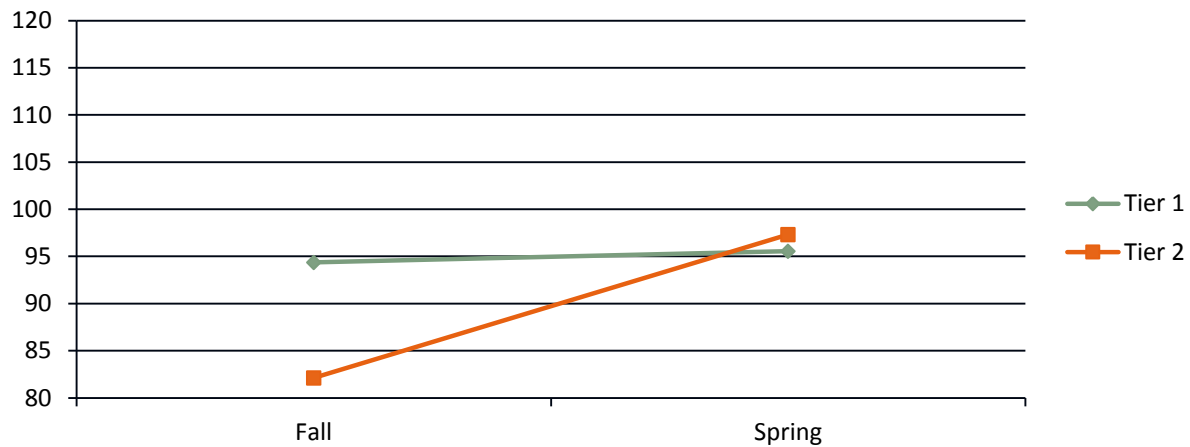


¹Standard Deviations: Tier 1: Fall=12.92, Spring=12.06; Tier 2: Fall=4.91, Spring=11.38.

Results of the analysis of WM Letter-Word Identification are presented in Figure 15. Children in Tier 1 increased significantly more over the course of the school year than children in Tier 2, who decreased slightly over the course of the year. Results of the analysis of WM Applied Problems are presented in Figure 16. Children in Tier 2, on average, increased over time, while scores for the children in Tier 1 remained relatively flat over time, on average. Both groups had similar scores, on average, at the end of the year.

⁵² Results of ANOVA simply tell you that there is a difference between the specified groups on the outcome variable. When more than two groups are specified, follow-up tests are required to determine which pairs of groups are significantly different. Tukey tests are one particularly conservative type of follow-up test.

Figure 16: Weighted WM Applied Problems Standard Scores over Time, by Income Tier Group¹

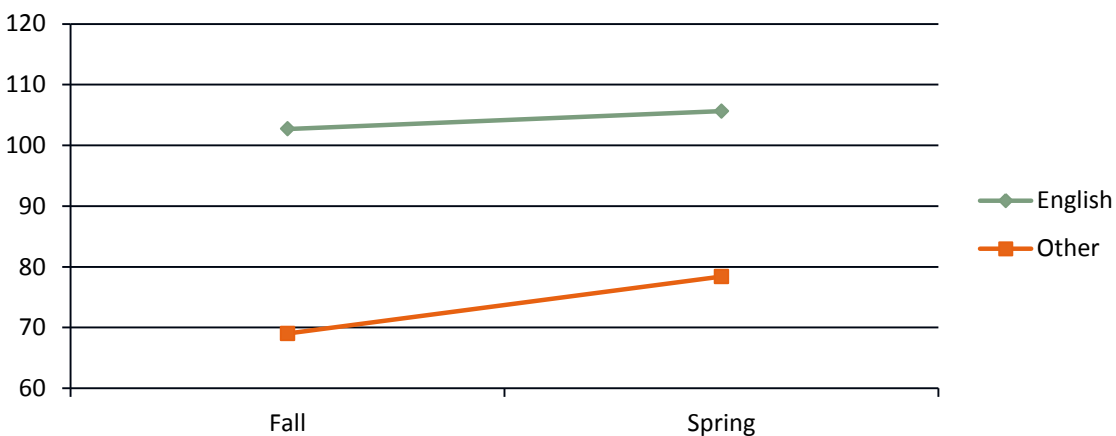


¹Standard Deviations: Tier 1: Fall=12.32, Spring=10.86; Tier 2: Fall=20.34, Spring=14.19.

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.⁵³ There were significant interactions between primary language group and time for PPVT,⁵⁴ DECA Initiative,⁵⁵ DECA Attachment,⁵⁶ and DECA Total Protective Factors.⁵⁷ Results of these analyses are presented in Figures 17-20.

Figure 17: Weighted PPVT Scores over Time, by Child Primary Language¹



¹Standard Deviations: English: Fall=20.25, Spring=16.67; Other: Fall=18.97, Spring=22.54.

⁵³ 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.

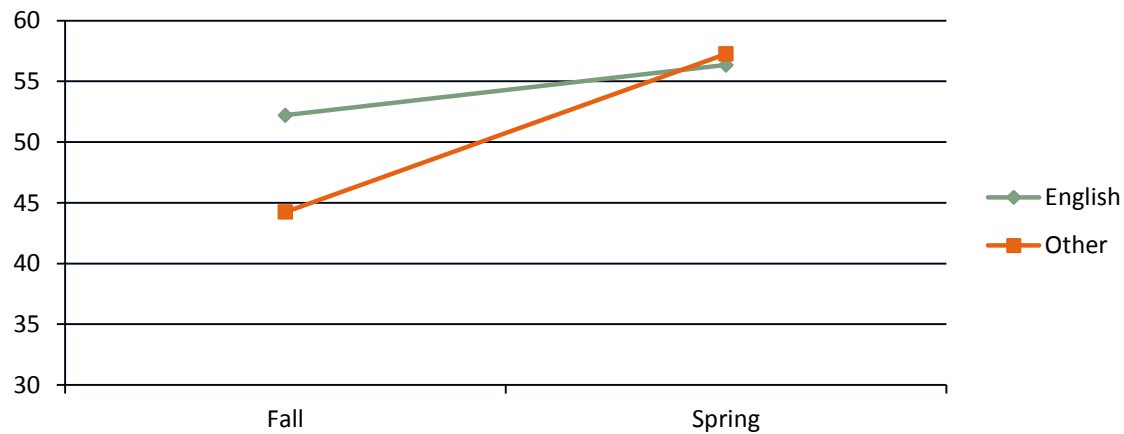
⁵⁴ $F(1,185)=10.76, p<.01$

⁵⁵ $F(1,176)=8.72, p<.05$

⁵⁶ $F(1,185)=4.07, p<.05$

⁵⁷ $F(1,178)=7.57, p<.01$

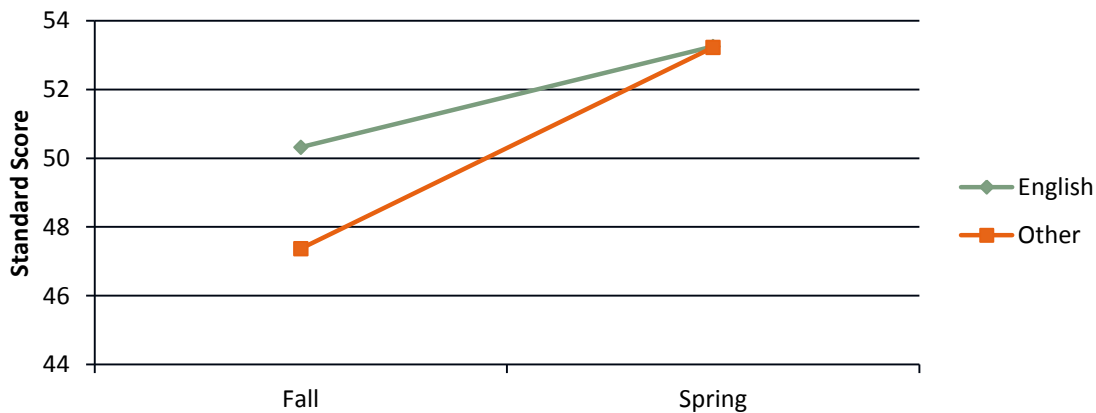
Figure 18: Weighted Teacher-Rated Initiative Scores over Time, by Child Primary Language¹



¹Standard Deviations: English: Fall=8.78, Spring=9.09; Other: Fall=7.77, Spring=8.37.

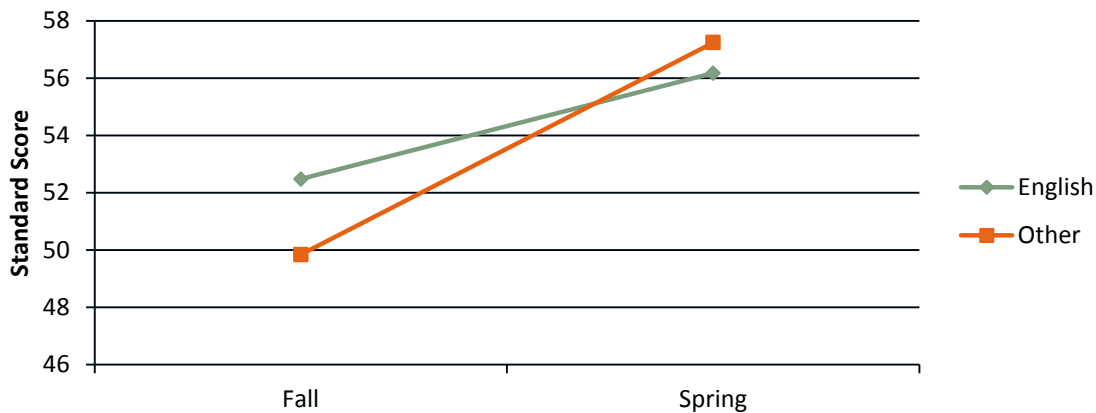
For all four assessments, the general pattern was that children whose primary language was not English tended to show larger increases in scores from fall to spring than their counterparts who spoke English as their primary language. For PPVT (see Figure 17), children whose primary language was English increased only slightly, while their counterparts with another primary language increased an average of about two-thirds of a standard deviation. For teacher rated DECA Initiative (see Figure 18), children whose primary language was English started the year with above average scores and increased about a half of a standard deviation over the course of the year. Children with another primary language increased even more over time, about one and a half standard deviations. For teacher-rated DECA Attachment (see Figure 19), children who spoke English as their primary language increased about one-third of a standard deviation over the course of the year. For children with another primary language, the gain was about twice that magnitude. For teacher-rated DECA Protective Factors, a similar pattern was observed. Children who spoke English as their primary language increased about a third of a standard deviation over the course of the year, while children who spoke another primary language increased about three-fourths of a standard deviation. For all three of these scales on the DECA, the gap that was observed in the fall was closed by the spring. For LWI, Applied Problems, Self-Control, and Behavioral Concerns, the child primary language by time interaction was non-significant, indicating that children progressed similarly in these areas over the course of their preschool year, regardless of their primary language.

Figure 19: Weighted Teacher-Rated Attachment Scores over Time, by Child Primary Language¹



¹Standard Deviations: English: Fall=8.07, Spring=8.58; Other: Fall=8.93, Spring=9.73.

Figure 20: Weighted Teacher-Rated Total Protective Factors Scores over Time, by Child Primary Language¹



¹Standard Deviations: English: Fall=8.82, Spring=8.60; Other: Fall=9.12, Spring=9.13.

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 and math), there were two significant correlations. Classroom Organization and Instructional Support were significantly and positively associated with WJ Letter-Word Identification (Classroom Organization, $r=.23$, $p<.01$; Instructional Support, $r=.30$, $p<.001$). That is, after taking into account children’s literacy skills in the fall, higher levels of Classroom Organization and Instructional Support were associated with higher vocabulary skills as assessed in English in the spring. There were

no significant correlations for Spanish assessments. For the teacher DECA ratings, there was a significant, positive association between Instructional Support and Total Protective Factors ($r=.23$, $p<.01$). That is, after controlling for Total Protective Factors in the fall, higher levels of Instructional Support were associated with higher ratings of Total Protective Factors in the spring. The association between Instructional Support and Behavioral Concerns was non-significant, as were the correlations between the Classroom Organization and the two DECA subscales.

RESULTS: ELEMENTARY SCHOOL

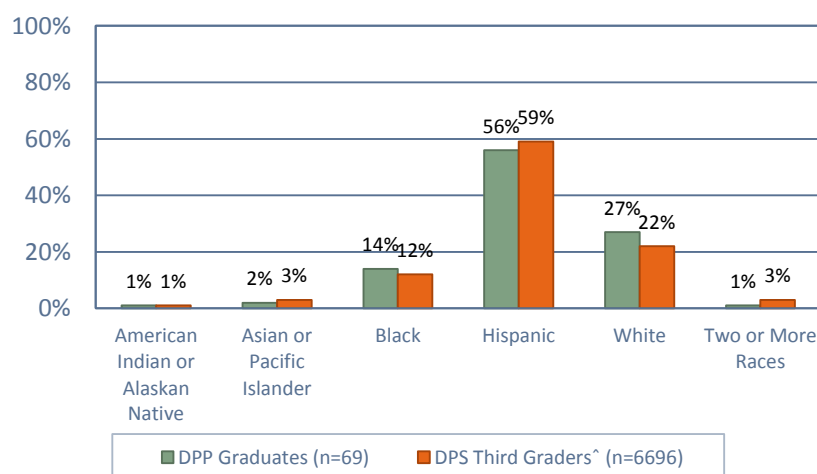
HOW SIMILAR ARE DPP GRADUATES TO THE POPULATION OF CHILDREN IN THE DISTRICT AS A WHOLE?

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.⁵⁸ 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 1

Figure 21 displays a comparison of the racial/ethnic composition of the sample of Cohort 1 DPP graduates with TCAP reading assessment data for school year 12-13 and the population of children enrolled in third grade as of fall 2012. The sample was fairly similar to the district as a whole in terms of race and ethnicity.

Figure 21: Race/Ethnicity of Cohort 1 DPP Graduates and All Third Graders Enrolled in DPS in Fall 2012*

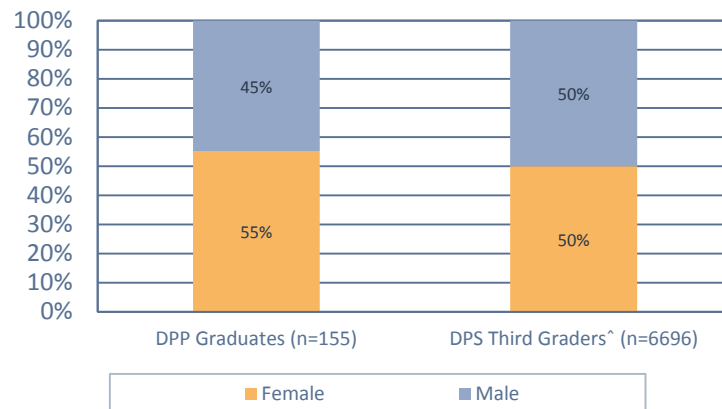


*This figure includes only DPP Graduates for whom TCAP reading data were available for the 12-13 school year and omits 3 children with missing or ambiguous race/ethnicity data in the database of DPP children provided to Clayton by ACS.

^The group of DPS Third Graders includes the DPP graduates. Figures do not sum to 100 because of rounding error.

⁵⁸ Available at: <http://www.cde.state.co.us/cdereval/pupilcurrentschool>

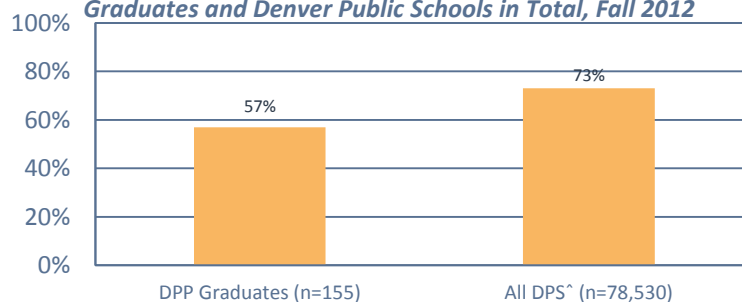
Figure 22: Gender of Cohort 1 DPP Graduates and All Third Graders Enrolled in DPS in Fall 2012*



*This chart includes only Cohort 1 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 22 presents the gender composition of Cohort 1 DPP graduates with reading assessment data and all third graders in DPS as of fall 2012. The district as a whole was split evenly between the genders. Cohort 1 DPP Graduates with spring 2013 reading assessment data were split a bit less evenly, with slightly more girls than boys, but still rather close to an even split.

Figure 23: Free/Reduced Lunch Status of Cohort 1 DPP Graduates and Denver Public Schools in Total, Fall 2012



^The group All DPS includes the DPP graduates.

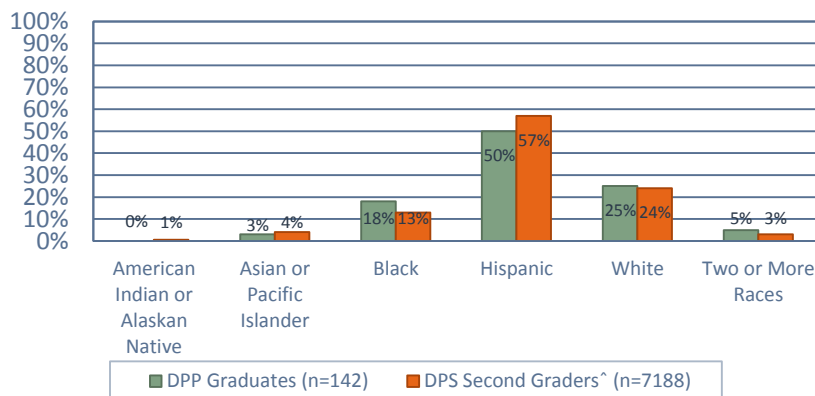
Finally, Figure 23 presents the proportion of children eligible for free or reduced lunch for the district as a whole and for the sample of DPP graduates. CDE does not provide free or reduced lunch data by grade level. As a result the comparison group in this figure is the entire district, from kindergarten through grade 12. Nearly three-quarters of the district as a whole qualified for free or reduced lunch. In contrast, only 57% of Cohort 1 graduates with reading assessment data qualified, suggesting that the DPP Cohort 1 sample is slightly wealthier than the district as a whole.

In sum, Cohort 1 children with spring 2013 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 2

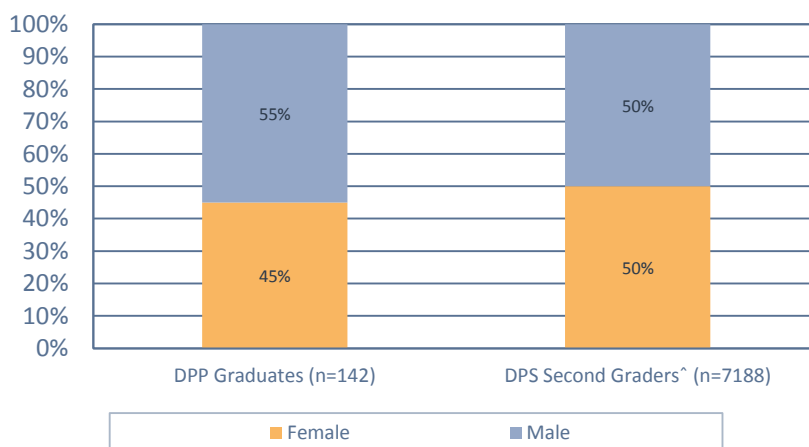
Figure 24 displays a comparison of the racial/ethnic composition of the sample of Cohort 2 DPP graduates with reading assessment data and the population of children enrolled in second grade in DPS as of fall 2012. The sample of DPP graduates includes more black children and fewer Hispanic children than the district as a whole. Otherwise, the racial and ethnic compositions of the two groups are similar.

Figure 24: Race/Ethnicity of Cohort 2 DPP Graduates and All Second Graders Enrolled in DPS in Fall 2012*



*This figure includes only Cohort 2 DPP Graduates for whom reading assessment data were available for the 12-13 school year.
 *The group of DPS Second Graders includes the DPP graduates. Percentages do not sum to 100 because of rounding error.

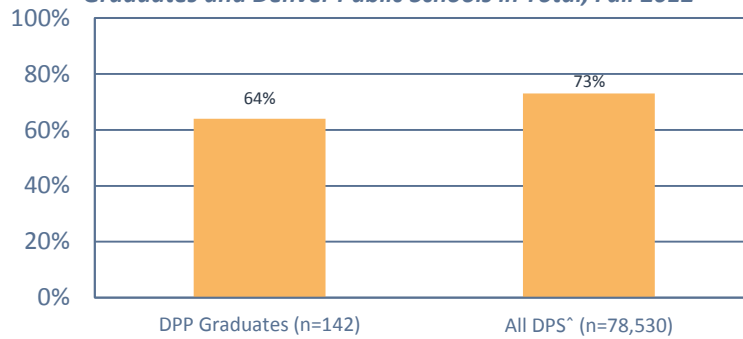
Figure 25: Gender of Cohort 2 DPP Graduates and All Second Graders Enrolled in DPS in Fall 2012*



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

Figure 25 presents the gender composition of Cohort 2 DPP graduates with reading assessment data and all second graders in DPS as of fall 2012. The district as a whole was split evenly between the genders. Cohort 2 DPP Graduates with spring 2013 reading assessment data were split a bit less evenly, with slightly more boys than girls, but still rather close to an even split.

Figure 26: Free/Reduced Lunch Status of Cohort 2 DPP Graduates and Denver Public Schools in Total, Fall 2012

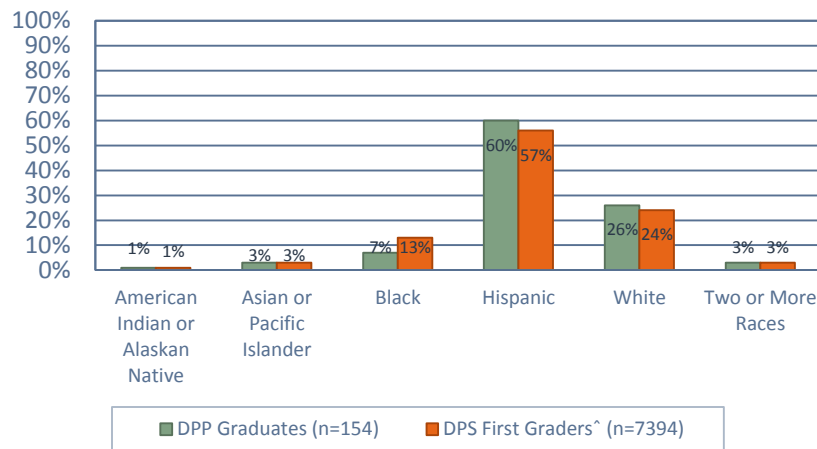


[^]The group All DPS includes the DPP graduates.

Finally, Figure 26 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 64% 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 2013 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 2 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.

Figure 27: Race/Ethnicity of Cohort 3 DPP Graduates and All First Graders Enrolled in DPS in Fall 2012*



^{*}This figure includes only Cohort 3 DPP Graduates for whom reading assessment data were available for the 12-13 school year. Percentages for DPP Graduates are weighted.

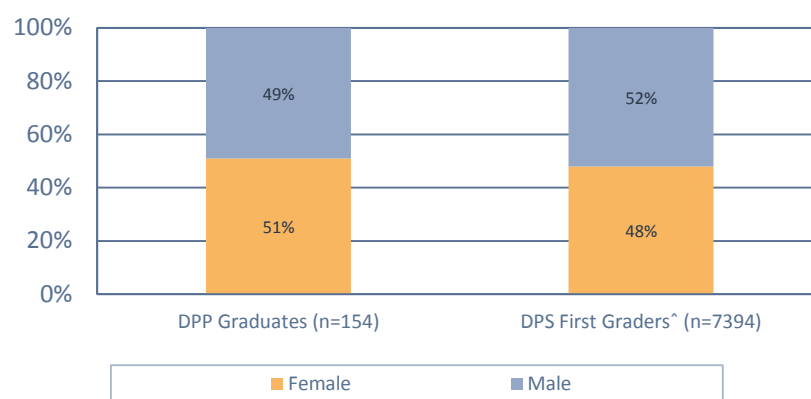
[^]The group of DPS First Graders includes the DPP graduates.

COHORT 3

Figure 27 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 first grade in DPS as of fall 2012. The racial and ethnic compositions of the two groups are remarkably similar.

Figure 28 presents the gender composition of Cohort 3 DPP graduates with reading assessment data and all first graders in DPS as of fall 2012. The district as a whole was split very close to evenly between the genders, with slightly more males than females. Cohort 3 DPP Graduates with spring 2013 reading assessment data were also split quite evenly, but with slightly more females than males

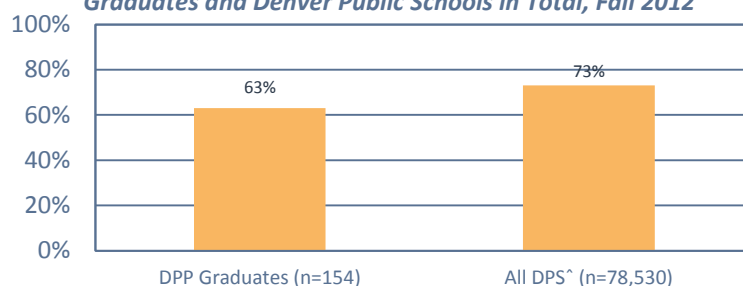
Figure 28: Gender of Cohort 3 DPP Graduates and All First Graders Enrolled in DPS in Fall 2012*



*This figure includes only Cohort 3 DPP Graduates for whom reading assessment data were available for the 12-13 school year.
^The group of DPS First Graders includes the DPP graduates.

Figure 29 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 3 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 29: Free/Reduced Lunch Status of Cohort 3 DPP Graduates and Denver Public Schools in Total, Fall 2012



^The group All DPS includes the DPP graduates.

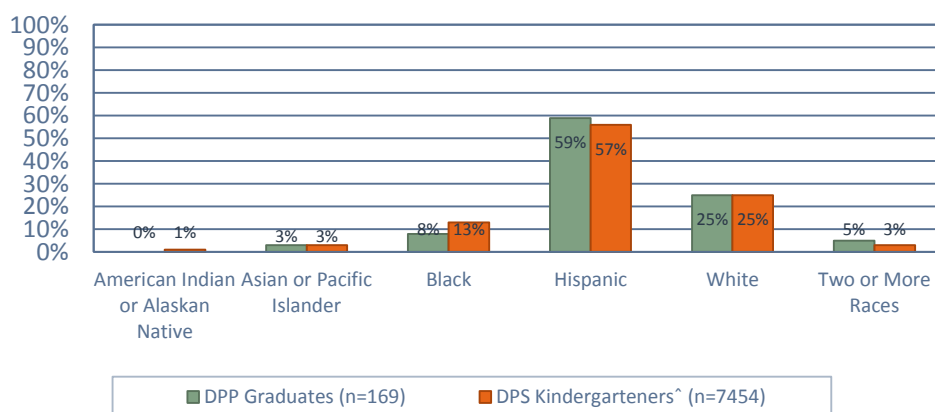
In sum, Cohort 3 children with spring 2013 reading assessment data were similar to the district in terms of their ethnic and gender composition. A smaller proportion of children in Cohort 3 qualified for free or reduced

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 4

Figure 30 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 kindergarten in DPS as of fall 2012. The group of Cohort 4 DPP graduates includes fewer black children than the district as a whole. Otherwise the racial and ethnic compositions of the two groups are remarkably similar.

Figure 30: Race/Ethnicity of Cohort 4 DPP Graduates and All Kindergarteners Enrolled in DPS in Fall 2012*

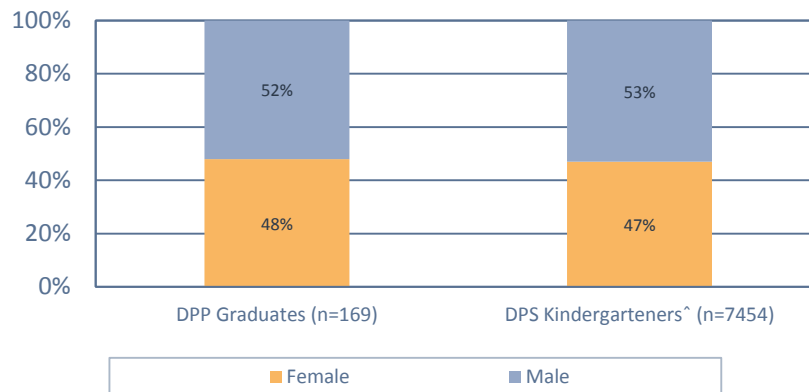


*This figure includes only Cohort 4 DPP Graduates for whom reading assessment data were available for the 12-13 school year. Percentages for DPP Graduates are weighted. Percentages do not sum to 100 because of rounding error.

^The group of DPS Kindergarteners includes the DPP graduates.

Figure 31 presents the gender composition of Cohort 4 DPP graduates with reading assessment data and all kindergarteners in DPS as of fall 2012. Both the group of Cohort 4 graduates with spring 2013 reading assessment data and the district as a whole were split very close to evenly between the genders.

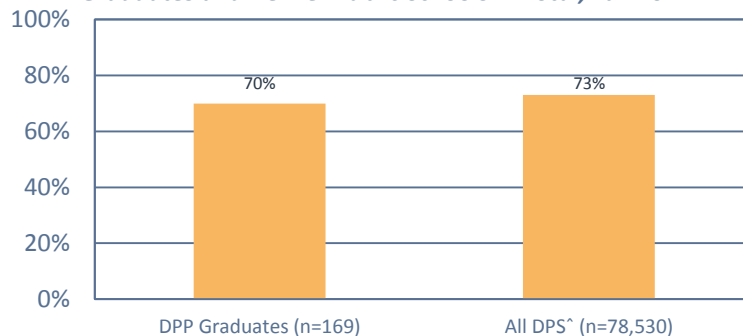
Figure 31: Gender of Cohort 4 DPP Graduates and All Kindergarteners Enrolled in DPS in Fall 2012*



*This figure includes only Cohort 4 DPP Graduates for whom reading assessment data were available for the 12-13 school year.
 ^The group of DPS Kindergarteners includes the DPP graduates.

Figure 32 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 just slightly lower than the district as a whole, which approached three-fourths qualifying for free or reduced lunch.

Figure 32: Free/Reduced Lunch Status of Cohort 4 DPP Graduates and Denver Public Schools in Total, Fall 2012



^The group All DPS includes the DPP graduates.

In sum, Cohort 4 children with spring 2013 reading assessment data were similar to the district in terms of their ethnic and gender composition. They were also similar in terms of the proportion of children who qualified for free or reduced lunch. Any differences between DPP children and the district as a whole may be due to the DPP program or other unmeasured factors.

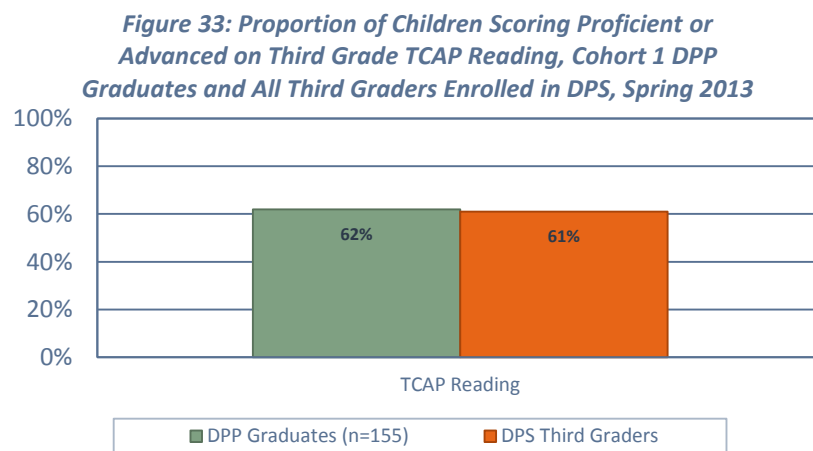
WHAT IS THE OVERALL READING PROFICIENCY OF DPP GRADUATES IN THE EARLY ELEMENTARY YEARS? HOW DOES THIS COMPARE TO THE DISTRICT AS A WHOLE?

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 1, second grade for Cohort 2, first grade for Cohort 3, and kindergarten for Cohort 4). 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 1

Figure 33 displays the proportion of Cohort 1 DPP graduates who scored proficient or advanced on the third grade TCAP reading assessment. All children were assessed in English. Sixty-two percent of Cohort 1 DPP graduates scored proficient or advanced, compared with 61% of the district as a whole.

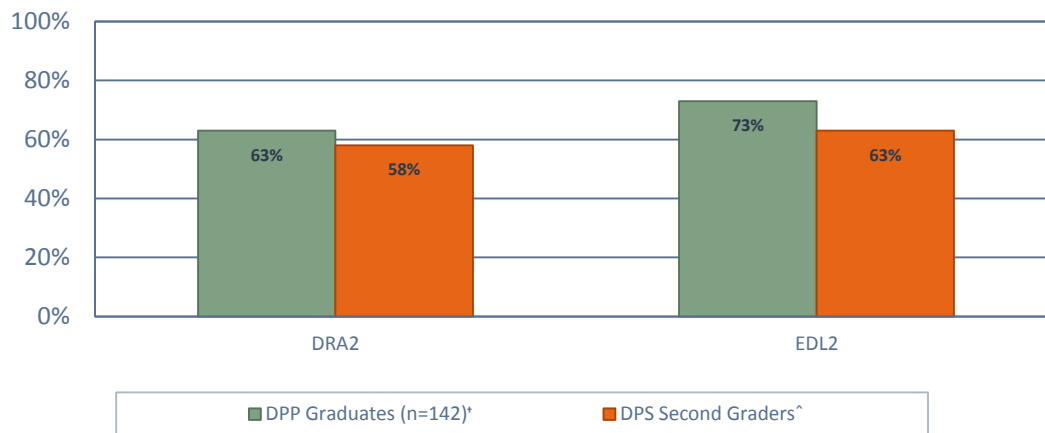


*The group of DPS third graders includes the DPP graduates.

COHORT 2

Figure 34 displays the proportion of Cohort 2 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. Nearly two-thirds 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 58% of second graders in the district as a whole. Only 28 DPP graduates were assessed using the EDL2. Of these, nearly three-quarters were reading on grade level compared to nearly two-thirds of the second graders assessed with EDL2 in the district as a whole.

Figure 34: Proportion of Children Reading On Grade Level in the Spring of the Second Grade Year, Cohort 2 DPP Graduates and All Second Graders Enrolled in DPS, Spring 2013*



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

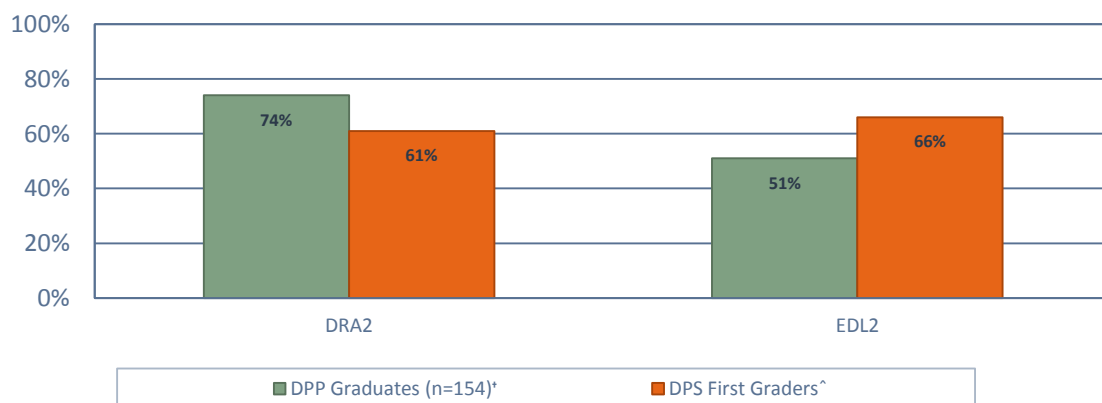
^The group of DPP Graduates includes 114 children assessed with the DRA2 and 28 assessed with the EDL2.

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

COHORT 3

Figure 35 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 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 (61%). About half 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 35: Proportion of Children Reading On Grade Level in the Spring of First Grade Year, Cohort 3 DPP Graduates and All First Graders Enrolled in DPS, Spring 2013*



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

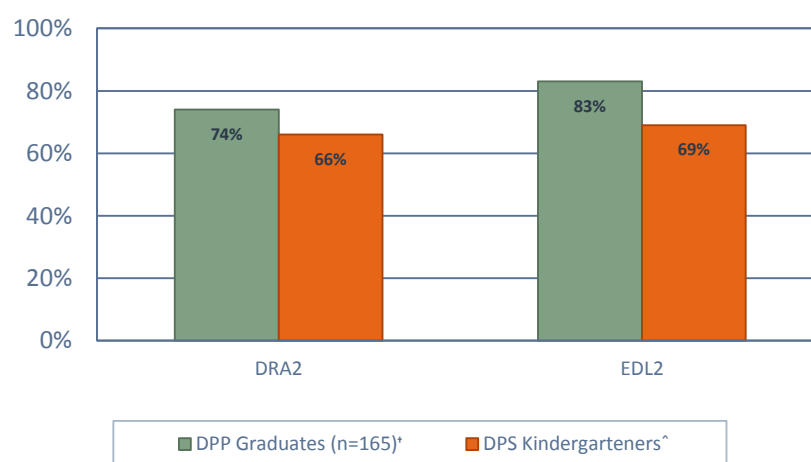
^The group of DPP Graduates includes 108 children assessed with the DRA2 and 46 assessed with the EDL2. Analyses are weighted.

^The group of DPS First Graders includes the DPP graduates.

COHORT 4

Figure 36 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 kindergarteners 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 kindergarten. This exceeds the proportion reading at or above grade level in the district as a whole (66%). Similarly, 83% of DPP graduates assessed in Spanish using the EDL2 were reading at or above grade level at the end of kindergarten. In contrast, just 69% of children in the district as a whole were reading at or above grade level as assessed by the EDL2.

Figure 36: Proportion of Children Reading On Grade Level in the Spring of Kindergarten Year, Cohort 4 DPP Graduates and All Kindergarteners Enrolled in DPS, Spring 2013*



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

^The group of DPP Graduates includes 123 children assessed with the DRA2 and 43 assessed with the EDL2. Analyses are weighted.

*The 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 1

As with the previous question, our analysis focused only on the children who were in the expected grades. In spring 2013, 155 Cohort 1 children were enrolled in third grade and had TCAP reading assessment data. Of these, 90 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 ($\chi^2=12.98$, $p<.001$). A one point increase in PPVT-4 standard score in preschool is associated with a 9% increase in the odds of scoring proficient or advanced on third grade TCAP. Letter-Word Identification also significantly predicted third grade TCAP proficiency ($\chi^2=12.26$, $p<.001$). For every one point increase in Letter-Word Identification standard score, there is a 12% increase in the odds of scoring proficient on the third grade TCAP. Finally, Applied Problems was a significant predictor of third grade reading proficiency ($\chi^2=9.95$, $p<.01$). For every one point increase in Applied Problems standard score, there was a 9% increase in the odds of scoring proficient on the third grade TCAP.

Sixty-five children with third grade reading assessment data in spring 2013 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. None of the preschool assessments, whether administered in English or Spanish, were significantly associated with scoring proficient on the TCAP.

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

COHORT 2

In spring 2013, 138 children were enrolled in second grade and had reading assessment data. Eighty-four 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 at .48 ($p<.0001$). Letter-Word Identification was correlated with DRA2 at .60 ($p<.0001$). Applied Problems was associated with DRA2 scores at .53 ($p<.001$).

Fifty-four Cohort 2 children with second grade reading assessment data in spring 2013 had been assessed in both English and Spanish during their preschool year. Twenty-six of these children were assessed in English using the DRA2 at the end of second grade. Twenty-eight 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 and WJ Letter-Word Identification administered in English. PPVT-4 was correlated .55 with DRA2 ($p<.01$) and Letter-Word Identification was correlated .50 with DRA2 ($p<.05$). DRA2 scores in second grade were not significantly associated with Applied Problems administered in English or any of the Spanish assessments.

For children assessed in Spanish in second grade, EDL2 scores were significantly correlated with Letter-Word Identification administered in both English and Spanish (English: $r=.58$, $p<.01$; Spanish: $r=.75$, $p<.0001$). There were no significant associations between EDL2 and PPVT-4, TVIP or Applied Problems in either language.

In sum, for Cohort 2, 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 was more complicated. 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 and literacy skills assessed in English at the end of the preschool year and children's reading scores at the end of second grade. For children who were assessed in Spanish at the end of the second grade year, literacy 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 3

In spring 2013, 154 children were enrolled in first grade and had reading assessment data. Seventy-seven 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 .45 ($p < .0001$). Letter-Word Identification was correlated with DRA2 at .51 ($p < .0001$). Applied Problems was associated with DRA2 scores at .31 ($p < .001$).

Seventy-seven Cohort 3 children with first grade reading assessment data in spring 2013 had been assessed in both English and Spanish during their preschool year. Thirty-one of these children were assessed in English using the DRA2 at the end of first grade. Forty-six of them 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 were strong, positive associations between DRA2 in first grade and preschool assessments administered in English. DRA2 was correlated with PPVT-4, English Letter-Word Identification, and English Applied Problems at .54, .67 and .65, respectively (all $p < .01$). There was also a strong pattern of association between DRA2 and the Spanish preschool assessments. DRA2 was correlated with TVIP, Spanish LWI, and Spanish Applied Problems at .48, .45, and .69, respectively (all $p < .05$).

For children assessed in Spanish at first grade, there was only one significant association between a preschool assessment and EDL2. TVIP was correlated .38 with EDL2 ($p < .05$).

In sum, for Cohort 2, there was a strong pattern of associations for children whose primary language was English and for developing bilingual children who were assessed in English 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. For developing Spanish-English bilingual children who were assessed in Spanish, there was a weaker pattern of association.

COHORT 4

In spring 2013, 169 children were enrolled in kindergarten and had reading assessment data. Ninety-six 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 .36 ($p < .001$). Letter-Word Identification was correlated with DRA2 at .48 ($p < .0001$). Applied Problems was associated with DRA2 scores at .43 ($p < .0001$).

Seventy-three of Cohort 4 children with kindergarten reading assessment data in spring 2013 had been assessed in both English and Spanish during their preschool year. Thirty of these children were assessed in English

using the DRA2 at the end of kindergarten. Forty-three of them 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 not a strong pattern of association between the DRA2 and preschool assessments in English or Spanish. For preschool assessments in English, the only significant association was between DRA2 and WJ Applied Problems ($r=.55$, $p<.01$). There were no significant associations between DRA2 and the Spanish preschool assessments. For children assessed in Spanish in kindergarten, there was only one significant association. EDL2 was correlated with Spanish Applied Problems at $.55$ ($p<.01$).

In sum, for Cohort 4, 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 not a strong pattern of associations between kindergarten readiness and reading skill at the end of kindergarten.

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.

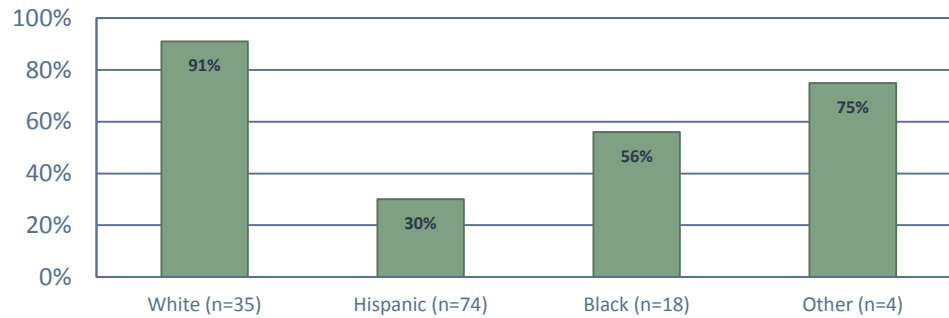
COHORT 1

TCAP

In spring 2013, among Cohort 1 DPP graduates, TCAP proficiency differed as a function of racial/ethnic group (see Figure 37).⁵⁹ The vast majority of the white children and three-fourths of those in the 'other' race/ethnicity category scored proficient or advanced on the 3rd grade reading TCAP. Hispanic and black children were slightly less likely to score proficient or advanced. About a third of Hispanic children and over half of black children scored proficient or advanced.

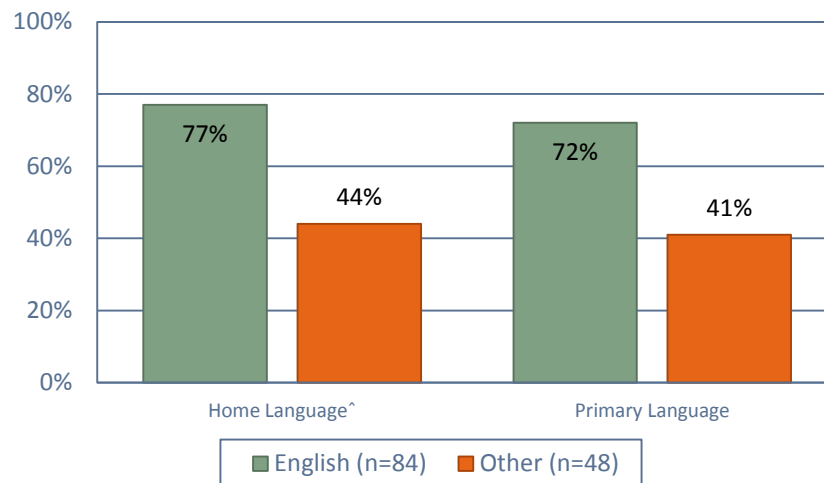
⁵⁹ $\chi^2_3=16.31$; $p<.01$

Figure 37: Percent of Children Scoring Proficient or Advanced on 3rd Grade Reading TCAP by Race/Ethnicity, Cohort 1, Spring 2013



Home language was also significantly associated with third grade reading proficiency (see Figure 38).⁶⁰ Over three-quarters of the children whose home language was English scored proficient or advanced on the TCAP reading. In contrast, less than half of children with another home language scored proficient or advanced. A similar pattern was observed for child primary language.⁶¹ Child gender was not associated with TCAP reading proficiency.

Figure 38: Percent of Children Scoring Proficient or Advanced on Third Grade TCAP Reading by Home Language and Child Primary Language, Cohort 1, Spring 2013



⁶⁰ $\chi^2_1=15.22, p<.0001$

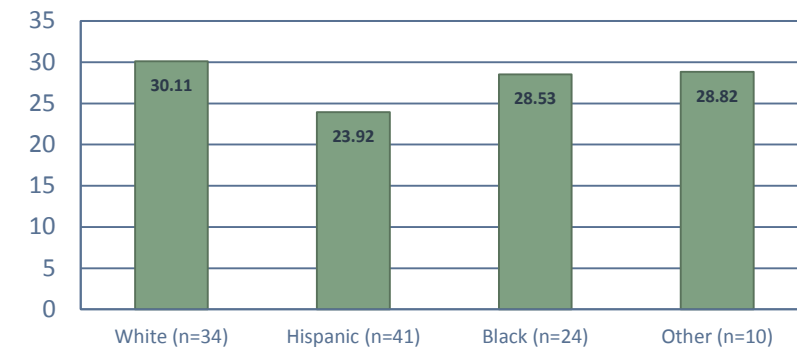
⁶¹ $\chi^2_1=14.18, p<.001$

COHORT 2

ENGLISH READING ASSESSMENT (DRA2)

In spring 2013, 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 39).⁶² 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.⁶³

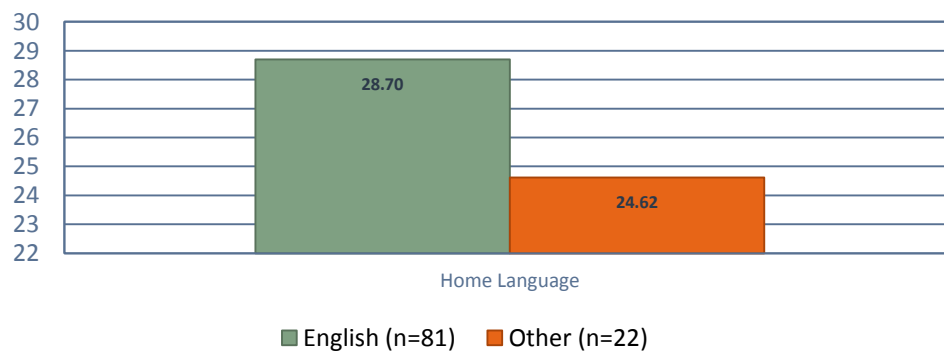
Figure 39: Second Grade Reading Level (DRA2) by Race/Ethnicity, Cohort 2, Spring 2013*



*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 40).⁶⁴ Children whose home language was English scored significantly higher than children with another home language. Children's primary language was not associated with DRA2 scores.

Figure 40: Second Grade Reading Level (DRA2) by Home Language, Cohort 2, Spring 2013*



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

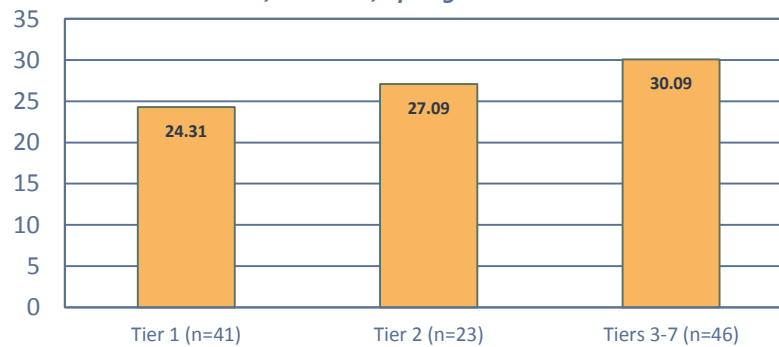
⁶² $F(3,107)=3.92, p<.05$

⁶³ A DRA2 score of 28 is considered "on grade level" for the end of second grade.

⁶⁴ Home Language: $t=2.17, df=101, p<.05$

There was also a significant difference by income tier (see Figure 41).⁶⁵ As income tier increased, so did average reading levels. Follow-up Tukey tests revealed that the group of children in Tiers 3-7⁶⁶ had significantly higher reading levels than children in Tier 1.

Figure 41: Second Grade Reading Level (DRA2) by Income Tier, Cohort 2, Spring 2013*



*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.⁶⁷

SPANISH READING ASSESSMENT (EDL2)

Unlike the English reading assessment, there was a significant difference in EDL2 scores by child gender (see Figure 42).⁶⁸ Boys scored significantly higher on the EDL2 than girls. 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. The effect for income was non-significant, as was the effect for region of the city.

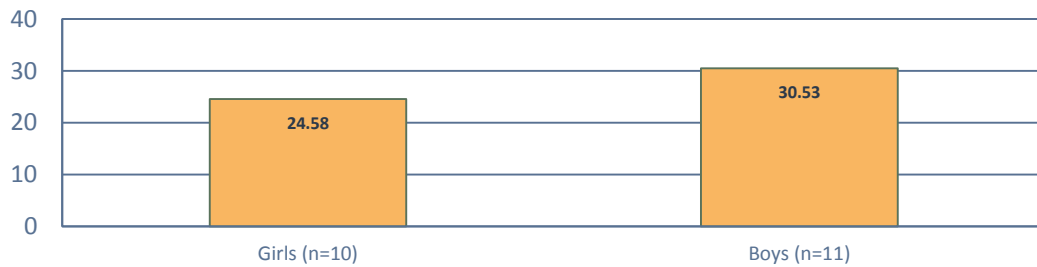
⁶⁵ Because of small sample sizes in some of the tiers, a collapsed version of income tier with three levels was used for this analysis: tier 1, tier 2 and tiers 3-7. $F(2,107)=5.28$, $p<.01$.

⁶⁶ The DPP income tier structure has changed over time. For cohorts 1 and 2, there was a 7-level income tier structure.

⁶⁷ $F(4,107)=1.62$, n.s.

⁶⁸ $t=2.99$, $df=19$, $p<.01$

Figure 42: Second Grade Reading Level (EDL2) by Gender, Cohort 2, Spring 2013*



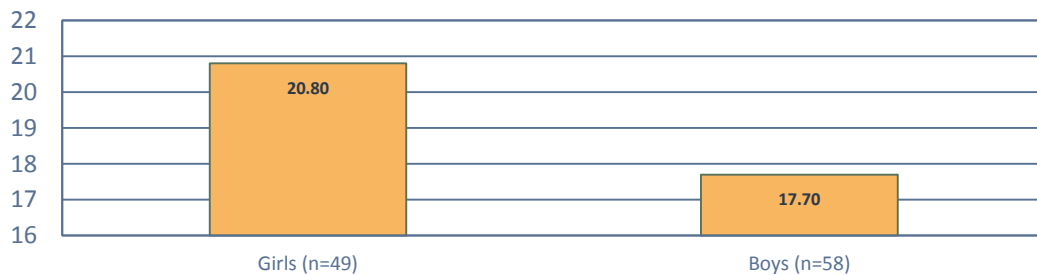
*A score of 28 is considered reading "on grade level" for the end of second grade.

COHORT 3

ENGLISH READING ASSESSMENT (DRA2)

For children assessed in English in first grade, there were significant associations between the DRA2 and three of the background characteristics: child gender, home language and income tier. Figure 43 displays the results for child gender. Girls scored significantly higher on the DRA2 than boys.⁶⁹ Figure 44 displays the results for home language. Children whose home language was English scored significantly higher on the DRA2 than children with another home language.⁷⁰ Figure 45 displays the results for income tier. As income tier increases, so do average DRA2 scores. Follow-up Tukey tests revealed that children in tiers 3-5 and tier 6 scored significantly higher than children in tier 1.⁷¹

Figure 43: First Grade Reading Level (DRA2) by Gender, Cohort 3, Spring 2013*



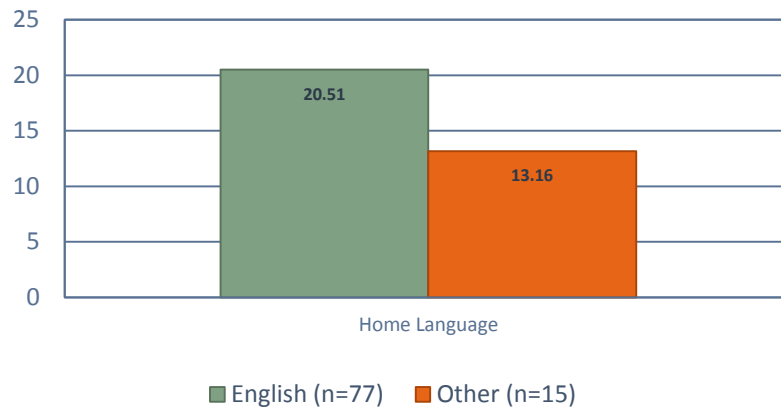
*A score of 16 is considered reading "on grade level" for the end of first grade.

⁶⁹ $t=2.01$, $df=105$, $p<.05$

⁷⁰ $t=3.90$, $df=90$, $p<.001$

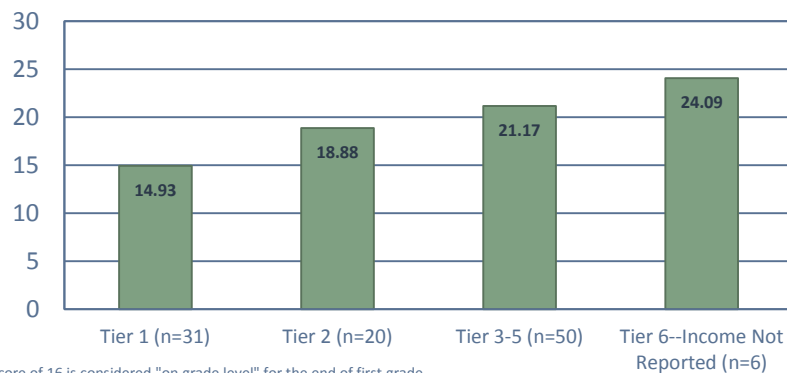
⁷¹ $F(3,106)=5.18$, $p<.01$

Figure 44: First Grade Reading Level (DRA2) by Child Primary Language and Home Language, Cohort 3, Spring 2013*



*A score of 16 is considered "on grade level" for the end of first grade.
 *83 children had English identified as their primary language, 19 had another language.
 *82 children had English identified as the home language, 21 had another language.

Figure 45: First Grade Reading Level (DRA2) by Income Tier, Cohort 3, Spring 2013*



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

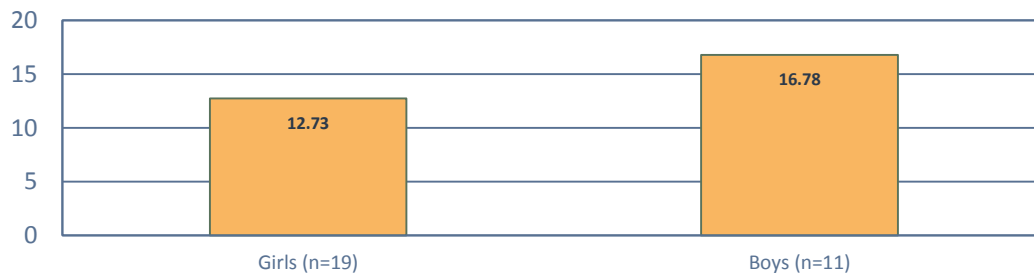
SPANISH READING ASSESSMENT (EDL2)

There were significant associations between first grade EDL2 and two of the demographic characteristics: child gender and 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. Figure 46 depicts the results for gender. Boys scored significantly higher on the EDL2, on average, than girls.⁷² Figure 47 depicts the results for tier level. All of the students who were assessed in Spanish

⁷² $t=2.10$, $df=28$, $p<.05$

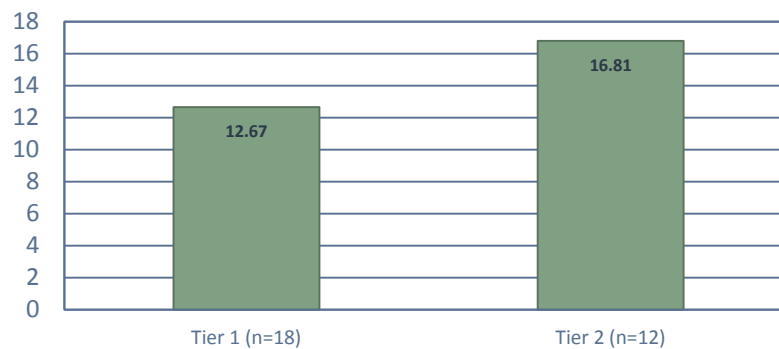
in first grade were either in tier 1 or tier 2 during their preschool year. Children in tier 2 scored significantly higher on the EDL2 than children in tier 1.⁷³

Figure 46: First Grade Reading Level (EDL2) by Gender, Cohort 3, Spring 2013*



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

Figure 47: First Grade Reading Level (EDL2) by Income Tier, Cohort 3, Spring 2013*



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

COHORT 4

ENGLISH READING ASSESSMENT (DRA2)

Kindergarten DRA2 scores varied significantly by race/ethnicity, home language, child language, and tier level. The results for race/ethnicity are presented in Figure 48. Follow-up Tukey tests revealed that Hispanic children scored significantly lower than white children on the DRA2.⁷⁴ The results for home language and child language are presented in Figure 49. Children whose home language was English scored significantly higher on the DRA2 than children with another home language.⁷⁵ Similarly, children whose primary language was English scored significantly higher, on average, on the DRA2 than children with another primary language.⁷⁶

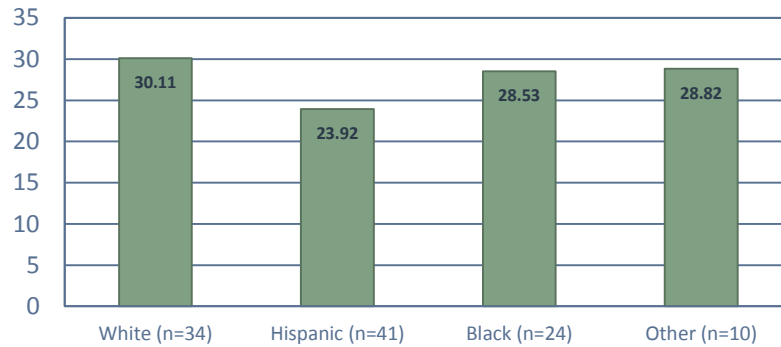
⁷³ $F(1,29)=4.71, p<.05$

⁷⁴ $F(3,118)=5.27, p<.01$

⁷⁵ $t=2.12, df=107, p<.05$

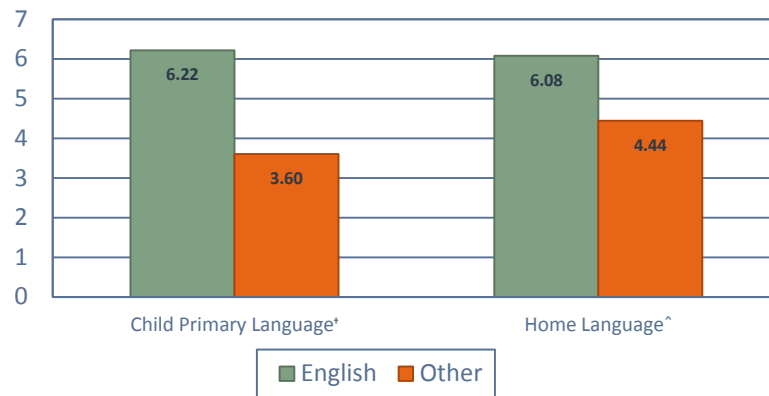
⁷⁶ $t=3.39, df=109, p<.01$

Figure 48: Kindergarten Reading Level (DRA2) by Race/Ethnicity, Cohort 4, Spring 2013*



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

Figure 49: Kindergarten Reading Level (DRA2) by Child Primary Language and Home Language, Cohort 4, Spring 2013*

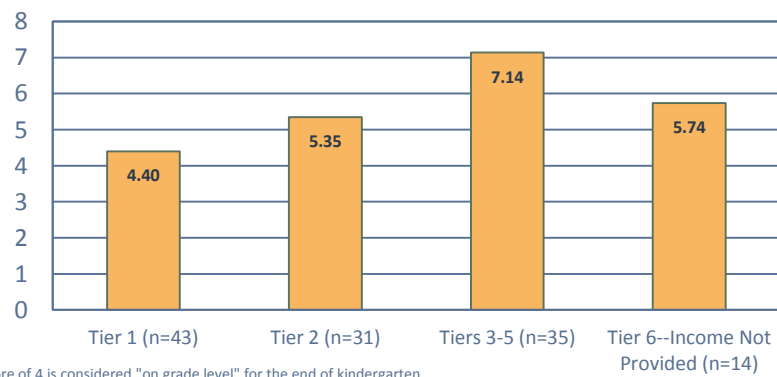


*A score of 4 is considered "on grade level" for the end of kindergarten.

*89 children had English identified as their primary language, 22 had another language.

*87 children had English identified as the home language, 22 had another language.

Figure 50: Kindergarten Reading Level (DRA2) by Tier Level, Cohort 4, Spring 2013*



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

There was also a significant difference by income tier (see Figure 50).⁷⁷ While all four 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 3-5 had significantly higher reading levels than children in Tier 1. The other pairwise comparisons were not statistically significant.

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 1

For Cohort 1, we were limited in our ability to address this question by the distribution of children in DPS and community sites. As described above, when we drew the sample for Cohort 1, we did not stratify by type of site. Reflective of the composition of children participating in DPP at the time of sampling, the Cohort 1 sample was comprised of 87% children from DPS sites. An analysis with such unequal group sizes is not ideal for detecting a statistical effect. We compared Cohort 1 children who had been enrolled in community preschools (n=17) with children who had been enrolled in DPS preschools (n=125) on third grade TCAP proficiency in reading administered during the spring of 2013. There was not a significant difference between these two groups.⁷⁸

COHORT 2

We compared Cohort 2 children who had been enrolled in community preschools (n=52) with children who had been enrolled in DPS preschools (n=56) on second grade reading assessments in English (DRA2) administered during the spring of 2013. There was not a significant difference between these two groups.⁷⁹ There was a significant difference by provider type in EDL2 scores.⁸⁰ The four children assessed with the EDL2 in second grade who had been enrolled in community sites tended to score lower on the EDL2 than children who had been enrolled in DPS preschools (n=17).⁸¹

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=54) on first grade reading assessments in English (DRA2) administered

⁷⁷ 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 3-5, and tier 6 (income not provided). $F(3,118)=4.36$, $p<.01$

⁷⁸ $\chi^2_1=.00$, n.s.

⁷⁹ $t=0.37$, $df=106$, n.s.

⁸⁰ $t=2.99$, $df=19$, $p<.01$

⁸¹ Community preschools: mean=20.0, sd=5.9; DPS preschools: mean=28.5, sd=4.9; 28 is considered "on grade level" for second grade.

during the spring of 2013. There was not a significant difference between these two groups.⁸² We were limited in our ability to test for a provider type difference in EDL2 by the fact that very few of the children who had been enrolled in community sites were assessed with the EDL2 (n=2).

COHORT 4

Among children assessed in English at the end of kindergarten with the DRA2, there was a significant difference between the 57 children who had attended community sites and the 62 children who had attended DPS preschools.⁸³ Children from community sites scored significantly higher than children who had attended DPS preschools.⁸⁴

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

As described in previous annual evaluation reports,⁸⁵ 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. 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 and 4, we examine the CLASS in lieu of the Qualistar rating.

COHORT 1

When we examined the distribution of star rating for children who had third grade TCAP data in spring 2013, we found that only 5% had been enrolled in preschools with less than a star 3 rating. Seventy percent of children had been enrolled in star 3 preschools and 24% 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 2

When we examined the distribution of star rating for children who had second grade DRA2 data in spring 2013, we found that only 6% had been enrolled in preschools with less than a star 3 rating. Sixty-eight percent of

⁸² $t=0.18$, $df=105$, n.s.

⁸³ $t=2.94$, $df=74.7$, $p<.01$

⁸⁴ Community preschools: mean=7.96, standard deviation=6.39; DPS preschools: mean=5.27, standard deviation=2.75.

⁸⁵ Klute, M. M. (2009). *Denver Preschool Program: Report on Child Outcomes, 2008-09 School Year*. Unpublished Report, October. Denver: Clayton Early Learning Institute.

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.

children had been enrolled in star 3 preschools and about 25% had been enrolled in star 4 preschools. In addition, the children who had been enrolled in lower quality programs tended to come from the higher income tiers. For example, among the children who had been enrolled in preschools with a lower star rating, 76% came from tiers 3-7. Because of this, we added in income tier as a covariate when we tested for the association between preschool quality and DRA2 scores. After controlling for income tier, the effect for star rating was non-significant.⁸⁶

When we examined the distribution of star rating for children with first grade EDL2 data, we found that 85% of them had been enrolled in star 3 preschools. Only three children were enrolled in star 4 preschools and just one child had been enrolled in a preschool with less than a star 3 rating. The lack of variability in star rating made it inappropriate to test for a difference by rating level.

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 DRA2. In all cases, the correlation between the measure of quality and the reading assessment score was non-significant. We did not compute correlations with EDL2 scores because there was so little variability in quality for sites attended by children assessed with the EDL2.

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

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

⁸⁶ $F(2,97)=0.92$, n.s.

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 2012-13 school year was the sixth 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 fully implement our evaluation design. This cohort was expected to be 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 first 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 exceeded that threshold by a small margin. 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: 61% 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 fewer than 5% of children. Teachers' ratings of behavioral concerns were rather low on average. Teachers identified behavioral concerns as an area of concern for about 11% of children. Parents identified protective factors as an area of concern for about 10% 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 fall below that threshold. Parents' ratings of protective factors also fell below that threshold. Parents identified behavioral concerns as an area of concern 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 45% 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 somewhat 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 study revealed an inconsistent pattern of results for income tier (defined by income adjusted for family size). Analyses of primary language groups revealed that children whose primary language is not English start the year lower and increase more over the course of the year than their primarily English-speaking counterparts on English vocabulary and the social-emotional assessments.

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 1, 2, 3, and 4 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 more pronounced for Cohorts 1, 2, and 3 than Cohort 4.

Cohort 1 children were compared to the population of third graders in DPS. DPP graduates scored proficient or advanced on the reading TCAP compared at a similar rate as the district as a whole.

Cohort 2 children were compared to all second graders in DPS. Among children whose reading was assessed in English in second grade, the proportion of DPP graduates who were reading at or above grade level was similar to the proportion of children in the district as a whole who were reading at or above grade level. Among children assessed in Spanish, DPP graduates were more likely to be reading at grade level than the district as a whole.

Cohort 3 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 smaller than in the district as a whole.

Cohort 4 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 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, we examined the association between CLASS observation scores and child outcomes. In the preschool year (Cohort 5) children 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. 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 the preschool year, the CLASS was related to literacy skills and protective factors.

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 children assessed in Spanish in first grade. 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 2013

Characteristic	Entire Sample, weighted ¹	By Provider Type, Unweighted		
		Community	DPS	Significance of Difference by Provider Type
Sex				$\chi^2_1=1.28$; n.s.
Female	48.7%	54.0%	46.0%	
Male	51.3%	46.0%	54.0%	
Ethnicity				$\chi^2_5=7.22$; n.s.
Hispanic	49.0%	39.0%	54.0%	
White (not of Hispanic origin)	30.33%	33.0%	29.0%	
African-American (not of Hispanic origin)	12.0%	18.0%	9.0%	
Multi-Racial	4.7%	6.0%	4.0%	
Asian/Pacific Islander	3.3%	4.0%	3.0%	
Native American	0.7%	0.0%	1.0%	
Child's Primary Language				$\chi^2_1=0.55$; n.s.
English	74.0%	76.0%	73.0%	
Another Language	25.3%	22.0%	27.0%	
Not Reported	0.7%	2.0%	0.0%	
Home Language				$\chi^2_1=1.83$; n.s.
English	72.0%	70.0%	73.0%	
Another Language	23.3%	16.0%	27.0%	
Not Reported	4.7%	14.0%	0.0%	
DPP Income Tier ²				$\chi^2_5=5.83$; n.s.
Tier 1	43.7%	47.0%	42.0%	
Tier 2	16.7%	14.0%	18.0%	
Tier 3	5.7%	5.0%	6.0%	
Tier 4	6.3%	3.0%	8.0%	
Tier 5	19.3%	18.0%	20.0%	
Tier 6—Income Not Reported	8.3%	13.0%	6.0%	
Star Level of Preschool				$\chi^2_4=4.18$; n.s.
Not Yet Rated	0.7%	2.0%	0.0%	
Provisional	0.3%	1.0%	0.0%	
Star 2	3.0%	3.0%	3.0%	
Star 3	56.0%	50.0%	59.0%	
Star 4	40.0%	44.0%	38.0%	
Region of the City				$\chi^2_4=6.41$; n.s.
Central	15.3%	20.0%	13.0%	
Northeast	23.3%	24.0%	23.0%	
Northwest	22.7%	28.0%	20.0%	
Southeast	10.7%	8.0%	12.0%	
Southwest	28.0%	20.0%	32.0%	

¹The weighted sample results are representative of the population of children enrolled in DPP in Fall 2012.

²DPP 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: DPP Income Tiers

Income Tier 1	
<i>Household Size</i>	<i>Annual Income Equal to or Less Than</i>
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		
<i>Household Size</i>	<i>Annual Income</i>	
	<i>More Than</i>	<i>Equal to or Less Than</i>
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		
<i>Household Size</i>	<i>Annual Income</i>	
	<i>More Than</i>	<i>Equal to or Less Than</i>
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		
<i>Household Size</i>	<i>Annual Income</i>	
	<i>More Than</i>	<i>Equal to or Less Than</i>
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		
<i>Household Size</i>	<i>Annual Income</i>	
	<i>More Than</i>	
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	