Denver Preschool Program: Report on Child Outcomes

2011-12 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
September 28, 2012

CONTENTS

EXECUTIVE SUMMARY

The Denver Preschool Program (DPP) is a taxpayer-funded initiative aimed at increasing access to high-quality preschool for all of Denver's children. DPP operates on the premise that preschool plays an important role in the academic and socioemotional 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 socio-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 and literacy skills. Spanish-speaking children also made progress in their 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 socio-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 socio-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 6% 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 2011-12 school year, nearly all children whose primary language was not English were from the lowest two income tiers as compared with about 50% 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 from the lowest income tiers and children whose primary language was not English tended to start the year lower than their counterparts from other groups on academic assessments administered in English. 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. That is, these children are on their way toward "catching up" to their peers from families from higher income tiers and those whose primary language is English. A similar pattern was observed for teacher-rated socioemotional skills and income. Children from lower income tiers tended to start lower and increase more over time than children from higher income tiers.

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

We followed three cohorts of DPP graduates who were enrolled in kindergarten, first and second grade during the 2011-12 school year. DPP graduates whose reading ability was assessed in English at the end of second grade were 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 were less 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 similar to 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 exceeded 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). This year, we did not find a strong pattern of associations between quality and child outcomes measured during the preschool year or reading skill at the end of kindergarten.

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. There was some evidence that children from higher-risk groups (living in or near poverty, speaking a language other than English primarily) made progress toward closing the achievement gap that was present at the beginning of the preschool year. 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 small group of children assessed in Spanish in second 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 all of Denver's children. DPP operates on the premise that preschool plays an important role in the academic and socioemotional 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 (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 beyond[1]. 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 their demographic counterparts who did not receive DPP tuition credits 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 2011-12 school year was the fifth year of the DPP program. During the first year of DPP's operation, we were not able to fully implement our evaluation design. The cohort from this first school year is best viewed as a pilot sample. This cohort was expected to be enrolled in third grade during the 2011-12 school year, the first grade in which students take the TCAP. This report provides preliminary answers to question 5 using data from this cohort. However, since the composition of this cohort is not representative of DPP participants as a whole, readers are urged to regard these results with caution. Question 5 will be begin to be addressed starting with the 12-13 school year, when the first full cohort of DPP children that were studied start to take TCAP assessments.

[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 for the child outcomes portion of the DPP evaluation includes 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				
	07-08	08-09	09-10	10-11	11-12
Cohort 0	Preschool	Kindergarten	1st Grade	2nd Grade	3rd Grade
Cohort 1		Preschool	Kindergarten	1st Grade	2nd Grade
Cohort 2			Preschool	Kindergarten	1st Grade
Cohort 3				Preschool	Kindergarten
Cohort 4					Preschool

COHORT 0

The DPP child outcomes evaluation study began during the first year of operation of the DPP program, the 07-08 school year. Our ability to carry out the evaluation plan as designed was limited in this first year by issues associated with the startup of the program.[2] As such, data from the 07-08 school year are best viewed as pilot data. The primary usefulness of data from this cohort is to test procedures and inform adjustments and improvements to the evaluation design, rather than generating results that can be used to inform the DPP program. To reflect the preliminary nature of this cohort, it is referred to as Cohort 0. The total sample size for this cohort was 121; 30 children were assessed in the winter of the preschool year and 118 children were assessed in the spring of the preschool year.

Our evaluation design involved obtaining reading assessment data from DPS. Prior to requesting these data, we needed to obtain Denver Public Schools ID numbers (DPS IDs) for children in our sample. Every child enrolled in DPP was assigned a DPS ID. DPS IDs were sent from DPS to Affiliated Computer Services (ACS), the contractor that handles enrollment of families and payment of tuition credits for DPP. We requested DPS IDs for Cohort 0 from ACS. Their records did not include a DPS ID for all children. For Cohort 0, we received DPS IDs and requested reading data from DPS for 114 children (97% of the original sample).

Cohort 0 children were expected to be in third grade during the 11-12 school year (see Table 1). We obtained 2011 Transitional Colorado Assessment Program (TCAP) reading assessment scores from DPS for 67 of these children (55% of the total sample, 59% of those for whom we had obtained DPS IDs).

COHORT 1

During the 2008-09 school year, 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

[2] Readers are referred to the first DPP Annual Evaluation Report for details about the evaluation in the first year. Augenblick, Palaich and Associates, and Clayton Early Learning Institute (2008). *An Evaluation of the Denver Preschool Program 2007-08.* Unpublished Report. Denver: Augenblick, Palaich and Associates.

those children in the fall and spring of their preschool year.[3] Henceforward, this cohort of children will be referred to as Cohort 1. The total sample size for Cohort 1 was 207; 200 children were assessed in the fall and spring of the preschool year. We were able to obtain DPS IDs from ACS for 200 of these children (97% of the original sample).

Cohort 1 children were expected to be in the second grade during the 11-12 school year (see Table 1). We obtained spring reading assessment data for 152 children (73% of the whole sample; 76% of those for whom we had obtained DPS IDs). Of these, 98% were in second grade as expected. Three children were in first grade. All three of these children were in kindergarten in spring 2011, suggesting that their parents delayed their entry into kindergarten.

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.

Cohort 2 children were expected to be in first grade during the 11-12 school year (see Table 1). We obtained reading data for 139 children (69% of the sample; 72% of the sample when sampling weights were applied). One child was in second grade, seven were in kindergarten, and all other children were in first grade as expected. The child who was in second grade had spring 2011 first grade assessment data. It appears that this child may have skipped kindergarten. For four of the seven kindergarteners, we had spring 2011 kindergarten reading assessment data, suggesting that these children repeated kindergarten. The remaining three kindergarteners did not have spring 2011 assessment data, suggesting that their parents delayed their entry into kindergarten.

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.

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

Cohort 3 children were expected to be in kindergarten during the 11-12 school year (see Table 1). We obtained reading data for 143 children (71% of the sample; 72% of children for whom we were able to obtain DPSIDs; 80% of the total sample when sampling weights were applied). All of the children were in kindergarten as expected.

COHORT 4

SAMPLING PLAN

As with previous years, we stratified our sample for Cohort 4 by type of provider. The result is two samples: a sample of children in community sites and a sample of children in DPS sites. Both of these samples are 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 are not applied.

During the DPP enrollment process, parents were asked if they would be willing to be contacted about participation in the evaluation study.[4] In August 2011, a sample of 100 children enrolled in community sites was drawn from the group of families that volunteered to participate (henceforth referred to as "volunteers"). In September 2011, a sample of 100 children enrolled in DPS 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 373 volunteers and the 90 non-volunteers for four variables. First, there was a significant difference between volunteers and non-volunteers on child's primary language.[5] Families with children who spoke English as their primary language were less likely to volunteer than families with children who primarily spoke another language. Seventy-eight percent of families with children who primarily spoke English volunteered compared with 90% of families with children who primarily spoke another language. Second, a similar pattern was observed for home language.[6] Seventy-nine percent of families with a home language of English volunteered compared with 90% of families with another home language. Third, there was a significant association between region of the city in which the child lived and volunteer status.[7] Follow up analyses revealed that this was primarily due the central and southeast regions of the city. Families living in the central region of the city were more likely to volunteer than families in other regions of the city. Ninety percent of families living in the central region of the city volunteered to be contacted, compared with 79%

[4] Information about the evaluation was provided on the DPP application, which was available in both English and Spanish.
[5] χ^2_1=6.20, p<.05
[6] χ^2_1=4.64, p<.05
[7] χ^2_4=14.02, p<.01

of families living in other regions of the city. In contrast, families residing in the southeast region of the city were less likely to volunteer than families living in other regions of the city. Sixty-eight percent of families living in the southeast volunteered to be contacted compared with 82% of families living in other regions. Finally, volunteer status was significantly associated with whether or not families opted out of providing income documentation during the enrollment process.[8] Families who opted out of providing income documentation were less likely to volunteer to be contacted about the evaluation study. Fifty-four percent of those who opted out of providing income documentation volunteered to be contacted compared with 82% of those who provided income documentation. Volunteers and non-volunteers were similar in terms of sex of the child, ethnicity, DPP income tier, and Qualistar rating of the preschool program. To adjust for these differences, the sampling frame was stratified by child language, region of the city (central, southeast, other), and income opt-out.[9] 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 August 2011 with respect to the variables examined. The sample was drawn with replacement; if a selected child was deemed ineligible for the study,[10] 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 2401 volunteers and 1038 non-volunteers on two variables.[11] First, a significant difference was detected for income tier level.[12] Follow-up analyses revealed that this was due to a difference between volunteers and non-volunteers in tier 5. Parents in tier 5 were significantly more likely to volunteer than families in other income tiers. Seventy-six percent of families in tier 5 volunteered compared with 69% of families in the other income tiers. A significant difference was also detected for ethnicity.[13] Follow-up analyses revealed that this effect was due to differences between volunteers and non-volunteers in two ethnic groups. Parents of white children were significantly more likely to volunteer to be contacted about the evaluation than parents of children from other ethnic groups. Seventy-four percent of parents of white children volunteered compared with 68% of parents of children from other ethnic groups. In contrast, parents of Asian children were significantly less likely to volunteer than parents of children from other ethnic groups. Fifty-two pecent of parents of Asian children volunteered compared with 70% of parents of children from other ethnic groups. To adjust for these differences, the sampling frame was stratified by income tier (tier 5 vs. other) and ethnicity (white, Asian, other). The proportion of children drawn from each stratum was adjusted to match the proportions in the population of children enrolled in DPP at the time of sampling. The result was a sample of 100 that was representative of the DPS site population as a whole in September 2011 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

[8] Families have an option of not providing documentation of their incomes. If they do so, they will receive the tuition credit associated with the highest income level. $\chi^2_1=11.26$, $p<.001$

[9] Home language and child primary language were strongly associated ($\chi^2_1=368.46$, $p<.0001$). Of the 411 families in the analysis, only 5 had a home language that differed from the child's primary language. Because these two variables were so strongly associated, it was not feasible to stratify on both variables. Instead, we only stratified on child primary language.

[10] Typically children were deemed ineligible because they were no longer enrolled in a DPP preschool at the time the family was contacted for participation.

[11] Because of the very large sample size and associated statistical power, a p-value of .01 was used for determining statistical significance.

[12] $\chi^2_5=16.13$, $p<.01$

[13] $\chi^2_6=34.85$, $p<.0001$

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, 11.9% of children enrolled in DPP were attending community sites and the remaining 88.1% were attending DPS sites.[14] The sample was divided evenly between community sites and DPS sites. As a result, the sampling design involved oversampling children from community sites. When analyzing data for the sample of 200 as a whole, it was important to weight the sample so that both program types had weights in the analysis that are comparable to each group's proportion of the total population. The result is an analysis of data that are representative of the DPP population as a whole.

SAMPLE CHARACTERISTICS

Characteristics of the fall sample are summarized in Table 2.[15] 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 about a tenth of the sample. Slightly over half of children spoke English as their primary language and in slightly over half of their homes, English was the primary language spoken. In terms of income, over two-thirds 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 2009. The upper bound for Tier 2 is equivalent to 185% of the Federal Poverty Guideline for 2009, which is also the cutoff for Free and Reduced Lunch. The next most common income tier was Tier 5. About six percent of families were assigned to the highest tier, Tier 6, because they opted out of the requirement to provide their income.

Nearly all (97%) of the children were enrolled in preschools with a 3 or 4 star rating. Nearly two-thirds of children were enrolled in star 3 preschools and slightly over a third of children were enrolled in star 4 preschools. Nearly a third of children resided in the northeast region of the city. Slightly over a quarter of children resided in the southwest region 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. The proportion of boys and girls was similar for the two provider types. There was a significant difference in the ethnic breakdown in the two types of sites. Follow-up analyses revealed that this was primarily due to differences in the distribution of Hispanic and white children in the two types of sites. There was a much larger percentage of Hispanic children in DPS sites, over twice the magnitude of the proportion of Hispanic children in community sites.[16] In contrast, the proportion of white children in DPS sites was about half the proportion of white children in community sites.[17] Provider type was also significantly associated with both child primary language and home language. DPS preschools tend to serve a population of children that is more diverse in terms of language. About four-tenths of children in DPS sites have a primary language other than English compared with about a tenth of children in community sites. Not surprisingly, a similar pattern was observed for home language.

---

[14] 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 ACS database to determine their provider type.

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

[16] $\chi^2_1=17.02$; $p<.0001$

[17] $\chi^2_1=18.14$; $p<.001$

Table 2: Cohort 4 Sample Characteristics Fall 2011[1]

Characteristic	Entire Sample, weighted[2]	By Provider Type, Unweighted		
		Community	DPS	Significance of Difference by Provider Type
Sex				χ^2_1=.50; ns.
Female	46.6%	51.0%	46.0%	
Male	53.4%	49.0%	54.0%	
Ethnicity				χ^2_4=21.73; p<.001
Hispanic	51.6%	26.0%	55.0%	
White (not of Hispanic origin)	28.4%	54.0%	25.0%	
African-American (not of Hispanic origin)	10.1%	11.0%	10.0%	
Multi-Racial	6.0%	6.0%	6.0%	
Asian/Pacific Islander	3.8%	2.0%	4.0%	
Child's Primary Language				χ^2_1=21.91; p<.0001
English	54.9%	84.0%	51.0%	
Another Language	33.9%	11.0%	37.0%	
Not Reported	11.2%	5.0%	12.0%	
Home Language				χ^2_1=23.23; p<.0001
English	55.3%	80.0%	52.0%	
Another Language	37.4%	11.0%	41.0%	
Not Reported	7.2%	9.0%	7.0%	
DPP Income Tier[3]				χ^2_5=30.50;p<.0001
Tier 1	44.4%	18.0%	48.0%	
Tier 2	24.0%	24.0%	24.0%	
Tier 3	0.6%	5.0%	0.0%	
Tier 4	8.9%	8.0%	9.0%	
Tier 5	15.9%	37.0%	13.0%	
Tier 6—Income Not Reported	6.2%	8.0%	6.0%	
Star Level of Preschool				χ^2_2=16.66; p<.001
Star 1	0.0%	0.0%	0.0%	
Star 2	3.0%	10.0%	2.0%	
Star 3	61.8%	38.0%	65.0%	
Star 4	35.3%	52.0%	33.0%	
Region of the City				χ^2_4=7.27; n.s.
Central	13.8%	20.0%	13.0%	
Northeast	30.2%	24.0%	31.0%	
Northwest	17.2%	26.0%	16.0%	
Southeast	10.1%	11.0%	10.0%	
Southwest	28.7%	19.0%	30.0%	

[1]Some percentages do not sum to 100 because of rounding error.
[2]The weighted sample results are representative of the population of children enrolled in DPP in Fall 2011.
[3]DPP Income Tiers are determined using family income and family size. Tier 1 is the lowest income. Details on the income tiers can be found in the appendix.

The association between income tier and provider type was significant. Follow-up analyses revealed that this was due to the distribution of children from Tiers 1, 3 and 5 across the provider types. DPS sites enrolled a much larger proportion of children from Tier 1 than community sites.[18] Nearly half of children enrolled in DPS sites were from Tier 1 compared to about a fifth in community sites. Conversely, community sites enrolled a higher proportion of children from Tier 5 than DPS sites.[19] Over a third of children in community sites were from Tier 5 compared with just 13% of children from DPS sites. Very few children in the sample were from Tier 3. However, as can be seen in Table 2, all of these children were enrolled in community sites.

Star rating of the preschool attended also varied by provider type. It was rare for any child in the sample to be in a preschool with a 2 star rating, but it was five times more likely for a child in a community site than in a DPS site.[20] However, children in community sites were also more likely be enrolled in preschools with a 4 star rating. In fact, over half of children from community sites were enrolled in 4 star sites, compared with a third of children in DPS preschools.[21] Conversely, children in enrolled in DPS preschools were more likely to be in a 3 star preschool. About two-thirds of children in DPS preschools were enrolled in 3-star sites compared with slightly over a third of children enrolled in community sites.[22]

There was no association between provider type and region of the city. The distribution of children across the city was similar for the two provider types.

In the spring 2012, there were three children that were lost to follow-up for the following reasons: one child moved out of Denver before the spring round and two 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.[23] As a result, the total sample size for the 2011-12 school year is 203.

REPRESENTATIVENESS OF THE SAMPLE

Analyses were conducted to test whether the sample selected was representative of the population of DPP children enrolled. 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 2011-12 school year. Each set of analyses are described in turn below.

[18] $\chi^2_1=20.35$; $p<.0001$
[19] $\chi^2_1=15.36$; $p<.0001$
[20] $\chi^2_1=5.67$; $p<.001$
[21] $\chi^2_1=7.39$; $p<.01$
[22] $\chi^2_1=14.59$; $p<.0001$
[23] 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.

FALL 2011

COMMUNITY SAMPLE

Children who were included in the community sample were compared to 363 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. There was a significant effect for tier level.[24] Follow-up analyses revealed that this effect was due to Tiers 1 and 5. In the population of DPP children enrolled in community sites, 26% of children were in Tier 1. In the sample, 16% of children were in Tier 1. Conversely, in the population, 27% of children were in Tier 5. In the sample, 36% of children were in Tier 3. The tests for differences in the remaining variables were all non-significant, indicating that the sample did not differ significantly from those not in the sample.[25] That is, the community sample was slightly wealthier but was otherwise representative of the population of children enrolled in community sites in August 2011.

DPS SAMPLE

Children who were included in the DPS sample were compared to 3339 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.[26] That is, the DPS sample was representative of the population of children enrolled in DPS sites in September 2011.

SUMMER 2012

COMMUNITY SAMPLE

Children who were included in the community sample were compared to 1574 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. Similar to the fall, there was a significant effect for tier level.[27] Follow-up analyses revealed that the sample underrepresented children from Tier 1[28] and overrepresented children from Tiers 2 and 5.[29] There were also significant effects for star rating of preschool, ethnicity, child language, and home language.[30] Follow-up analyses revealed that the star rating effect was due to an overrepresentation of children in star 4 preschools and an underrepresentation of children in star 3 preschools.[31] Follow-up analyses revealed that

---

[24] χ^2_5=11.88, p<.05

[25] Gender: χ^2_1=0.35, n.s.; ethnicity: χ^2_6=11.41, n.s.; Qualistar rating: χ^2_3=2.30, n.s.; home language: χ^2_1=0.74, n.s.; child primary language: χ^2_1=0.23, n.s.; region of the city: χ^2_4=1.02, n.s.

[26] Gender: χ^2_1=0.30, n.s.; ethnicity: χ^2_6=5.43, n.s.; income tier: χ^2_6=6.33, n.s.; Qualistar rating: χ^2_2=1.22, n.s.; home language: χ^2_1=0.02, n.s.; child primary language: χ^2_1=0.02, n.s.; region of the city: χ^2_4=2.78, n.s.

[27] χ^2_5=45.90, p<.0001

[28] 19% of children in the sample were from tier 1 compared with 48.9% in the population as a whole.

[29] 25% of children in the sample were from tier2 compared with 14.5% in the population as a whole; 37% of the in the sample were from tier 5 compared with 19.9% in the population as a whole.

[30] Star Rating: χ^2_4=19.74, p<.001; Ethnicity: χ^2_6=24.76, p<.001; home language: χ^2_1=5.54, p<.05; child primary language: χ^2_1=4.56. p<.05

[31] 56% of children in the sample attended star 4 preschools compared with 36% in the population as a whole. Conversely, 36% of children in the sample attended star 3 preschools compared with 54% in the population as a whole.

the ethnicity effect was because the sample included more white children and fewer Hispanic children than were included in the population at the end of the school year.[32] By the end of the school year, 79% the population of DPP children enrolled in community sites had English as the identified home language. Eight-nine percent of the community sample had English as their identified home language. Similarly, by the end of the school year, 80% of the population of DPP children enrolled in community sites spoke English as their primary language. Eighty-eight percent of the children in the community sample spoke English as their primary language.

These effects were due, in large part, to the increase in enrollment over the course of the year in community sites. At the time of sampling, 463 children were enrolled. Over the course of the year, 1210 additional children from community sites enrolled. A relatively small proportion of these later enrolling children were enrolled in Star 4 preschools (32%) compared with children who were enrolled at the time of sampling (46%). A large proportion of these children were Hispanic (45%), changing the racial/ethnic distribution of enrolled children. A quarter of this group of later enrolling children had home languages and primary languages that were not English, changing the language distribution in the population.

The tests for differences in the remaining variables were all non-significant, indicating that the sample did not differ significantly from those not in the sample.[33] In sum, as in the fall, the sample was somewhat wealthier than the population as a whole in summer 2012. Further, because later enrolling children were more likely to be Hispanic and non-English speaking and less likely to be enrolled in a Star 4 preschool, the sample includes more white and English speaking children and children enrolled in Star 4 preschools than the population as a whole in summer 2012.

DPS SAMPLE

Children who were included in the DPS sample were compared to 3853 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.[34] That is, the DPS sample was representative of the population of enrolled children in DPS sites at the end of the school year.

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.

Teachers, after providing informed consent, 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

[32] The sample included 54% white and 26% Hispanic children compared with 32% white and 39% Hispanic in the population of enrolled children at the end of the school year.

[33] Gender: χ^2_1=.08, n.s.; region of the city: χ^2_4=5.16, n.s.

[34] Gender: χ^2_1=..48, n.s.; Ethnicity: χ^2_6=2.85, n.s.; income tier: χ^2_5=5.81, n.s.; Qualistar rating: χ^2_3=.45, n.s.; home language: χ^2_1=.28, n.s.; child primary language: χ^2_1=.29, n.s.; region of the city: χ^2_4=3.66, n.s.

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 socioemotional development in January 2012. 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: Sample sizes by data collection type, Fall 2011 and Spring 2012

Data Collection Activity	Fall 2011	Spring 2012
Standardized Assessments—English	200	200
Standardized Assessments—Spanish	52	54
DECA—Teacher Report	199 (99.5%)	199 (99.5%)
DECA—Parent Report	200 (99% of the full sample n=203)	
Classroom Observations[1]	169 (83% of the full sample n=203)	

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

Table 3 presents the total sample sizes for each data collection activity. About a quarter of the children in the sample spoke Spanish and completed assessments in Spanish as well as English. Response rates for the teacher surveys were excellent, with nearly all teachers completing the survey in the fall and spring. Response rate for the parent survey and classroom observations were also excellent.

MEASURES

PRESCHOOL YEAR

ARCHIVAL DATA

Information about demographic characteristics was obtained from ACS, the contractor that handles enrollment and tuition payments for the Denver Preschool Program. Information about program quality was obtained from Qualistar Colorado, which is responsible for conducting quality ratings of sites. Reading assessment data for kindergarten, first grade, second grade, and third were obtained from Denver Public Schools.

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.[35]

[35] 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

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 report. 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.

Table 4: Areas of Child Development Assessed

Area Assessed	Name of Assessment	Acronym	Language of Assessment
Receptive Vocabulary	Peabody Picture Vocabulary Test-4[36]	PPVT	English
	Test de Vocabulario en Imagenes Peabody[37]	TVIP	Spanish
Literacy Skills	Woodcock-Johnson III Achievement Battery,[38] Letter-Word Identification Subtest	WJ LWI	English
	Batería III Woodcock-Muñoz,[39] Letter-Word Identification Subtest	WM LWI	Spanish
Math Skills	Woodcock-Johnson III Achievement Battery, Applied Problems Subtest	WJ AP	English
	Batería III Woodcock-Muñoz, Applied Problems Subtest	WM AP	Spanish
Socioemotional Development	Devereaux Early Childhood Assessment[40]	DECA	English or Spanish

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

---

of State-Wide Early Education Programs (SWEEP): Preliminary descriptive report. Chapel Hill, NC: National Center for Early Development & Learning.

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

[37] Dunn, L. M., Lugo, D. E., Padilla, E. R., & Dunn, L. M. (1986). Test de Vocabulario en Imagenes Peabody (TVIP). Minneapolis: Pearson Assessments.

[38] 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.

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

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

teacher-child interactions.[41] The 2011-12 school year was the second 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 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 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 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)[42] and its Spanish language counterpart Evaluación del Desarrollo de la Lectura (EDL2).[43] 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 instructional system designed to help teachers pinpoint children's reading level and design differentiated instruction to meet the needs of all children in their classroom.[44] The assessment yields a reading level for each child. In kindergarten, a reading level of 4 is considered reading on grade level.[45] 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.[46] 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

[41] Pianta, R. C., LaParo, K. M., & Hamre, B. K. (2008). Classroom Assessment Scoring System (CLASS) Manual, Pre-K. Baltimore, MD: Brookes.
[42] Beaver, J. M., & Carter, M. A. (2006). *The Developmental Reading Assessment, Second Edition (DRA2).* Upper Saddle River, NJ: Pearson.
[43] Ruiz, O.A. & Cuesta, V. M. (2007). *Evaluación del desarrollo de la lectura.* Parsippany, NJ: Pearson Education, Inc.
[44] *K-8 Technical Manual, Developmental Reading Assessment, Second Edition* (2009). Upper Saddle River, NJ: Pearson Education, Inc.
[45] Prior to the 2010-11 school year, a reading level of 3 was considered on grade level for kindergarten.
[46] For more information about TCAP, visit http://www.cde.state.co.us/assessment/GeneralInfo.asp

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 higher scores on Behavioral Concerns and Self-Control. Once again there was substantial variability in all of the scores.

Table 5: Weighted Descriptive Statistics for Child Outcome Measures

Variable	**Fall 2011**				**Spring 2012**			
All Children	N	Mean	Standard Deviation	Range	N	Mean	Standard Deviation	Range
Standardized Assessments								
PPVT Standard Score	200	91.82	23.75	25-148	200	94.94	21.75	39-157
WJ LWI Standard Score	200	99.84	12.52	73-159	200	102.45	12.10	68-168
WJ AP Standard Score	200	105.49	14.91	71-136	200	105.80	12.40	64-139
Teacher-Rated DECA								
Initiative T-Score	199[1]	51.47	8.35	33-72	199	56.51	7.61	35-72
Self-Control T-Score	199	55.94	8.38	28-72	199	59.42	7.65	30-72
Attachment T-Score	199	49.47	8.31	31-72	199	52.86	8.57	33-72
Total Protective Factors T-Score	199	51.81	8.22	31-72	199	56.52	7.37	34-72
Behavioral Concerns T-Score	191	48.85	8.93	31-72	192	46.94	7.71	31-72
Parent-Rated DECA								
Initiative T-Score	--	--	--	--	200	51.80	9.00	28-72
Self-Control T-Score	--	--	--	--	200	54.89	8.69	28-72
Attachment T-Score	--	--	--	--	200	48.34	10.69	28-72
Total Protective Factors T-Score	--	--	--	--	200	51.41	9.09	28-72
Behavioral Concerns T-Score	--	--	--	--	200	57.15	9.89	28-72
Spanish-Speaking Children Only								
Standardized Assessments								
TVIP Standard Score	52	89.87	19.28	63-118	53	92.02	19.11	55-118
WM LWI Standard Score	51	97.36	14.13	75-119	54	103.13	13.85	76-131
WM AP Standard Score	52	94.23	13.38	64-118	54	98.93	14.33	62-131

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

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 about 2.5 standard deviations higher on the PPVT than their counterparts with another primary language. For LWI and AP, children whose primary language was English scored about 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, about two standard deviations in magnitude. Differences between primary language groups for LWI and AP were slightly smaller than the fall, but still statistically significant. For LWI, the difference between language groups was nearly one standard deviation in magnitude. For AP, the difference between the groups was nearly two-thirds of a standard deviation.

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

Assessment	Primary Language						t
	English			Another Language			
	N	Mean	SD	N	Mean	SD	
PPVT Standard Score	135	106.40	14.86	48	68.56	17.01	15.84***
WJ LWI Standard Score	135	106.00	9.71	48	90.90	9.89	10.04***
WJ AP Standard Score	135	112.30	10.90	48	95.81	15.51	7.66***

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

[1]Information 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 Round[1]

Assessment	Primary Language						t
	English			Another Language			
	N	Mean	SD	N	Mean	SD	
PPVT Standard Score	133	109.50	15.07	50	77.74	17.51	12.20***
WJ LWI Standard Score	133	108.80	13.58	50	95.48	14.02	5.88***
WJ AP Standard Score	133	109.80	10.57	50	100.40	11.27	5.27***

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

[1]Information 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 139 classrooms in 97 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.[47]

[47] It is important to keep in mind that all of the preschool quality information provided here is based on only a sample of 97 preschools where the children in the sample were enrolled. For information on the quality of all preschool programs participating in DPP during the 11-12 school year, readers are referred to the annual evaluation report prepared by Augenblick, Palaich and Associates.

QUALISTAR RATING AND ACCREDITATION

96 of the 97 preschools were Qualistar rated. Of those that were Qualistar rated, 94 were center-based sites and two were home-based. Detailed information about the quality of these preschools was provided to Clayton Early Learning from Qualistar. One site became eligible for DPP because they had obtained Accreditation from the National Association for the Education of Young Children (NAEYC). For this site, the only quality information that is available is the number of stars.[48] Figure 1 presents the breakdown of programs by star level. About two-thirds of the programs had 3 stars. Approximately one-quarter of programs had 4 stars. No preschools had a provisional rating or a rating of one star, indicating that very few programs were of the lowest quality.

Figure 1: Star Level of Programs Attended by Children in the Sample (n=97 programs)

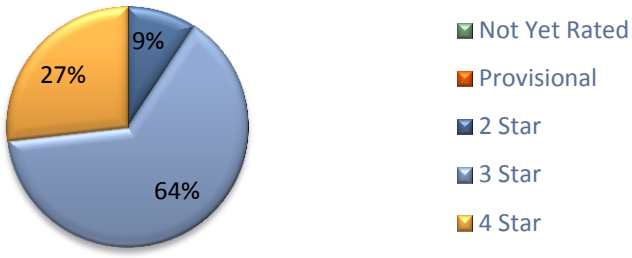


Table 8 presents descriptive statistics for the five component areas of the Qualistar rating for the 96 sites with a Qualistar rating.[49] Sites were strongest, on average, in the areas of Family Partnerships and Adult-to-Child Ratios and Groups Size. Family Partnerships was a particularly strong area, with programs earning, on average, over 90% of the possible points for this area. While scores in this area covered a wide range (0-10), very few programs earned very low scores on this component. One program earned no points for this area, four 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 4 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 or no 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 as few as 4 points and others earning all of the possible points. A very small proportion of programs earned the 2 points for having earned an accreditation.

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

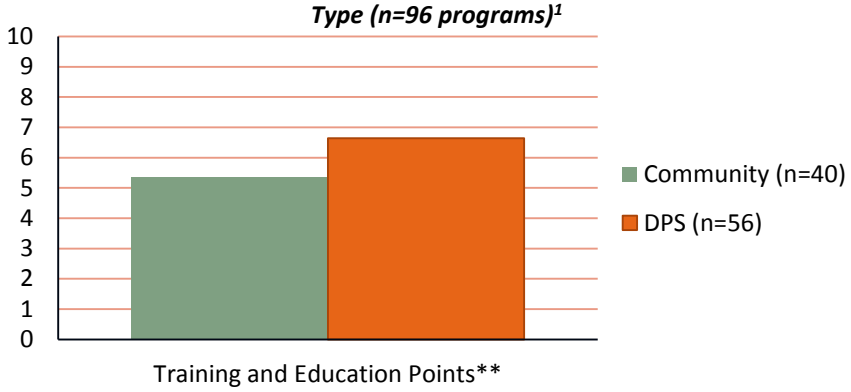
Component	Possible Range	Mean	SD	Range
Learning Environment	0-10	6.85	1.58	4-10
Family Partnerships	0-10	9.17	1.68	0-10
Training and Education	0-10	6.10	1.73	0-10
Adult-to-Child Ratios and Group Size	0-10	8.72	1.43	4-10
Accreditation	0-2	.02	0.20	0-2

[48] 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 also receive a DPP Quality Rating of 3 stars.

[49] More information about the five component areas of the Qualistar rating is available at: http://www.qualistar.org/qualistar-rating-components.html

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).[50] On average, DPS programs earned significantly more points than community preschools in this area.

Figure 2: Qualistar Rating Training and Education Points for Programs Attended by Children in the Sample, by Provider Type (n=96 programs)[1]



**p<.01
[1]Standard deviations: Community=1.92, DPS=1.37

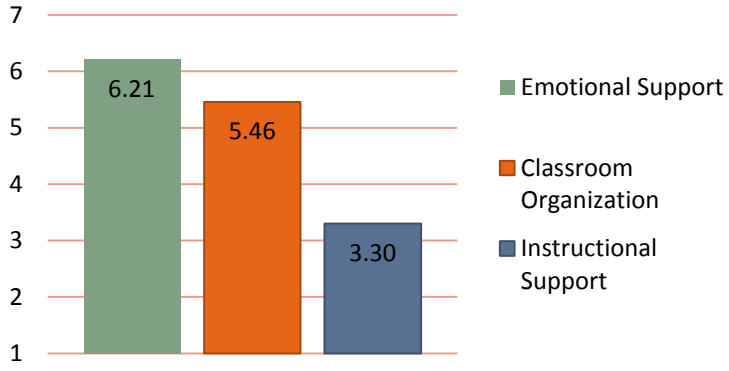
Analyses were conducted to test whether any of the child and family background characteristics were associated with Total Qualistar Rating Points. Total Rating Points was not associated with child primary language, home language or ethnicity.[51] There was a significant association between tier level and number of rating points earned by the child's preschool.[52] Follow-up Tukey tests revealed that children from Income Tier 1 were enrolled in preschools that earned a significantly higher number of rating points than children from Income Tiers 3-5.[53] This difference was of relatively small size, however, just 1.33 points. There was also a significant association between region of the city and the number of rating points earned by the child's preschool.[54] Follow-up Tukey tests revealed that children residing in the northwest region of the city tended to be enrolled in programs earning the lowest number of rating points on average. The average number of points earned by programs attended by children who live in the northwest region was significantly lower than what was earned by programs attended by children in the southwest, northeast, and southwest regions of the city.[55] These differences were of small magnitude, however, between 2.5 and 2.9 points.[56]

CLASS OBSERVATIONS

Figure 3 displays the mean scores for the 123 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 in all of these areas 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, average scores for

[50] t=3.65, df=66.33, p<.001
[51] Child primary language: F(1,173)=2.57, n.s.; primary home language: F(1,175)=3.87, n.s.; ethnicity: F(4,188)=2.35, n.s.
[52] F(2,189)=3.16, p<.05
[53] Income Tier 1, mean=32.42, sd=2.81; Income Tier 2, mean=31.86, sd=3.30; Income Tiers 3-5, mean=31.09, sd=3.41
[54] F(4,189)=6.16, p<.001
[55] northwest, mean=29.72, sd=3.27; northeast, mean=32.56, sd=2.63; southeast, mean=32.25, sd=2.75; southwest, mean=32.59, sd=3.52; central, mean=31.18, sd=2.81
[56] Rating points: Central Region-mean=31.90, sd=2.48; Northeast Region-mean=31.76, sd=3.10; Northwest Region-mean=29.78, sd=3.47; Southeast Region-mean=31.05, sd=2.55; Southwest Region-mean=31.51, sd=3.85

Figure 3: Average CLASS Domain Scores (n=123 Classrooms)



Classroom Organization tended to be in the 4.5-5 point range, and average scores for Instructional Support tended to be 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, nearly 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, very few classrooms scored in the high range, and the remainder were split fairly evenly between the low and middle-ranges.

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 Emotional Support and Classroom Organization were significantly higher, on average, in DPS classrooms than in community-based preschool classrooms.[57] There was a trend toward a significant difference between the two provider types in Instructional Support as well.[58] Differences between the two provider types were, on average, about one-third of a point in magnitude for all three scales.

Figure 4: Distribution of Scores for Emotional Support (n=123 Classrooms)

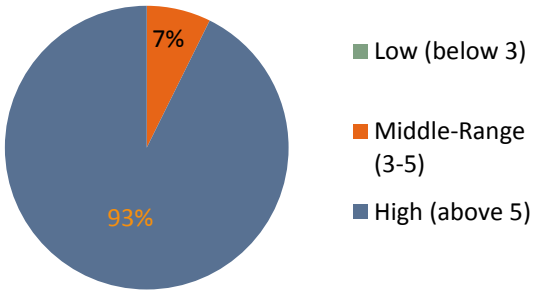


Figure 5: Distribution of Scores for Classroom Organization (n=123 Classrooms)

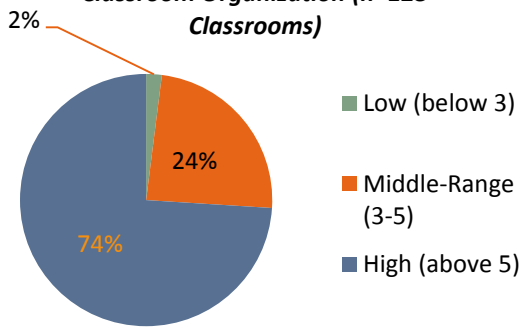
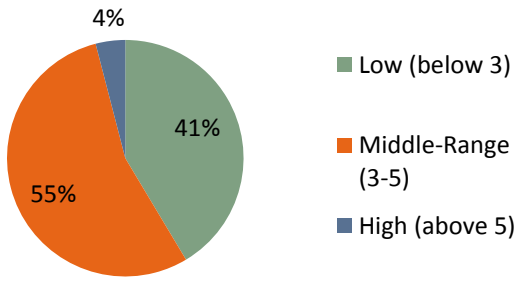


Figure 6: Distribution of Scores for Instructional Support (n=123 Classrooms)*

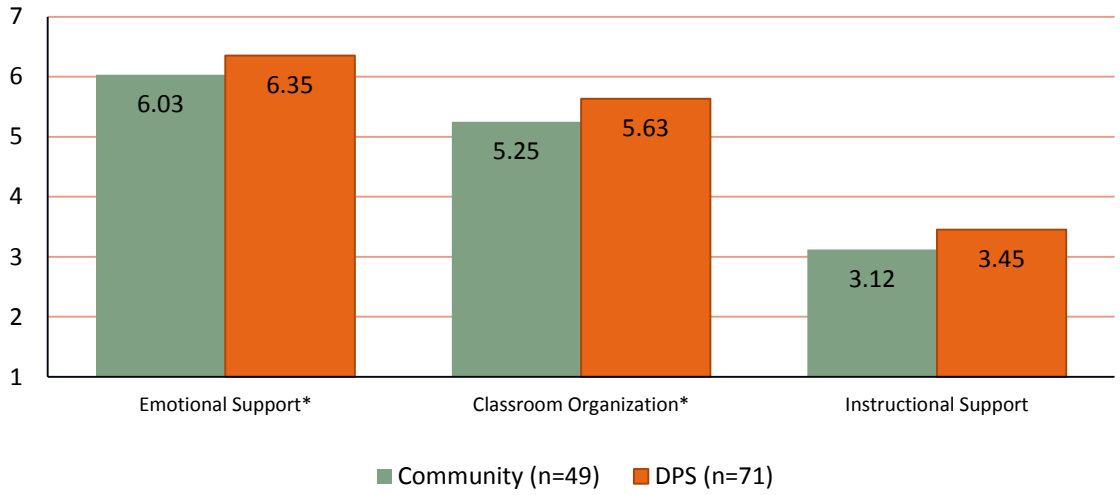


*Percentages do not sum to 100 because of rounding error.

[57] Emotional Support—t=2.54, df=118, p<.05; Classroom Organization—t=2.22, df=79.92, p<.05.
[58] T=1.89, df=118, p<.10

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 ratio was 3.2 children for every adult and the largest ratio was 10.8 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.[59] Finally, we examined whether CLASS domain scores were associated with the size of the preschool program. We computed correlations between the number of preschool classrooms at the site and CLASS domain scores. These correlations were also non-significant.[60]

Figure 7: CLASS Domain Scores by Provider Type



*p<.05
^Standard deviations: Emotional Support—Community=0.74, DPS=0.61; Classroom Organization—Community=1.05, DPS=0.74; Instructional Support—Community=0.91, DPS=1.00

As with the Qualistar Rating, we conducted analyses to test for associations between CLASS domain scores and child and family background characteristics. A small but significant association was observed between income tier and Emotional Support.[61] Follow up Tukey tests revealed that the children in Tier 1 tended to be enrolled in classrooms with higher levels of Emotional Support than children in Tiers 3-5. The difference between these groups was relatively small, less than a third of a point.[62] Child primary language was significantly associated with Instructional Support.[63] Children whose primary language was English tended to be enrolled in classrooms demonstrating higher levels of Instructional Support, about a half of a point higher, than children whose primary language was something other than English.[64] Home language was significantly associated with both Classroom Organization and Instructional Support.[65] Children whose home language was English tended to be enrolled in

[59] Ratio with Emotional Support, r=.06; ratio with Classroom Organization, r=.00; ratio with Instructional Support, r=.05; all non-significant.
[60] Number of preschool classrooms with Emotional Support, r=.14; number of preschool classrooms with Classroom Organization, r=.05; number of preschool classrooms with Instructional Support, r=.14.
[61] F(3,168)=2.81, p<.05
[62] Means for Emotional Support: Tier 1, mean=6.48, sd=.70; Tier 2, mean=6.23, sd=.67; Tiers 3-5, mean=6.20, sd=.55; Tier 6 (income not reported), mean=6.18, sd=.57
[63] F(1,153)=9.74, p<.01
[64] Means for Instructional Support by child primary language: English, mean=3.64, sd=1.01; Another Language, mean=3.15, sd=0.89
[65] Classroom Organization: F(1,154)=4.88, p<.05; Instructional Support: F(1,154)=9.13, p<.01

classrooms demonstrating higher levels of Classroom Organization and Instructional Support than children with a another home language, on average. The magnitude of the difference was about a fourth of a point for Classroom Organization and nearly a half of a point for Instructional Support.[66] There were significant differences by racial/ethnic group in Instructional Support.[67] Follow-up Tukey tests revealed that Hispanic children tended to be enrolled in classrooms demonstrating lower levels of Instructional Support than white and black children. The magnitude of the difference between the means for Hispanic and black children was nearly 1 point. The difference between Hispanic and white children was slightly more than half of a point.[68] Finally, Instructional Support differed by region of the city.[69] Follow-up Tukey tests revealed that children residing in the southeast region of the city tended to be enrolled in classrooms demonstrating higher levels of Instructional Support than children residing in the southwest and northwest regions of the city. The magnitude of these differences was about 1 point.[70]

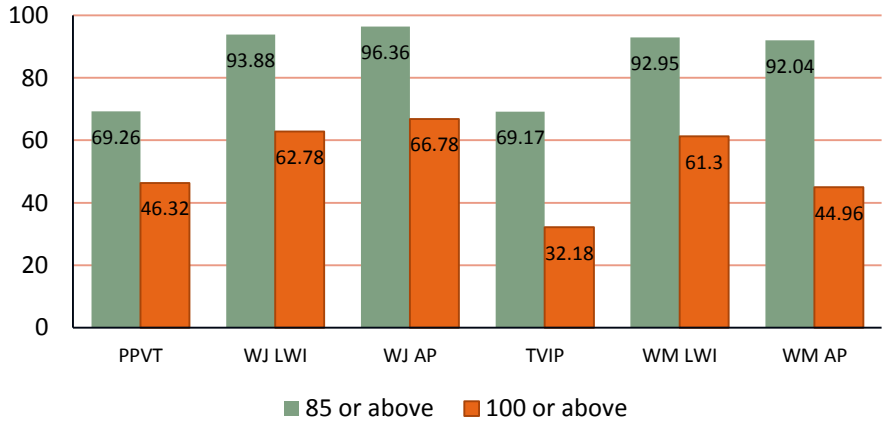
KINDERGARTEN READINESS

STANDARDIZED ASSESSMENTS

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

For the English assessments, the vast majority of children (over 90%) 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

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



[66] Means for Classroom Organization by home language: English, mean=5.66, sd=.81; Another Language, mean=5.39, sd=.74; Means for Instructional Support by home language: English, mean=3.65, sd=1.00; Another Language, mean=3.19, sd=.93

[67] F(3,167)=7.54, p<.0001

[68] Means for Instructional Support by child race/ethnicity: Black, mean=4.10, sd=1.62; Hispanic, mean=3.16, sd=0.89; White, mean=3.72, sd=0.85; Other, mean=3.72, sd=0.74

[69] F(4,168)=3.80, p<.01

[70] Means for Instructional Support by region of the city: Central, mean=3.43, sd=.84; Northeast, mean=3.54, sd=.95; Northwest, mean=3.08, sd=.70; Southeast, mean=4.19, sd=1.20; Southwest, mean=3.30, sd=1.05

Problems. In contrast, only about 70% 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. The vast majority of children whose primary language was English (94%) scored 85 or above on the PPVT as compared with a relatively small proportion of children whose primary language was not English (29%).[71] A similar, but less pronounced pattern was observed for WJ LWI. Nearly all children (99%) whose primary language was English scored 85 or above on WJ LWI as compared with 87% of children whose primary language was not English.[72] There was not a significant association between scoring 85 or above on WJ Applied Problems and primary language.[73] 98% of children whose primary language was English scored 85 or above compared with 95% of children whose primary language was not English.

A more pronounced pattern of results emerged when a score of 100 was used as the cutoff. For PPVT, three-quarters of children whose primary language was English earned a score of 100 or greater as compared with just 5% of children with another primary language.[74] For WJ LWI, 81% of children whose primary language was English scored 100 or greater as compared with 34% of children with another primary language.[75] Finally, for WJ Applied Problems, 81% of children whose primary language was English earned scores of 100 or above compared with 46% of children whose primary language was something other than English.[76]

For assessments administered in Spanish, scores were once again stronger for LWI and Applied Problems than for vocabulary (TVIP). Approximately nine-tenths of children scored 85 or above on WM LWI and Applied Problems while over two-thirds of children scored 85 or above on the TVIP. 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. Over 80% of children met or exceeded the cutoff of 85 in at least one language in the area of receptive vocabulary (i.e., PPVT or TVIP). Nearly all children met or exceeded the cutoff of 85 in at least one language on the literacy assessment (WJ-LWI or WM-LWI) and the math assessment (WJ-AP or WM-AP). When a score of 100 was used as a cutoff, over half of children met or exceeded this benchmark for vocabulary; approximately three-quarters met or exceeded this benchmark for both literacy and math.

---

[71] χ^2_1=84.20, p<.0001
[72] χ^2_1=12.64, p<.001
[73] χ^2_1=1.59, n.s.
[74] χ^2_1=83.05, p<.0001
[75] χ^2_1=40.98, p<.0001
[76] χ^2_1=23.08, p<.0001

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 50 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, nearly three-quarters of children met or exceeded this threshold in at least one language. For literacy and math, over 90% of children met or exceeded this threshold.

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

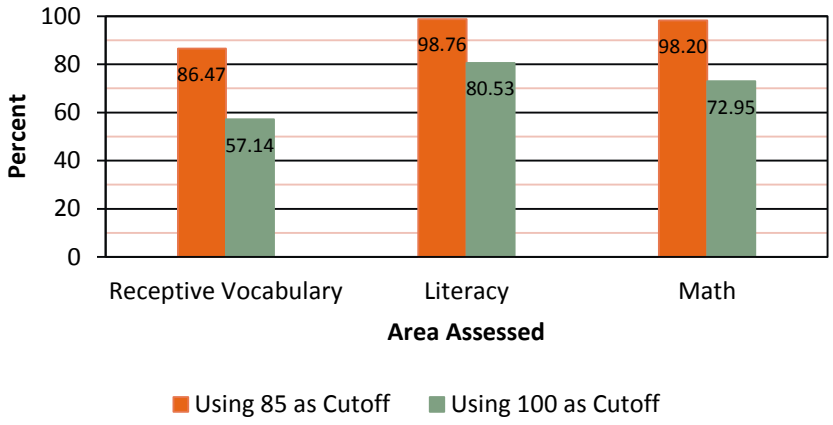
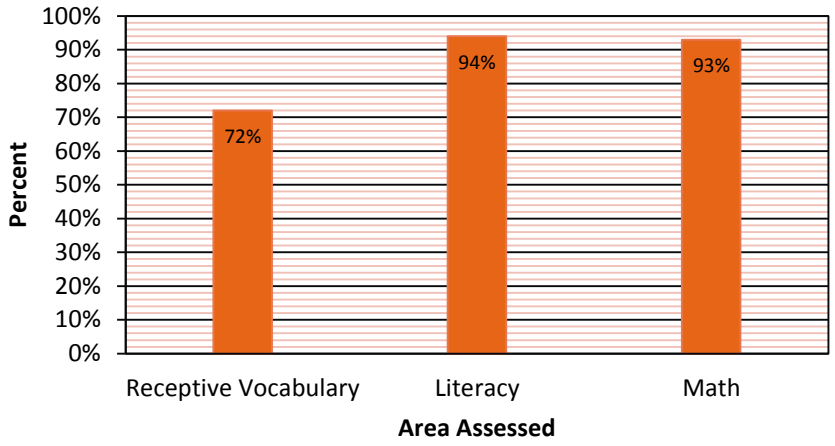


Figure 10: Weighted Percent of Children Scoring 92.5 or Above on Spring Standardized Assessments in Spanish or English

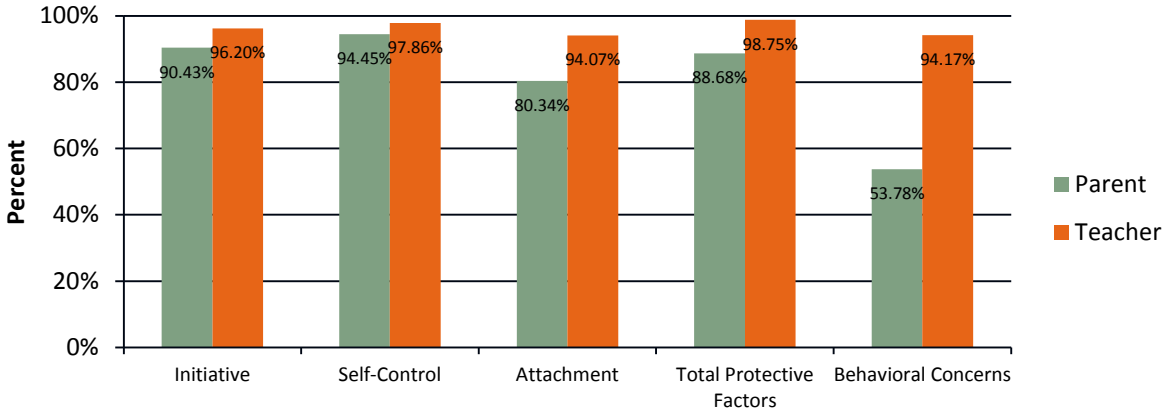


PARENT AND TEACHER SURVEYS

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 80% of children in the typical or strength range for Attachment and slightly over

half in the typical range for Behavioral Concerns. Teachers rated over 90% of children in the typical or strength range for all of the DECA subscales.

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, 5.0 (sd=10.3), 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 third of the sample, the teacher's rating was significantly greater than the parent's rating. For a relatively small proportion of the sample, the parent's rating was significantly greater than teacher's rating.

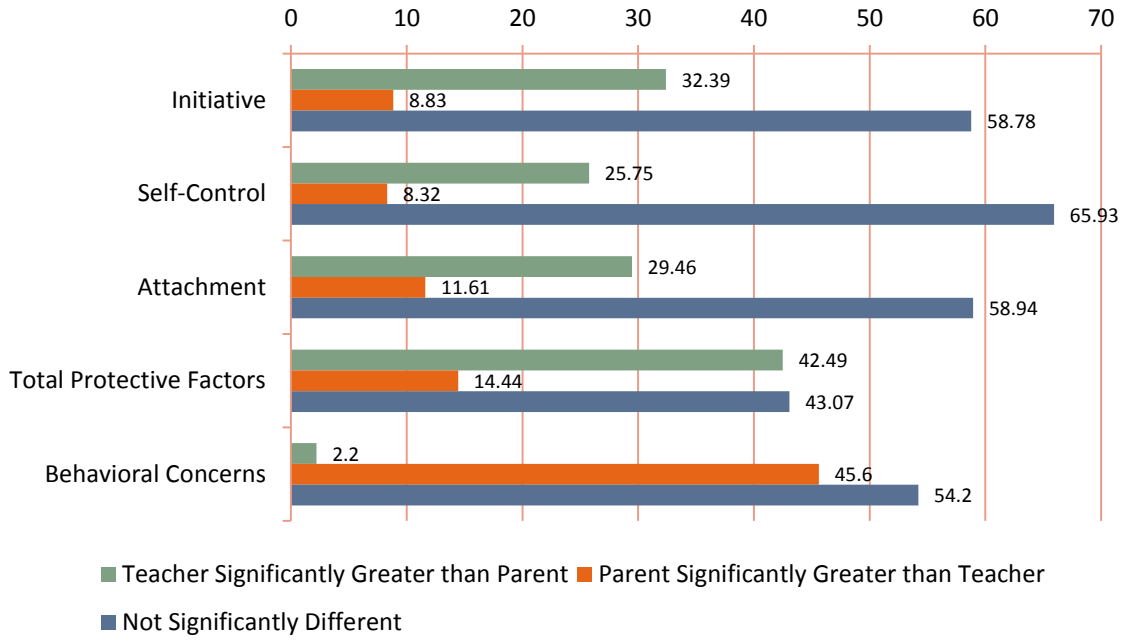
For Self-Control, a difference of 10 is needed to conclude that there is a significant difference between the parent and teacher rating. The average difference between teachers' and parents' reports, 4.7 (sd=10.7), did not exceed this threshold, indicating that, on average, teachers' and parents' ratings did not differ. As displayed in Figure 12, for about two-thirds of children, teachers' and parents' reports did not significantly differ. For about a quarter of children, teachers rated children significantly higher than did parents. Only a small proportion 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 parent's and teacher's scores, 4.7 (sd=12.2), did not exceed this threshold. Once again, for over 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 parent's and teacher's ratings. Across the sample, the average difference between ratings for Protective Factors was 5.3 (sd=10.1), which was below that threshold. However, for over four-tenths of

children, the teacher's rating was significantly greater than the parent's rating (see Figure 12). For a similar proportion of children, there was not a significant difference between raters.

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 10.4 (sd=11.4), which did not reach this threshold. For over 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 about 2.5 to 3.5 times more often as 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 3 points, or about one-fifth of a standard deviation. For Applied Problems, the increase was not significant. 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, change over time was not significant for the TVIP. However, significant and large increases were observed for WM LWI

and WM AP. For WM LWI, on average, scores increased nearly 6 points, two-fifths of a standard deviation. For WM AP, on average, scores increased over four points, or about a quarter 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	197	91.67 (23.90)	94.93 (21.91)	4.54***
WJ-LWI	197	99.69 (12.56)	102.48 (12.18)	5.19***
WJ-AP	197	105.47 (14.96)	105.74 (12.48)	0.33
Standardized Assessments—Spanish				
TVIP	51	89.95 (19.38)	91.97 (19.46)	1.19
WM-LWI	51	97.36 (14.13)	103.07 (14.20)	4.95***
WM-AP	52	94.23 (13.38)	98.57 (14.29)	3.27**
Teacher Survey				
Initiative T-Score[1]	196	51.29 (8.26)	56.43 (7.63)	6.05***
Self-Control T-Score	196	55.77 (8.29)	59.38 (7.56)	5.01***
Attachment T-Score	196	49.43 (8.37)	52.78 (8.57)	4.03***
Total Protective Factors T-Score	196	51.61 (8.09)	56.43 (7.30)	5.76***
Behavioral Concerns T-Score	183	49.13 (8.96)	46.94 (7.79)	2.18*

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

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

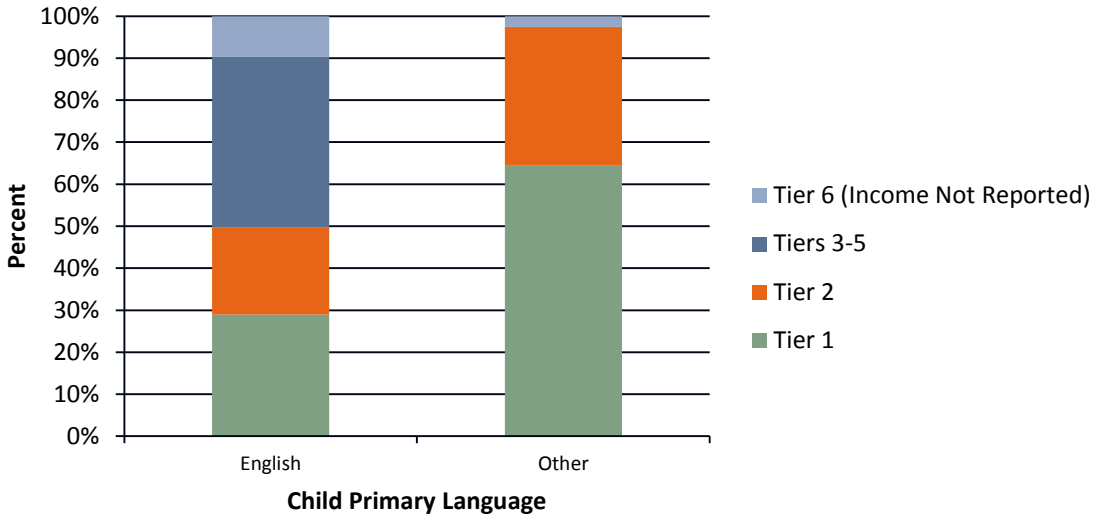
Significant improvements were also observed in all of the teachers' ratings on the DECA over the course of the school year. Change over time was significant and positive for all of the Protective Factors (Initiative, Self-Control, and Attachment, as well as Total Protective Factors). Teachers also reported significant decreases in Behavioral Concerns over the course of the school year. Of particular note are the increases in Initiative and Total Protective Factors (over one-third 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).[77] It is important to note that these two background characteristics, income tier and child's primary language, are strongly associated (see Figure 13).[78] 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 ANOVAs[79] 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 WJ LWI,[80] WJ AP,[81] DECA Initiative,[82] DECA Attachment,[83] and DECA Total Protective Factors.[84] Results of these analyses are depicted in Figures 14-18.

Figure 14 shows average WJ LWI 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

[77] 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 each of these categories.

[78] $\chi^2_3=67.58$, $p<.0001$

[79] 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.

[80] $F(3,193)=11.89$, $p<.0001$

[81] $F(3,193)=3.18$, $p<.05$

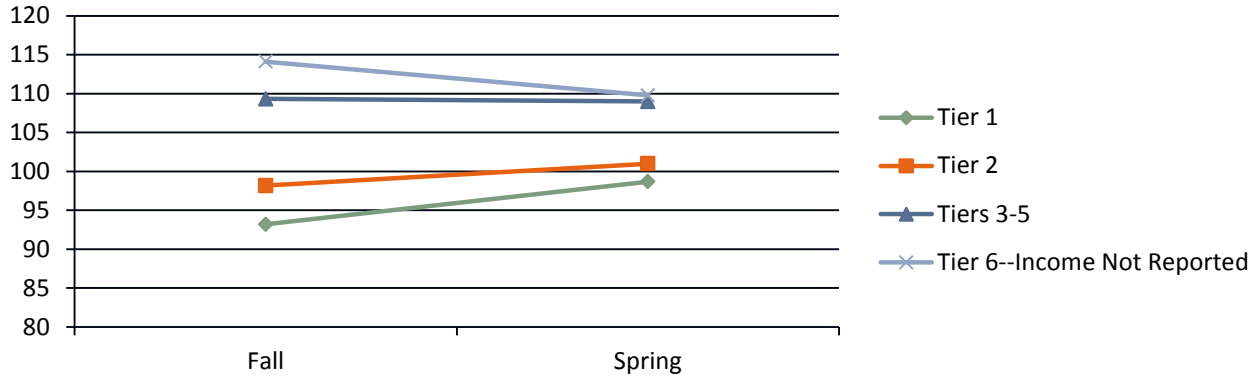
[82] $F(3,192)=2.68$, $p<.05$

[83] $F(3,192)=3.34$, $p<.05$

[84] $F(3,192)=3.16$, $p<.05$

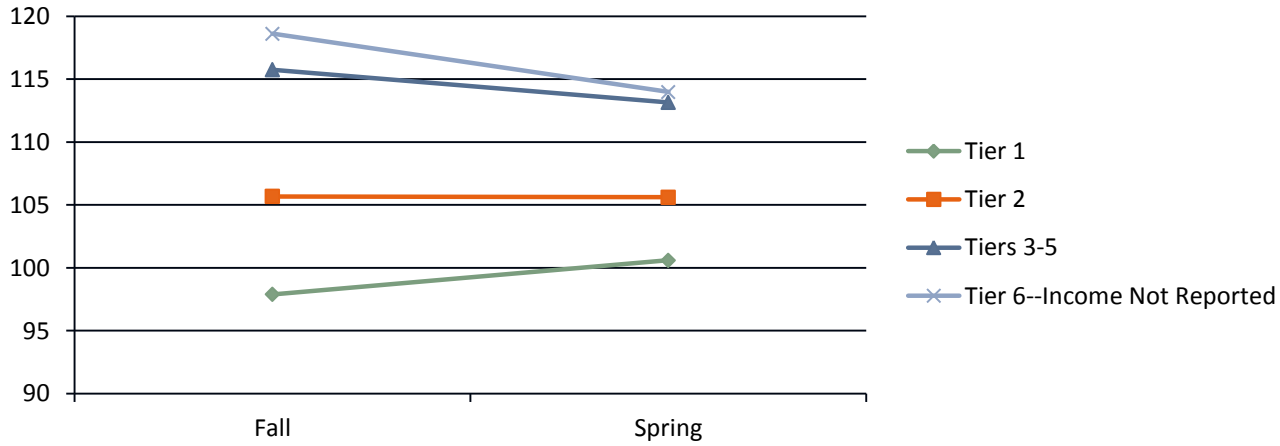
the higher tier groups. Follow-up Tukey[85] tests revealed that children from Tier 1 increased significantly more over time than children in Tiers 3-5 and children in Tier 6, the group that did not report income. Pairwise comparisons of the other groups were not significant.

Figure 14: Weighted WJ Letter-Word Identification Standard Scores over Time, by Income Tier Group[1]



[1]Standard Deviations: Tier 1: Fall=12.26, Spring=13.26; Tier 2: Fall=9.54, Spring=10.69; Tiers 3-5: Fall=8.36, Spring=10.26; Tier 6: Fall=9.68, Spring=8.74.

Figure 15: Weighted WJ Applied Problems Standard Scores over Time, by Income Tier Group[1]



[1]Standard Deviations: Tier 1: Fall=16.96, Spring=14.62; Tier 2: Fall=12.74, Spring=10.11; Tiers 3-6: Fall=7.86, Spring=9.03; Tier 6: Fall=9.19, Spring=5.40.

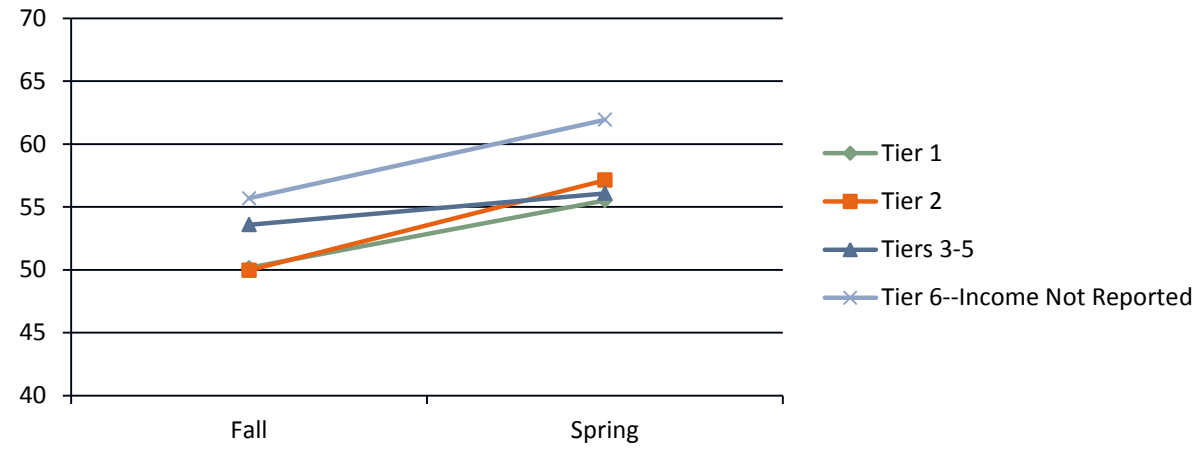
Results of the analysis of WJ AP are presented in Figure 15. Follow-up Tukey tests revealed a significant difference in change over time between Tier 1 and Tiers 3-5. Children in Tier 1 increased significantly more over

[85] 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.

the course of the school year than children in Tiers 3-5. Children Tiers 3-5 decreased slightly over the course of the year, but still finished the year with very high scores, on average.

Results of the analysis of DECA Initiative are presented in Figure 16. Follow-up Tukey tests revealed that Tier 2 was significantly different than the Tiers 3-5 group. Children in Tier 2, on average, increased over time, while scores for the Tiers 3-5 remained relatively flat over time, on average.

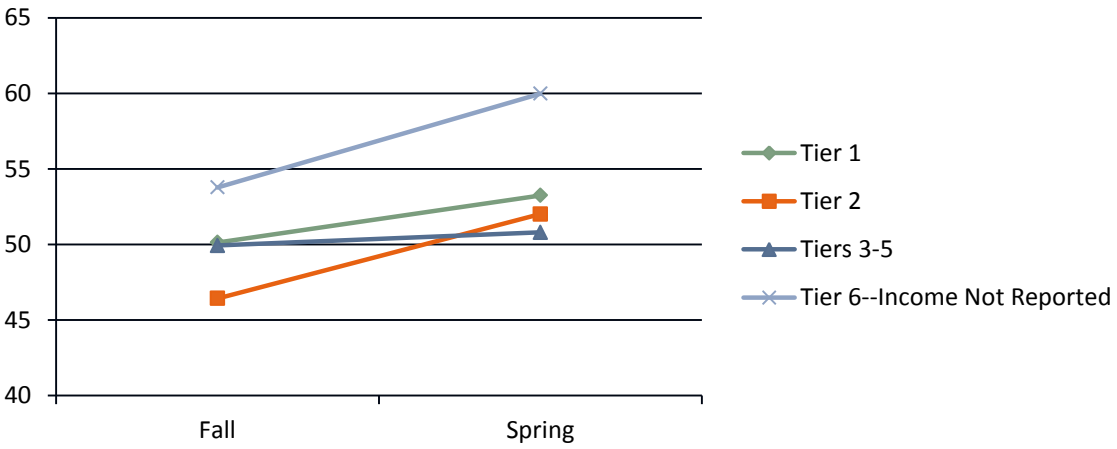
Figure 16: Weighted Teachers' Ratings of Children's Initiative over Time, by Income Tier Group[1]



[1]Standard Deviations: Tier 1: Fall=10.09, Spring=8.62; Tier 2: Fall=6.98, Spring=7.74; Tiers 3-5: Fall=6.92, Spring=6.27; Tier 6: Fall=6.11, Spring=6.88.

Results for teachers' ratings on DECA Attachment over time are presented in Figure 17. Follow-up Tukey tests revealed a significant difference between Tier 2 and the Tier 3-5 group. Children in Tier 2 increased over time, on average, in teacher ratings of Attachment, while scores for children in Tiers 3-5 remained stable over time.

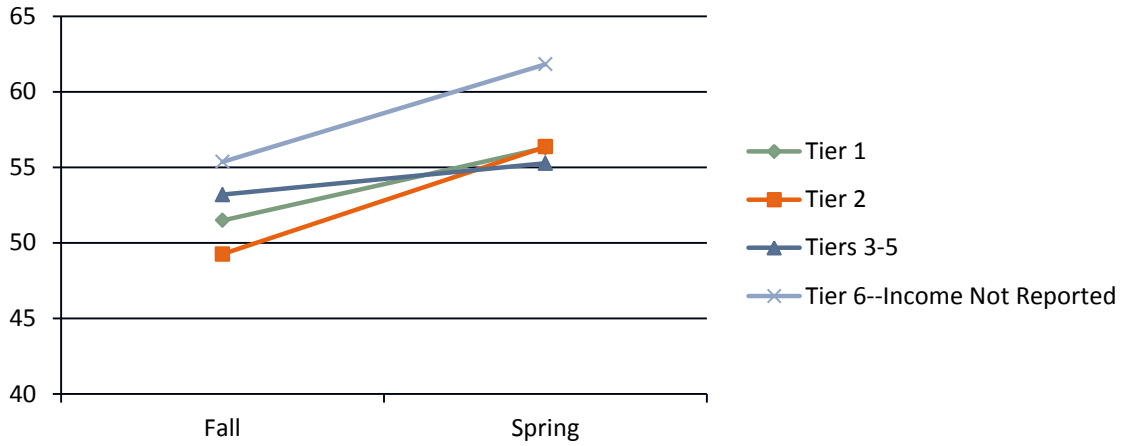
Figure 17: Weighted Teachers' Ratings of Children's Attachment over Time, by Income Tier Group[1]



[1]Attachment was measured with the DECA. Standard Deviations: Tier 1: Fall=9.12, Spring=9.03; Tier 2: Fall=7.68, Spring=7.24; Tiers 3-5: Fall=7.74, Spring=8.06; Tier 6: Fall=5.59, Spring=10.20.

Results for teachers' ratings on DECA Total Protective Factors over time are presented in Figure 18. Follow-up Tukey tests revealed a significant difference between Tier 2 and the Tier 3-5 group. Children in Tier 2 increased over time, on average, in teacher ratings of Attachment, while scores for children in Tiers 3-5 remained relatively stable over time.

Figure 18: Weighted Teachers' Ratings of Children's Total Protective Factors over Time, by Income Tier Group[1]



[1]Attachment was measured with the DECA. Standard Deviations: Tier 1: Fall=9.12, Spring=9.03; Tier 2: Fall=7.68, Spring=7.24; Tiers 3-5: Fall=7.74, Spring=8.06; Tier 6: Fall=5.59, Spring=10.20.

The time by income tier group interaction was non-significant for all of the remaining variables tested: all of the standardized assessments administered in Spanish and the remainder of the teacher-rated DECA scales. This indicates that children progressed in a similar fashion, on average, on each of these assessments regardless of income tier group.

CHILDREN'S PRIMARY LANGUAGE

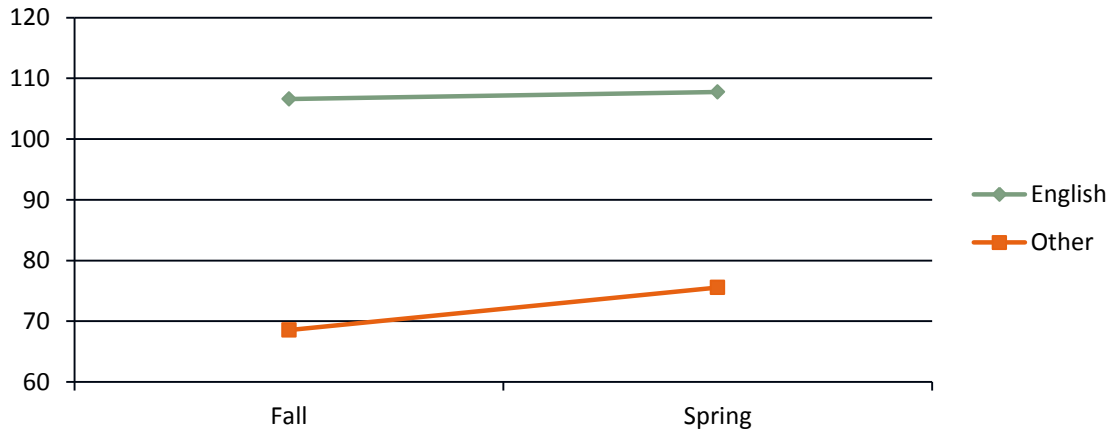
A series of Repeated Measures ANOVAs was conducted with primary language predicting scores over time on assessments administered in English and teacher-rated DECA.[86] There were significant interactions between primary language group and time for PPVT[87] and WJ AP.[88] Results of these analyses are presented in Figures 19 and 20.

[86] 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.

[87] $F(1,178)=16.23$, $p<.0001$

[88] $F(1,178)=18.89$, $p<.0001$

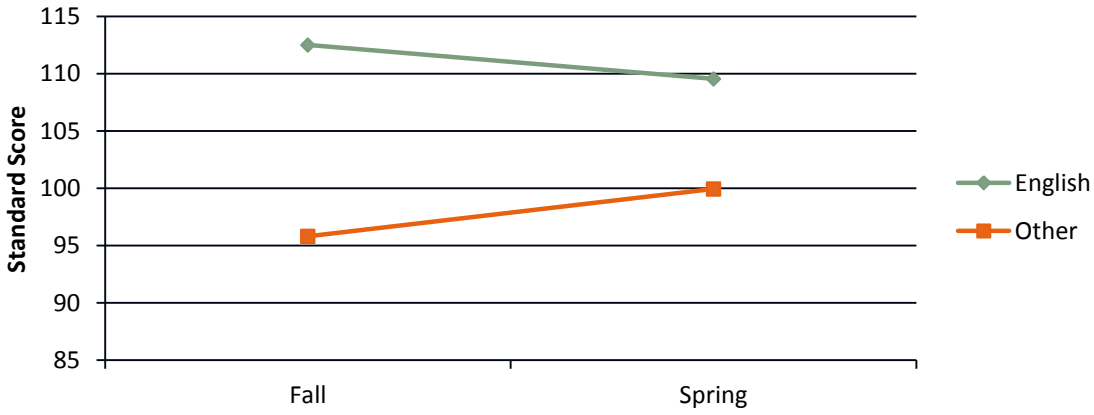
Figure 19: Weighted PPVT Scores over Time, by Child Primary Language[1]



[1]Standard Deviations: English: Fall=17.01, Spring=18.76; Other: Fall=14.97, Spring=14.20.

For both 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 19), children whose primary language was English increased only slightly, while their counterparts with another primary language increased an average of about half of a standard deviation. For WJ AP (see Figure 20), children whose primary language was English started the year with very high scores and decreased slightly over the course of the year. In contrast, children with another primary language increased by about one-third of a standard deviation on average. For LWI and all of the DECA subscales, 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 20: Weighted Woodcock-Johnson AP Scores over Time, by Child Primary Language[1]



[1]Standard Deviations: English: Fall=15.51, Spring=13.67; Other: Fall=10.86, Spring=9.92.

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.

The pattern of findings for these analyses was weak and inconsistent. For the English academic assessments (vocabulary, literacy and math), there were two significant correlations. Unexpectedly, Classroom Organization was significantly and negatively associated with PPVT-4 (r=-.16, p<.05). That is, after taking into account children's vocabulary skills in the fall, higher levels of Classroom Organization were associated with lower vocabulary skills as assessed in English in the spring. In contrast, Classroom Organization was positively associated with Letter-Word Identification assessed in English (r=.20, p<.05). After taking into account literacy skills in the fall, higher levels of Classroom Organization were associated with higher literacy skills assessed in English in the spring. There were no significant correlations for Spanish assessments or teacher DECA ratings.

RESULTS: ELEMENTARY SCHOOL

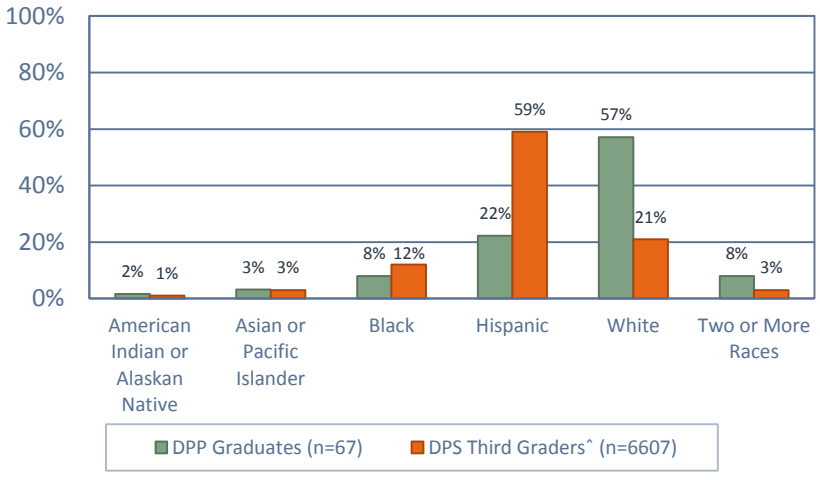
HOW SIMILAR ARE DPP GRADUATES TO THE POPULATION OF CHILDREN IN THE DISTRICT AS A WHOLE?

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.[89] CDE provides information about free and reduced lunch status for the district as a whole as well as gender and race/ethnicity for each grade level.

COHORT 0

Figure 21 displays a comparison of the racial/ethnic composition of the sample of Cohort 0 DPP graduates with TCAP reading assessment data for school year 11-12 and the population of children enrolled in third grade as of fall 2011. As expected based on the nature of the sample (a sample of convenience), Cohort 0 children have a dramatically different racial and ethnic composition than the district

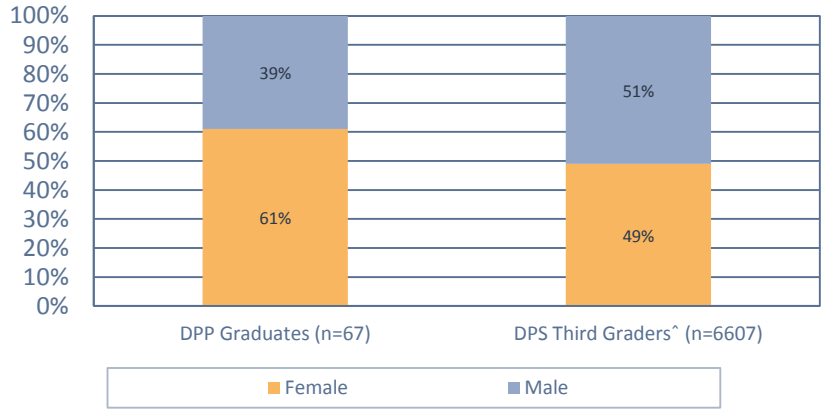
Figure 21: Race/Ethnicity of Cohort 0 DPP Graduates and All Third Graders Enrolled in DPS in Fall 2011*



*This figure includes only DPP Graduates for whom TCAP reading data were available for the 11-12 school year and omits 4 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.

[89] Available at: http://www.cde.state.co.us/cdereval/rvprioryearpmdata.htm

*Figure 22: Gender of Cohort 0 DPP Graduates and All Third Graders Enrolled in DPS in Fall 2011**



*This chart includes only Cohort 0 DPP Graduates for whom TCAP reading assessment data were available for the 11-12 school year.
^The group of DPS Third Graders includes the DPP graduates.

as a whole. Notably, Cohort 0 children are nearly three times as likely to be white than children in the district as a whole. They are about a third as likely to be Hispanic as children in the district as a whole.

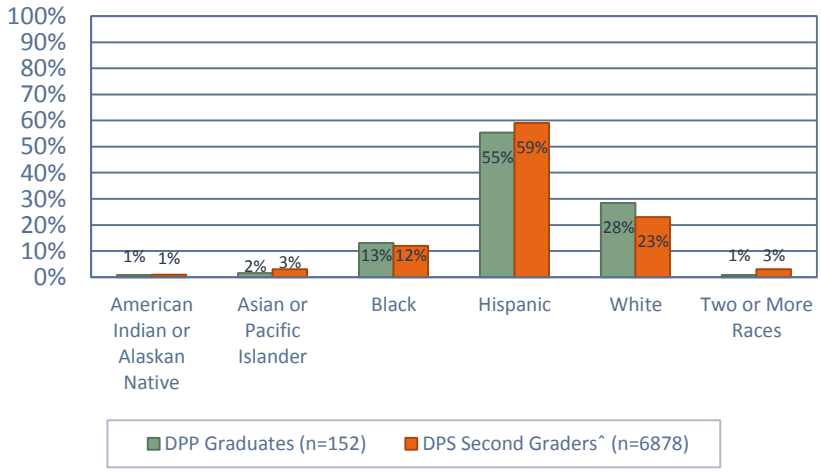
Figure 22 displays the gender composition for these same groups of children. Again, we see that, four years after DPP, Cohort 0 children differ markedly from the district's third graders as a whole. While the group of third graders enrolled in DPS in fall 2011 were evenly divided between boys and girls. Cohort 0 children with reading assessment data in spring 2012 were comprised of nearly two-thirds girls and more than one-third boys.

In sum, the Cohort 0 sample was not selected via random selection as originally planned. Instead, this cohort is comprised of most of the very first families that signed up for DPP when the program was in its infancy. It is not surprising that this sample is so dissimilar to the population enrolled in DPS as a whole. As a result, one should use extreme caution when interpreting comparisons of reading assessment scores for this group to the district as a whole. Any differences observed could be due to participation in DPP, factors related to the demographic differences between the two groups, or unmeasured characteristics.

COHORT 1

Figure 23 displays a comparison of the racial/ethnic composition of the sample of Cohort 1 DPP graduates with reading assessment data and the population of children enrolled in second grade in DPS as of fall 2011. The racial and ethnic compositions of the two groups are remarkably similar.

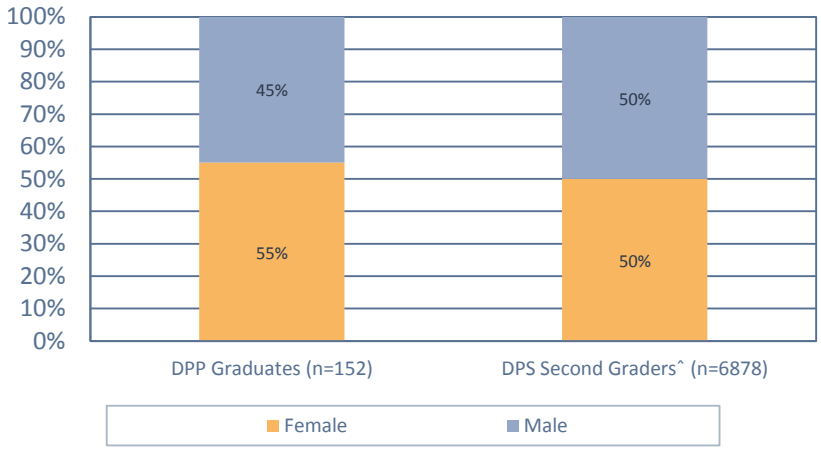
*Figure 23: Race/Ethnicity of Cohort 1 DPP Graduates and All Second Graders Enrolled in DPS in Fall 2011**



*This figure includes only Cohort 1 DPP Graduates for whom reading assessment data were available for the 11-12 school year and omits 22 DPP graduates with missing or ambiguous race/ethnicity data in the database of DPP children provided to Clayton by ACS. ^The group of DPS Second Graders includes the DPP graduates. Percentages sum to 101 because of rounding error.

Figure 24 presents the gender composition of Cohort 1 DPP graduates with reading assessment data and all second graders in DPS as of fall 2011. The district as a whole was split evenly between the genders. Cohort 1 DPP Graduates with spring 2012 reading assessment data were split a bit less evenly, with slightly more girls than boys, but still rather close to an even split.

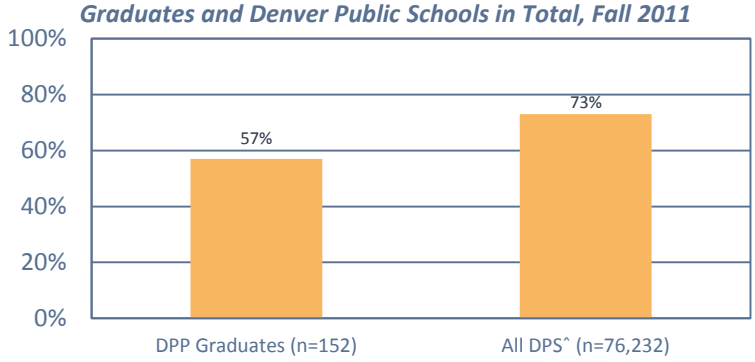
Figure 24: Gender of Cohort 1 DPP Graduates and All Second Graders Enrolled in DPS in Fall 2011*



*This figure includes only Cohort 1 DPP Graduates for whom reading assessment data were available for the 11-12 school year.
^The group of DPS Second Graders includes the DPP graduates.

Finally, Figure 25 presents the proportion of children eligible for free and reduced lunch for the district as a whole and for the sample of DPP graduates.[90] CDE does not provide free and 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 and 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.

Figure 25: Free/Reduced Lunch Status of Cohort 1 DPP Graduates and Denver Public Schools in Total, Fall 2011



^The group All DPS includes the DPP graduates.

In sum, Cohort 1 children with spring 2011 reading assessment data were similar to the district in terms of their ethnic and gender composition. A smaller proportion of children in Cohort 1 qualified for free and 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.

[90] We were not able to examine free and reduced lunch for Cohort 0 because our evaluation data file did not include information about family size. Income cut offs for free and reduced lunch vary by the size of the family.

COHORT 2

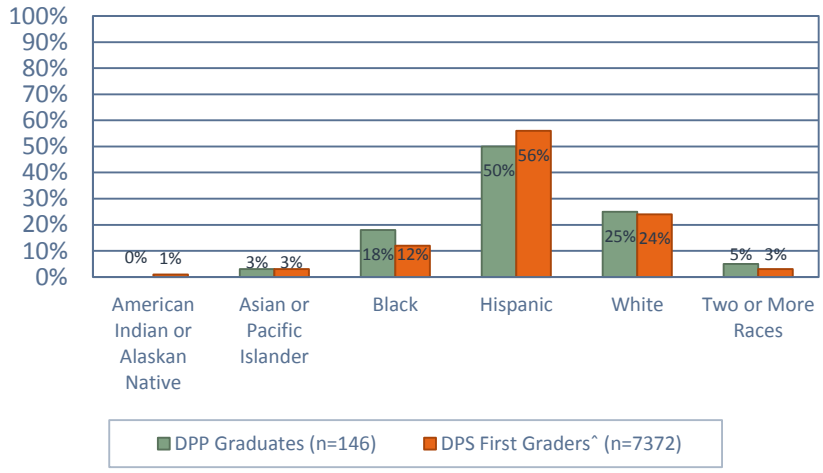
Figure 26 displays a comparison of the racial/ethnic composition of the sample of Cohort 2 DPP graduates with reading assessment data and the population of children enrolled in first grade in DPS as of fall 2011. The racial and ethnic compositions of the two groups are remarkably similar.

Figure 27 presents the gender composition of Cohort 2 DPP graduates with reading assessment data and all first graders in DPS as of fall 2011. The district as a whole was split very close to evenly between the genders. Cohort 2 DPP Graduates with spring 2012 reading assessment data were split a bit less evenly, with slightly fewer girls than boys, but still rather close to an even split.

Figure 28 presents the proportion of children eligible for free and reduced lunch for the district as a whole and for the sample of DPP graduates. About two-thirds of Cohort 2 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.

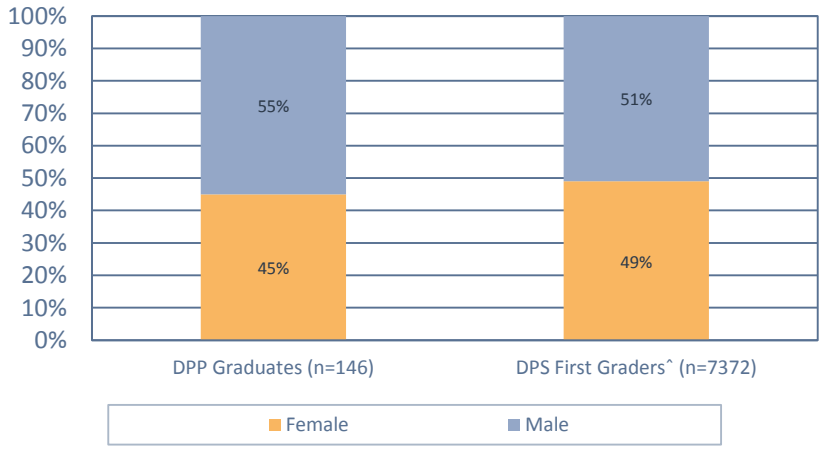
In sum, Cohort 2 children with spring 2012 reading assessment data were similar to the district in terms of their ethnic and gender composition. A smaller proportion of children in Cohort 2 qualified for free and reduced lunch than for the district as whole, suggesting that

Figure 26: Race/Ethnicity of Cohort 2 DPP Graduates and All First Graders Enrolled in DPS in Fall 2011*



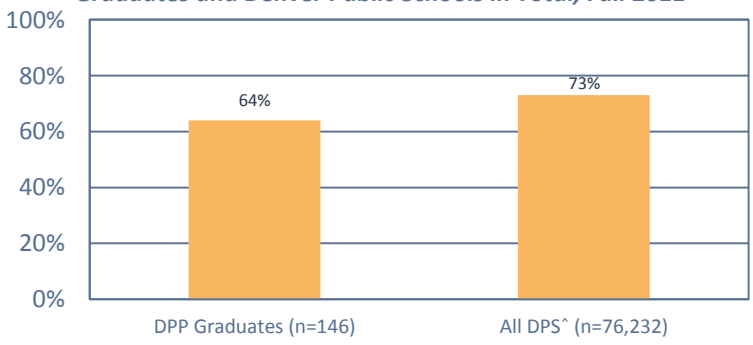
*This figure includes only Cohort 2 DPP Graduates for whom reading assessment data were available for the 11-12 school year. Percentages for DPP Graduates are weighted and do not sum to 100 because of rounding error.
^The group of DPS First Graders includes the DPP graduates. Percentages do not sum to 100 because of rounding error.

Figure 27: Gender of Cohort 2 DPP Graduates and All First Graders Enrolled in DPS in Fall 2011*



*This figure includes only Cohort 2 DPP Graduates for whom reading assessment data were available for the 10-11 school year.
^The group of DPS First Graders includes the DPP graduates.

Figure 28: Free/Reduced Lunch Status of Cohort 2 DPP Graduates and Denver Public Schools in Total, Fall 2011



^The group All DPS includes the DPP graduates.

this sample might be slightly wealthier. The magnitude of the difference between proportions qualifying for free and reduced lunch was much smaller than was observed with Cohort 1, however. 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.

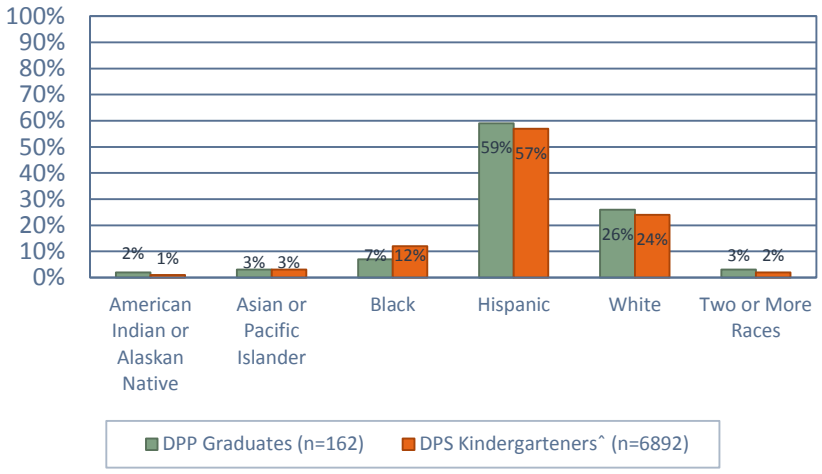
COHORT 3

Figure 29 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 kindergarten in DPS as of fall 2011. The racial and ethnic compositions of the two groups are remarkably similar.

Figure 30 presents the gender composition of Cohort 3 DPP graduates with reading assessment data and all kindergarteners in DPS as of fall 2011. Both the group of Cohort 3 graduates with spring 2012 reading assessment data and the district as a whole were split very close to evenly between the genders.

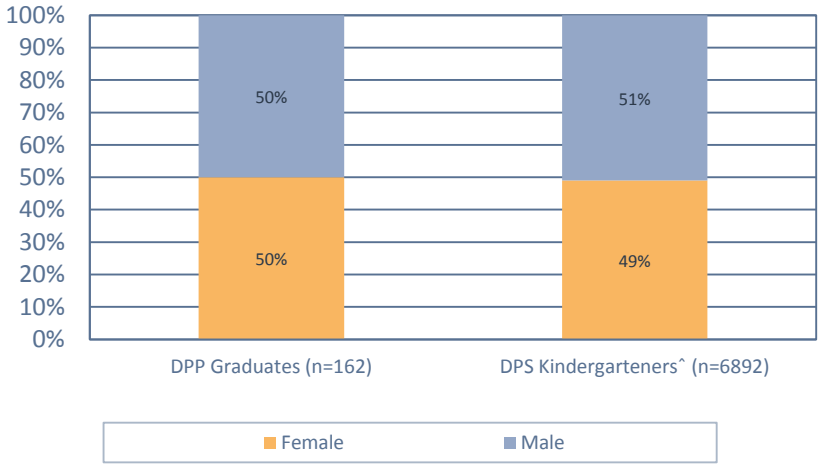
Figure 31 presents the proportion of children eligible for free and reduced lunch for the district as a whole and for the sample of DPP graduates. Two-thirds of Cohort 3 graduates with reading assessment data qualified for free or reduced

Figure 29: Race/Ethnicity of Cohort 3 DPP Graduates and All Kindergarteners Enrolled in DPS in Fall 2011*



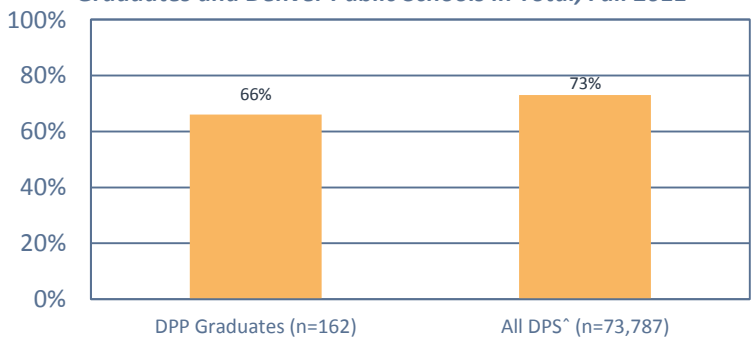
*This figure includes only Cohort 3 DPP Graduates for whom reading assessment data were available for the 11-12 school year. Percentages for DPP Graduates are weighted.
^The group of DPS Kindergarteners includes the DPP graduates. Percentages do not sum to 100 becuase of rounding error.

Figure 30: Gender of Cohort 3 DPP Graduates and All Kindergarteners Enrolled in DPS in Fall 2011*



*This figure includes only Cohort 2 DPP Graduates for whom reading assessment data were available for the 10-11 school year.
^The group of DPS Kindergarteners includes the DPP graduates.

Figure 31: Free/Reduced Lunch Status of Cohort 3 DPP Graduates and Denver Public Schools in Total, Fall 2011



^The group All DPS includes the DPP graduates.

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

In sum, Cohort 3 children with spring 2012 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 and reduced lunch than for the district as whole, suggesting that this sample might be slightly wealthier. The magnitude of the difference between proportions qualifying for free and reduced lunch was much smaller than was observed with Cohort 1, however. 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.

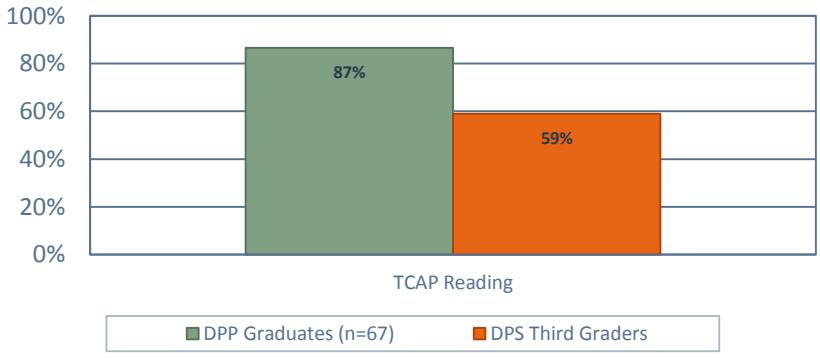
WHAT IS THE OVERALL READING PROFICIENCY OF DPP GRADUATES IN THE EARLY ELEMENTARY YEARS? HOW DOES THIS COMPARE TO THE DISTRICT AS A WHOLE?

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 0, second grade for Cohort 1, first grade for Cohort 2, and kindergarten for Cohort 3). 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 0

Figure 32 displays the proportion of Cohort 0 DPP graduates who scored proficient or advanced on the third grade TCAP reading assessment. All children were assessed in English. 87% of Cohort 0 DPP graduates scored proficient or advanced, compared with 59% of the district as a whole. When examining these results, one should keep in mind the demographic differences between the Cohort 0 DPP graduates and the district as a whole. The differences observed could be due to the DPP program, factors related to those demographic differences, or unmeasured characteristics.

Figure 32: Proportion of Children Scoring Proficient or Advanced on Third Grade TCAP Reading, Cohort 0 DPP Graduates and All Third Graders Enrolled in DPS, Spring 2012

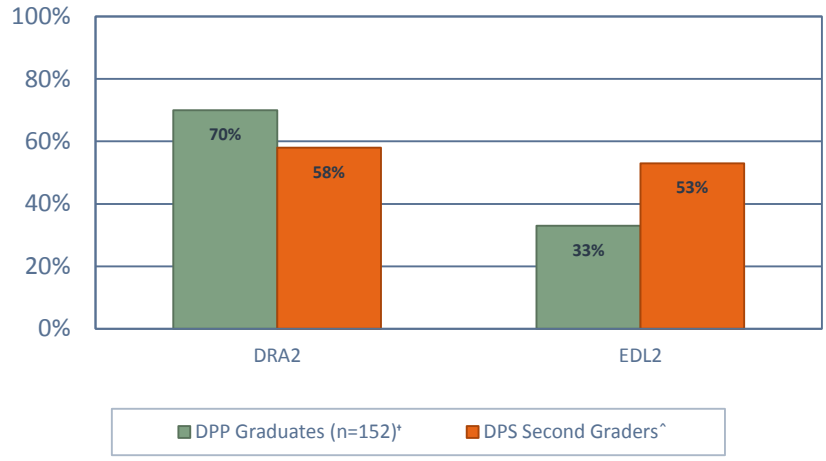


ˆThe group of DPS third graders includes the DPP graduates.

COHORT 1

Figure 33 displays the proportion of Cohort 1 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 2012. Over 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 15 DPP graduates were assessed using the EDL2. Of these 15, only a third were reading on grade level compared to slightly over half of the second graders assessed with EDL2 in the district as a whole.

Figure 33: Proportion of Children Reading On Grade Level in the Spring of the Second Grade Year, Cohort 1 DPP Graduates and All Second Graders Enrolled in DPS, Spring 2012*

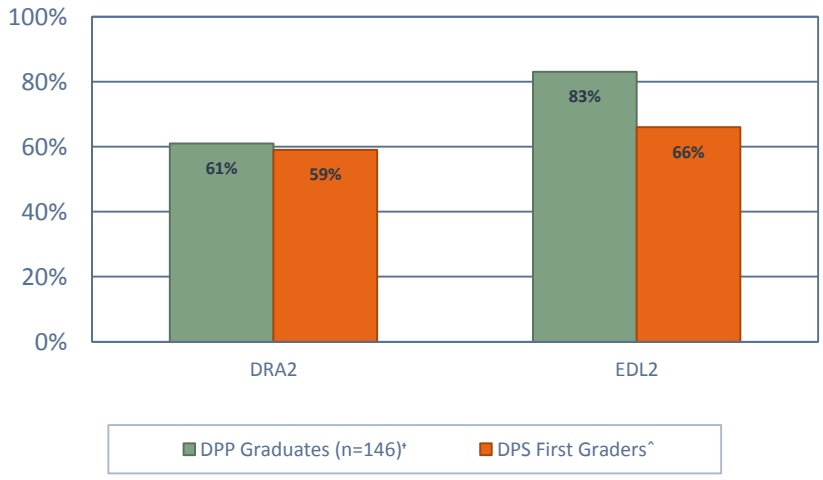


*A score of 28 is considered reading "on grade level" for the end of second grade.
ᵗThe group of DPP Graduates includes 137 children assessed with the DRA2 and 15 assessed with the EDL2.
ˆThe group of DPS second graders includes the DPP graduates.

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 first graders in the district as a whole in spring 2012. Sixty-one percent of DPP graduates assessed in English with the DRA2 were reading at or above grade level at the end of first grade. This is similar to the proportion reading at or above grade level in the district as a whole (59%). Eighty-three percent of DPP graduates assessed in Spanish using the EDL2 were reading at or above grade level at the end of first grade. In contrast, just two-

Figure 34: Proportion of Children Reading On Grade Level in the Spring of First Grade Year, Cohort 2 DPP Graduates and All First Graders Enrolled in DPS, Spring 2012*



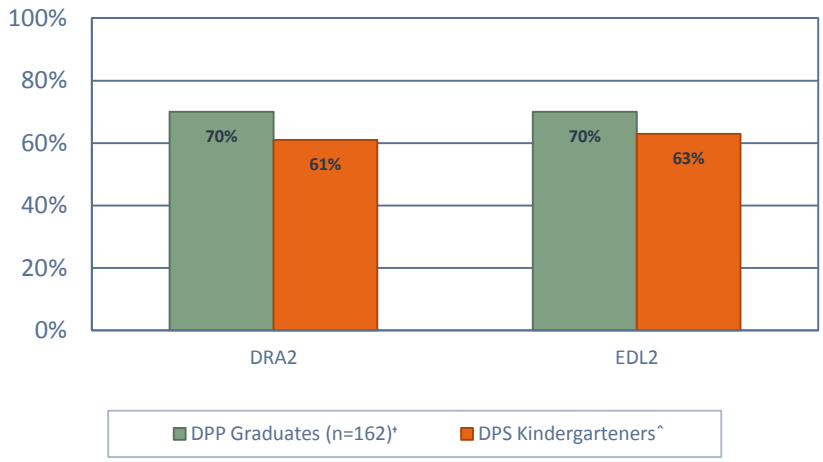
*A score of 16 is considered "on grade level" for the end of first grade.
ᵗThe group of DPP Graduates includes 114 children assessed with the DRA2 and 32 assessed with the EDL2. Analyses are weighted.
ˆThe group of DPS First Graders includes the DPP graduates.

thirds of children in the district as a whole were reading at or above grade level as assessed by the EDL2.

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 kindergarteners in the district as a whole in spring 2012. Seventy percent of DPP graduates assessed in English with the DRA2 were reading at or above grade level at the end of kindergarten. This exceeds the proportion reading at or above grade level in the in the district as a whole (61%). Similarly, 70% of DPP graduates assessed in Spanish using the EDL2 were reading at or above grade level at the end of kindergarten. In contrast, less than 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 Kindergarten Year, Cohort 3 DPP Graduates and All Kindergarteners Enrolled in DPS, Spring 2012*



*A score of 4 is considered "on grade level" for the end of kindergarten.
ᵗThe group of DPP Graduates includes 118 children assessed with the DRA2 and 45 assessed with the EDL2. Analyses are weighted.
ˆThe group of DPS Kindergarteners includes the DPP graduates.

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

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

COHORT 0

As with the previous question, our analysis focused only on the children who were in the expected grades. In spring 2012, 76 Cohort 0 children were enrolled in third grade and had TCAP reading assessment data. Of these, 57 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 students scored proficient or advanced on the TCAP or not. PPVT-4 scores in preschool significantly predicted third grade TCAP proficiency (χ^2=8.44, p<.01). A one point increase in PPVT-4 standard score in preschool is associated with a 13% increase in the odds of scoring proficient or advanced on third grade TCAP. Letter-Word Identification also significantly predicted third grade TCAP proficiency (χ^2=4.65, p<.05). for every one point increase in Letter-Word Identification standard score, there is a 13% increase in the odds of

scoring proficient on the third grade TCAP. Finally, Applied Problems was a significant predictor of third grade reading proficiency (χ^2=4.21, p<.05). For every one point increase in Applied Problems standard score, there was a 14% increase in the odds of scoring proficient on the third grade TCAP.

Ten children with third grade reading assessment data in spring 2011 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. This sample size is too small to permit us to examine consistency over time for this subgroup of children.

In sum, for predominantly English-speaking children, there was evidence for a strong association between kindergarten readiness and third grade reading proficiency. The subgroup of children identified as Spanish speakers in this cohort was not large enough to examine.

COHORT 1

In spring 2012, 149 children were enrolled in second grade and had reading assessment data. Ninety-one 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 .52 (p<.0001). Letter-Word Identification was correlated with DRA2 at .52 (p<.0001). Applied Problems was associated with DRA2 scores at .40 (p<.001). The magnitude of the associations of DRA2 with PPVT-4 and Letter-Word Identification was similar to what was observed in first grade.[91] The association between second grade DRA2 and Applied Problems was somewhat smaller than what was observed in first grade.

Fifty-eight Cohort 1 children with second grade reading assessment data in spring 2012 had been assessed in both English and Spanish during their preschool year. Forty-four of these children were assessed in English using the DRA2 at the end of second grade. Fourteen 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 all three of the English preschool assessments. PPVT-4 was correlated .38 with DRA2 (p<.05), Letter-Word Identification was correlated .35 with DRA2 (p<.05) and Applied Problems was correlated .54 with DRA2 (p<.001). DRA2 scores in second grade were not significantly associated with TVIP or Spanish Letter-Word Identification, but they were significantly associated with the Spanish Applied Problems assessment (r=.31, p<.05). This pattern of associations differs markedly from the pattern of results observed for this cohort last year, when they were in first grade. Last year, there were no significant associations between DRA2 and the English assessments.

For children assessed in Spanish in second grade, there were no significant associations between the EDL2 scores and the preschool assessments administered in English. This differs from last year, when the children were

[91] For more information about the analyses of Cohort 1 for the first grade year, readers are referred to a previously prepared report. Klute, M. M. (2011). *Denver Preschool Program: Report on Child Outcomes, 2010-11 School Year.* Unpublished Report. Denver: Clayton Early Learning Institute.

in first grade. Last year, there was a strong pattern of associations between the English preschool assessments and EDL2 scores. For Spanish preschool assessments, Applied Problems was correlated with EDL2 at .67 ($p<.01$). There was also a trend toward a significant association between EDL2 scores and TVIP ($r=.52$, $p<.10$). This is somewhat similar to last year, but the pattern of effects is weaker.

In sum, for Cohort 1, 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 was a strong pattern of associations between kindergarten readiness assessed in English and children's reading scores at the end of second grade. In addition, their preschool math abilities, and to some extent, their preschool literacy skills assessed in Spanish were associated with second grade reading assessed in English. For children who were assessed in Spanish at the end of the second grade year, there was a weak pattern of evidence for an association between kindergarten readiness assessed in Spanish and reading skills assessed in Spanish at the end of the second grade year. It is noteworthy that, for developing bilinguals, the pattern of effects is second grade differs markedly from the pattern of effects observed at the end of first grade.

COHORT 2

In spring 2012, 141 children were enrolled in first grade and had reading assessment data. Sixty percent of these children were assessed only in English in preschool and were assessed with the DRA2 in the first grade year. Associations between the preschool assessments and first grade DRA2 scores were strong and significant. PPVT-4 scores in preschool were correlated with first grade DRA2 scores at .53 ($p<.0001$). Letter-Word Identification was correlated with DRA2 at .66 ($p<.0001$). Applied Problems was associated with DRA2 scores at .52 ($p<.001$). The magnitude of these associations was similar to associations observed for this cohort last year when the children were in kindergarten.

Forty percent of Cohort 2 children with first grade reading assessment data in spring 2012 had been assessed in both English and Spanish during their preschool year. Forty-three percent of these children were assessed in English using the DRA2 at the end of first grade. Fifty-seven percent 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 English vocabulary and literacy assessments. DRA2 was correlated with PPVT-4 and English Letter-Word Identification at .57 and .69, respectively (both $p<.01$). The association with preschool math skills assessed in English was non-significant as were all of the correlations between the DRA2 and Spanish preschool assessments. This pattern of results differs from what was observed last year, when these children were in kindergarten. Last year, we observed no significant associations between kindergarten DRA2 and preschool assessments in English. Similar to last year, there were no significant associations between preschool assessments in either language and the Spanish language EDL2 at the end of first grade.

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 is strongly associated with reading assessment data at the end of kindergarten. For developing Spanish-English bilingual children, there was some evidence of associations between first grade language and literacy skills and reading skill

at the end of first grade for children assessed in English. There were no associations for developing bilingual children assessed in Spanish in first grade.

COHORT 3

In spring 2012, 162 children were enrolled in kindergarten and had reading assessment data. Fifty-two percent 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 .46 ($p<.0001$). Letter-Word Identification was correlated with DRA2 at .55 ($p<.0001$). Applied Problems was associated with DRA2 scores at .33 ($p<.01$). The magnitude of these associations was somewhat weaker than was observed last year when Cohort 2 was in kindergarten.

Forty-eight percent of Cohort 3 children with kindergarten reading assessment data in spring 2012 had been assessed in both English and Spanish during their preschool year. Forty-three percent of these children were assessed in English using the DRA2 at the end of kindergarten. Fifty-seven percent 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 a strong pattern of association between the DRA2 and preschool assessments in both English and Spanish. For preschool assessments in English, DRA2 was correlated with PPVT-4 at .66 ($p<.001$), Letter-Word Identification at .67 ($p<.001$), and Applied Problems at .56 ($p<.01$). For preschool assessments in Spanish, DRA2 was correlated with TVIP at .51 ($p<.01$), Letter-Word Identification at .53 ($p<.01$) and Applied Problems at .79 ($p<.0001$). For children assessed in Spanish in kindergarten there were no significant associations.

In sum, for Cohort 3, there was a strong pattern of associations for children whose primary language was English. For these children, there is strong evidence that kindergarten readiness at the end of preschool is strongly associated with reading assessment data at the end of kindergarten. For developing Spanish-English bilingual children, there was a strong pattern of associations between kindergarten readiness and reading skill at the end of kindergarten for children whose kindergarten reading skill was assessed in English, but not for those who were assessed in Spanish.

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

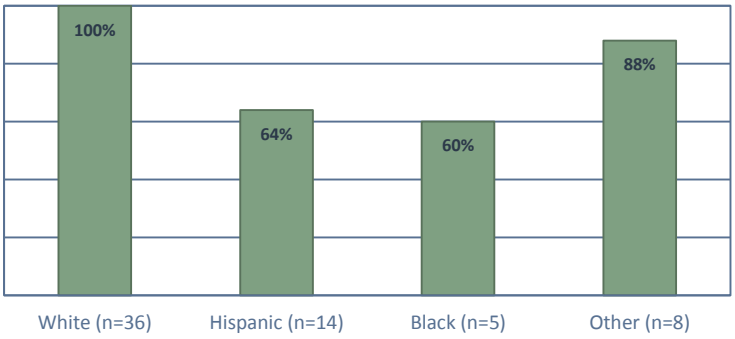
To address this question, we examined the associations between demographic characteristics and reading scores for each cohort. For Cohort 0, we examined child gender, race/ethnicity, home language and child primary language. We were unable to examine family income because income information for Cohort 0 was unreliable. For Cohorts 1, 2, and 3 we examined child gender, race/ethnicity, home language, child primary language, income tier and region of the city.

COHORT 0

TCAP

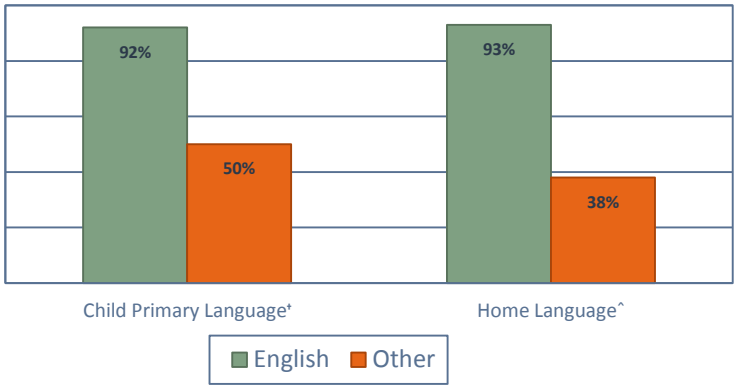
In spring 2012, among Cohort 0 DPP graduates, TCAP proficiency differed as a function of racial/ethnic group (see Figure 36).[92] White children and those in the 'other' race/ethnicity category were more likely to score proficient or advanced on the 3rd grade TCAP than Hispanic and black children. It is noteworthy that all of the white children scored proficient or advanced.

Figure 36: Percent of Children Scoring Proficient or Advanced on 3rd Grade Reading TCAP by Race/Ethncity, Cohort 0, Spring 2012



When we examined child primary language and home language, we found that both of these variables were significantly associated with third grade reading proficiency (see Figure 37).[93] Children who spoke English as their primary language were nearly twice as likely to score proficient or advanced on the TCAP reading assessment, which is administered in English. Similarly, children whose home language is English were nearly 2.5 times more likely to score proficient or advanced on the third grade TCAP reading. These language differences may, in part, explain the racial/ethnic differences described above, as all of the children with primary or home languages other than English were Hispanic.

Figure 37: % of Children Scoring Proficient or Advanced on Third Grade TCAP Reading by Child Primary Language and Home Language, Cohort 0, Spring 2012



*59 children had English identified as their primary language, 8 had another language.
^59 children had English identified as the home language, 8 had another language.

[92] χ^2_3=15.29, p<.01

[93] Child Primary Language: χ^2_1=10.45, p<.001; Home Language: χ^2_1=18.81, p<.0001

COHORT 1

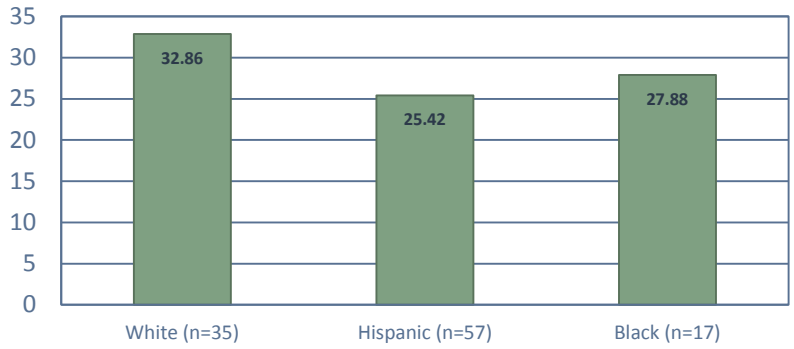
ENGLISH READING ASSESSMENT (DRA2)

In spring 2012, there was not a significant difference between boys and girls on the DRA2.[94] There was, however, a difference in DRA2 scores by race/ethnicity (see Figure 38).[95] Follow-up Tukey tests revealed that white children, whose scores, on average, exceeded grade level expectations, scored significantly higher than Hispanic children, whose average score fell short of grade level expectations.[96] Black children's scores, on average, fell just slightly below grade level expectations.

DRA2 scores also differed by children's primary language and home language (see Figure 39).[97] Children whose primary language was English scored significantly higher than children with another primary language. A similar pattern was observed for home language.

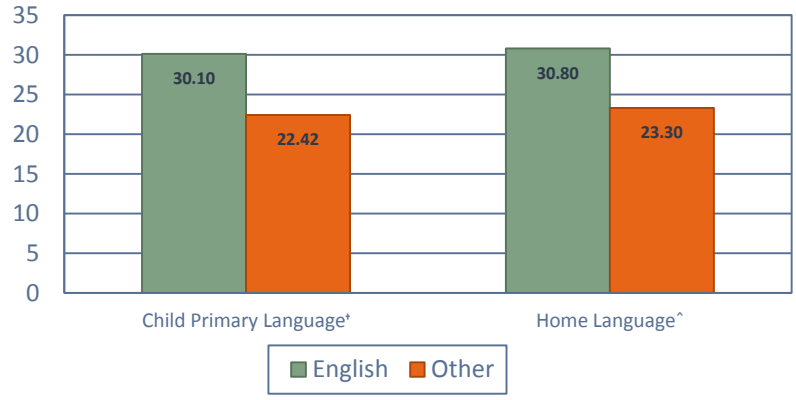
There was also a significant difference by income tier (see Figure 40).[98] As income tier increased, so did average reading levels. Follow-up Tukey

Figure 38: Second Grade Reading Level (DRA2) by Race/Ethnicity, Cohort 1, Spring 2012*



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

Figure 39: Second Grade Reading Level (DRA2) by Child Primary Language and Home Language, Cohort 1, Spring 2012*



*A score of 28 is considered reading "on grade level" for the end of second grade.
†103 children had English identified as their primary language, 31 had another language.
^83 children had English identified as the home language, 30 had another language.

[94] t=0.74, df=133, n.s.

[95] This analysis omitted 2 Asian children, one Native American child, and one child classified as multi-racial. Even combining these groups would have resulted in a group too small for analysis. F(2,108)=10.06, p<.0001

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

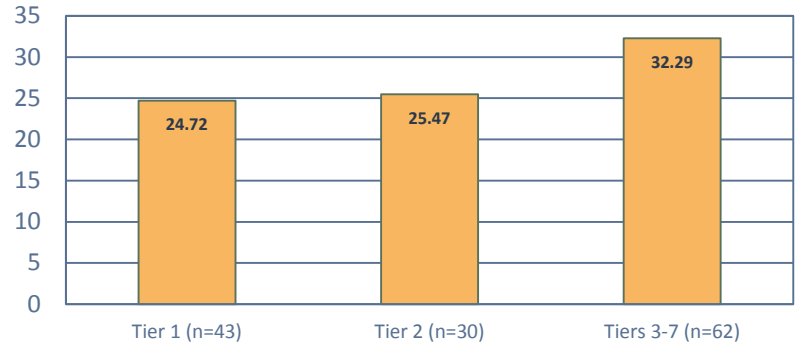
[97] Child Primary Language: t=5.00, df=132, p<.0001; Home Language: t=4.70, df=111, p<.0001.

[98] 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(2,134)=16.58, p<.0001.

tests revealed that the group of children in Tiers 3-7 had significantly higher reading levels than children in both of the other income tier groups. The difference between Tier 1 and Tier 2 was not statistically significant.

It is not suprising that there were effects for both the language variables and income tier. As discussed at length in our annual evaluation report on data from 08-09 school year,[99] income and language were closely related in this sample at the time when children enrolled in DPP. In our full Cohort 1 sample, over 90% of children whose primary language was something other than English were from Income Tiers 1 and 2. Less than half of the children whose primary language was English were from these lowest two tiers. As a result, in this sample, it is impossible to disentangle the effects of income and language. The effects for each of these variables just described, are possibly due to the co-occurence of these two factors.

*Figure 40: Second Grade Reading Level (DRA2) by Income Tier, Cohort 1, Spring 2012**



*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.[100]

SPANISH READING ASSESSMENT (EDL2)

As with the English reading assessment, there was not a significant difference in EDL2 scores by child gender.[101] 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.[102]

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

[100] $F(4,133)=1.87$, n.s.

[101] $t=0.42$, df=12, n.s.

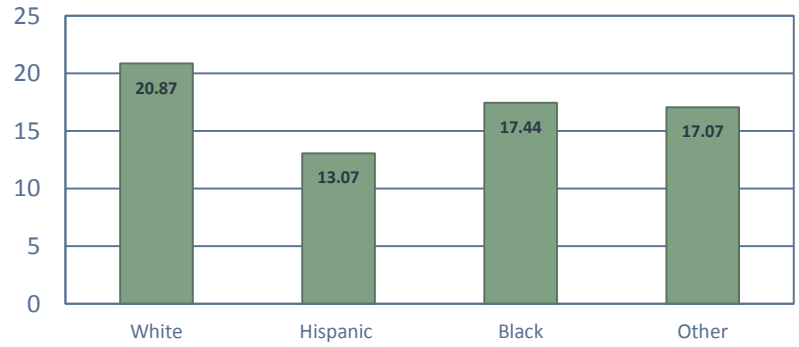
[102] Income tier: $F(2,13)=0.95$, n.s.; region of the city: $F(3,13)=0.75$, n.s.

COHORT 2

ENGLISH READING ASSESSMENT (DRA2)

In spring 2012, there was not a significant difference between boys and girls on the DRA2.[103] There was, however, a difference in DRA2 scores by race/ethnicity (see Figure 41).[104] Follow-up Tukey tests revealed a significant difference between white students, who were, on average, exceeding grade-level expectations for reading and Hispanic children, who were, on average, reading below grade level.[105] Black students and students from other racial/ethnic groups were also reading slightly above grade level, on average.

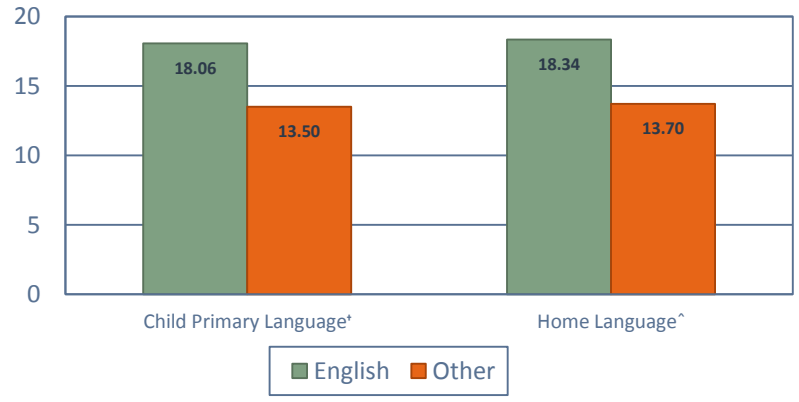
Figure 41: First Grade Reading Level (DRA2) by Race/Ethnicity, Cohort 2, Spring 2012*



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

DRA2 scores differed by children's primary language and home language (see Figure 42).[106] Children who spoke English as their primary language scored signficantly higher than children with another primary language. A similar pattern was observed for home language. On average, children who spoke English as their primary language or had English as their home language exceeded grade level expectations for reading. Children who had other primary or home languages tended to be reading below grade level expectations, on average.

Figure 42: First Grade Reading Level (DRA2) by Child Primary Language and Home Language, Cohort 2, Spring 2012*



*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.

[103] t=0.29, df=105, n.s.

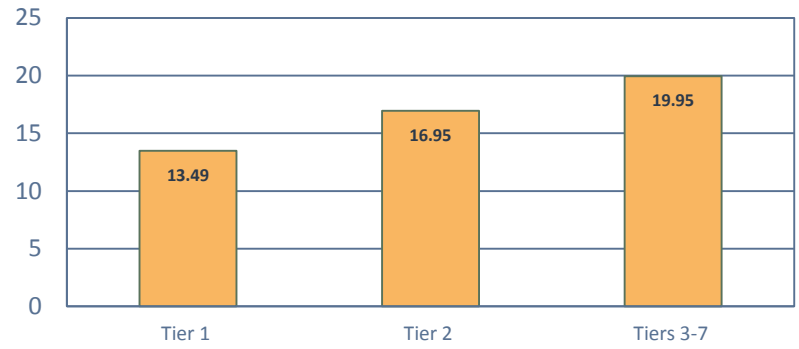
[104] F(3,106)=7.54, p<.001

[105] A DRA2 score of 16 is considered "on grade level" for the end of first grade.

[106] Child Primary Language: t=2.44, df=100, p<.05; Home Language: t=3.37, df=49.29, p<.01

There was also a significant difference by income tier (see Figure 43).[107] As income tier increased, so did average reading levels. Follow-up Tukey tests revealed that the group of children in Tiers 3-7 had significantly higher reading levels than children in Tier 1. The other pairwise comparisons were not statistically significant.

*Figure 43: First Grade Reading Level (DRA2) by Income Tier, Cohort 2, Spring 2012**



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

As with Cohort 1, it is not suprising that there were effects for both home language and income tier, as these variables are strongly associated.[108] In our full Cohort 2 sample, over 80% of children whose primary language was something other than English were from Income Tiers 1 and 2. About half of the children whose primary language was English were from these lowest two tiers. As a result, in this sample, it is impossible to disentangle the effects of income and language. The effects for each of these variables just described, are possibly due to the co-occurence of these two factors.

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.[109]

SPANISH READING ASSESSMENT (EDL2)

As with the English reading assessment, there was not a significant difference in EDL2 scores by child gender.[110] 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.[111]

[107] 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. $F(2,104)=8.53$, $p<.001$

[108] For a more detailed discussion, readers are referred to: Klute, M. M. (2009). *Denver Preschool Program: Report on Child Outcomes, 2009-10 School Year.* Unpublished Report, November. Denver: Clayton Early Learning Institute.

[109] $F(4,106)=1.47$, n.s.

[110] $t=0.88$, $df=22$, n.s.

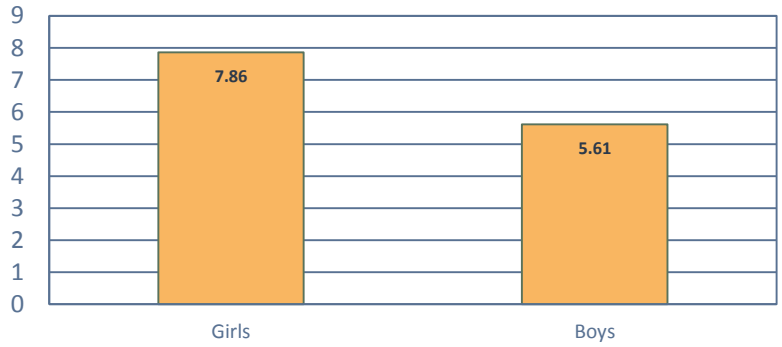
[111] Income tier: $F(2,23)=2.64$, n.s.; region of the city: $F(4,23)=0.06$, n.s.

COHORT 3

ENGLISH READING ASSESSMENT (DRA2)

In contrast to previous cohorts, for cohort 3 there was not a difference in DRA2 scores by race/ethnicity in spring 2012.[112] There was, however, a significant difference between boys and girls on the DRA2 (see Figure 44).[113] On average, both groups were exceeding grade-level expectations for reading, but girls scored significantly higher than boys. There was not a significant

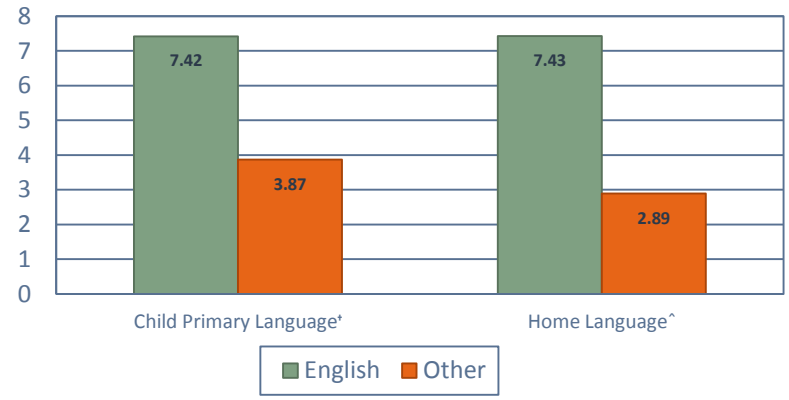
Figure 44: Kindergarten Reading Level (DRA2) by Child Gender, Cohort 3, Spring 2012*



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

DRA2 scores differed by both children's primary language and home language (see Figure 45).[114] Children whose primary language was English far exceeded the grade level expectations for reading and they scored significantly higher than children with another primary language, who on average, approached grade level expectations. A similar, but more pronounced pattern was observed for home language.

Figure 45: Kindergarten Reading Level (DRA2) by Child Primary Language and Home Language, Cohort 3, Spring 2012*



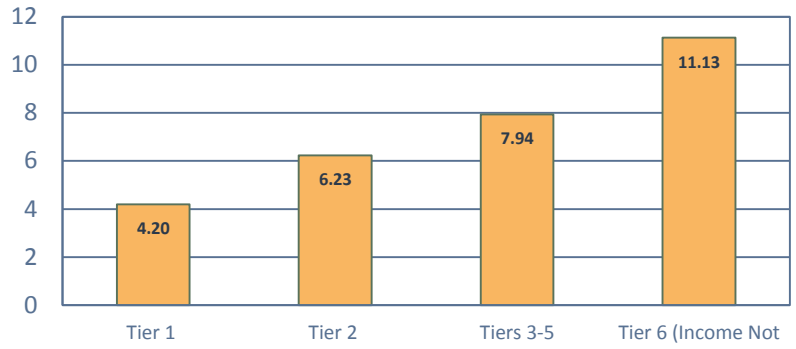
*A score of 4 is considered "on grade level" for the end of kindergarten.
'86 children had English identified as their primary language, 18 had another language.
^81 children had English identified as the home language, 17 had another language.

There was also a significant difference by income tier (see Figure 46).[115] 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 and the group of children in Tier 6 (Income Not Provided) had significantly higher reading levels than children in Tier 1. The other pairwise comparisons were not statistically significant.

[112] $F(3,112)=1.18$, n.s.
[113] $t=2.28$, df=82.80, $p<.05$
[114] Child Primary Language: $t=3.08$, df=102, $p<.05$; Home Language: $t=6.63$, df=91.90, $p<.0001$
[115] 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). $F(3,112)=6.14$, $p<.001$

As with previous cohorts, it is not suprising that there were effects for both home language and income tier, as these variables are strongly associated.[116] In our full Cohort 3 sample, over 90% of children whose primary language was something other than English were from Income Tiers 1 and 2. About 40% of the children whose primary language was English were from these lowest two tiers. As a result, in this sample, it is impossible to disentangle the effects of income and language. The effects for each of these variables just described, are possibly due to the co-occurence of these two factors.

*Figure 46: Kindergarten Reading Level (DRA2) by Income Tier, Cohort 3, Spring 2012**



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

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.[117]

SPANISH READING ASSESSMENT (EDL2)

There was not a significant difference in EDL2 scores by child gender.[118] 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.[119]

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

COHORT 0

We compared Cohort 0 children who had been enrolled in community preschools (n=39) with children who had been enrolled in DPS preschools (n=28) on third grade TCAP proficiency in reading administered during the spring of 2011. There was not a significant difference between these two groups.[120]

[116] For a more detailed discussion, readers are referred to: 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.
[117] $F(4,112)=0.85$, n.s.
[118] $t=0.71$, df=27, n.s.
[119] Income tier: $F(1,28)=0.02$, n.s.; region of the city: $F(3,28)=1.22$, n.s.
[120] $\chi^2_1=.81$, n.s.

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. 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. Nonetheless, we attempted the analysis to compare children who had been enrolled in community sites (n=17) to those who had been enrolled in DPS sites (n=106). The difference in means for these two groups was not statistically significant.[121]

Fourteen Cohort 1 DPP graduates were assessed in Spanish with the EDL2 at the end of second grade. Unexpectedly, every one of these children had been enrolled in a DPS preschool. As a result it was not possible to test for mean differences by provider type for Cohort 1 children assessed in Spanish.

COHORT 2

We compared Cohort 2 children who had been enrolled in community preschools (n=51) with children who had been enrolled in DPS preschools (n=56) on first grade reading assessments in English (DRA2) administered during the spring of 2012. There was not a significant difference between these two groups.[122] There was a significant difference by provider type in EDL2 scores.[123] The five children assessed with the EDL2 in first 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=19).[124]

COHORT 3

We compared Cohort 3 children who had been enrolled in community preschools (n=54) with children who had been enrolled in DPS preschools (n=59) on kindergarten reading assessments in English (DRA2) administered during the spring of 2012. There was not a significant difference between these two groups.[125] 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).

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

We were only able to examine this question for Cohorts 1, 2 and 3, as we did not gather program quality data for Cohort 0 because it was a pilot year. As described in the annual evaluation reports for 08-09, 09-10 and 10-11,[126] there was very limited variability in the star rating of the preschools attended by DPP children in these

[121] t=0.13, df=121, n.s.
[122] t=0.21, df=105, n.s.
[123] t=2.26, df=22, p<05
[124] Community preschools: mean=12.4, sd=6.07; DPS preschools: mean=17.68, sd=4.28; 16 is considered "on grade level" for first grade.
[125] t=1.02, df=111, n.s.
[126] 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.

cohorts. Very few children were enrolled in programs with less than a star 3 rating (4% of the Cohort 1 sample, 7% of the Cohort 2 sample, 9% of Cohort 3 sample). 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 Cohort 3 only, we examine the CLASS in lieu of the Qualistar rating.

COHORT 1

When we examined the distribution of star rating for children who had second grade DRA2 data in spring 2012, we found that, similar to the preschool year, only 4% had been enrolled in preschools with less than a star 3 rating. Sixty-nine percent of children had been enrolled in star 3 preschools and 26% had been enrolled in star 4 preschools. There was no difference between these groups on their second grade DRA2 scores.[127] When we examined the distribution of star rating for children with second grade EDL2 data, we found that 64% of them had been enrolled in star 3 preschools. 29% of children were enrolled in star 4 preschools. Just one child had been enrolled in a preschool with less than a star 3 rating, making it impossible to include this rating level in the analysis. The analysis of second grade EDL2 scores comparing children who had been enrolled in star 3 preschools to those who had been enrolled in star 4 preschools was also non-significant.[128]

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 and EDL2 scores. Only one of these six correlations was significant. Training and education points was correlated with DRA2 scores at the end of second grade at .25 ($p<.01$).

COHORT 2

When we examined the distribution of star rating for children who had first grade DRA2 data in spring 2012, we found that, similar to the preschool year, only 10% had been enrolled in preschools with less than a star 3 rating. Sixty-one percent of children had been enrolled in star 3 preschools and 29% 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 start rating, 82% 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.[129]

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.

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.

[127] $F(2, 133)=1.46$, n.s.

[128] $F(1, 12)=0.60$, n.s.

[129] $F(2,96)=2.13$, n.s.

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. Neither of the correlations with the DRA2 was significant. For children assessed with the EDL2, both correlations were significant and in an unexpected direction. Classroom Organization was correlated with EDL2 score at -.43 ($\underline{p}<.05$). Instructional Support was correlated with EDL2 score at -.47 ($\underline{p}<.05$). When interpreting these correlations, one should keep in mind that there were just 25 children included in this analysis.

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?

4. Do children who received DPP tuition credits compare favorably with their demographic counterparts who did not receive DPP tuition credits 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)?

As described above, only the first evaluation cohort, which was not representative of DPP participants,has taken the TCAP. Because of this we are limited in our ability to address question 5 this year.

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

Children did make significant progress in their academic and socio-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 and literacy skills. Spanish-speaking children also made progress in their 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 socio-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 socio-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: 57% 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 6% of children. Parents identified protective factors as an area of concern for about 12% of children and behavioral concerns as an area of concern for nearly half of children. The DECA, the socio-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 fell slightly below that threshold. Parents identified behavioral concerns as an area of concern for about many more children that 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 parents. 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 that children from lower income tiers (defined by income adjusted for family size) started lower and made larger gains in two of the three academic assessments in English, but there was no association for the assessments in Spanish. Children from Tier 1 tended to increase more rapidly than children in Tiers 3-6. Finally, with respect to teachers' ratings of Protective Factors, children in Tier 2 increased significantly more than children in Tiers 3-5.

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 two of the three academic assessments. There were no significant associations with the socioemotional variables.

QUESTION 4: DO CHILDREN WHO RECEIVED DPP TUITION CREDITS COMPARE FAVORABLY WITH THEIR DEMOGRAPHIC COUNTERPARTS WHO DID NOT RECEIVE DPP TUITION CREDITS ON ASSESSMENTS ADMINISTERED BY DENVER PUBLIC SCHOOLS IN KINDERGARTEN AND BEYOND?

Cohorts 1, 2, and 3 were demographically similar to the populations of children in 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 Cohort 1 than Cohort 2 and 3.

Cohort 1 children were compared to the population of second graders in DPP. Among children whose reading was assessed in English in second grade, the proportion of children in Cohort 1 who were reading at or above grade level exceeded the proportion in the district as a whole. Among children assessed in Spanish, the proportion of DPP graduates reading at or above grade level was smaller than the district as a whole, but this group was rather small in size.

A different pattern was observed for the group of DPP graduates who were enrolled in first grade during the 11-12 school year. 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 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 kindergarteners in DPP. 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. The same pattern was observed for children whose reading was assessed in Spanish at the end of the kindergarten year.

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

In the preschool year, there was not strong evidence for an association between classroom quality and children's kindergarten readiness in the academic domains or their socioemotional development.

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 first or second grade. For Cohort 3, we examined the association between CLASS observation scores and reading skill in kindergarten. For children assessed in English, there was no association. For children assessed in Spanish, there were some significant associations in an unexpected direction.

SUMMARY AND FUTURE DIRECTIONS

This evaluation described children's progress during the course of their DPP preschool year. In general, children progressed in their language and literacy skills as assessed in English and literacy and math skills in Spanish at a rate which exceeded what would be expected simply because of maturation. Children demonstrated positive changes in their socioemotional 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 the preschool year or with reading scores in kindergarten.

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. There was some evidence that children from higher-risk groups (living in or near poverty, speaking a language other than English primarily) made progress toward closing the achievement gap that was present at the beginning of the preschool year. 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 small group of children assessed in Spanish in second 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 2012

	Entire Sample, weighted[1]	By Provider Type, Unweighted		
Characteristic		Community	DPS	Significance of Difference by Provider Type
Sex				χ^2_1=.72; n.s.
Female	45.7%	51.0%	45.0%	
Male	54.3%	49.0%	55.0%	
Ethnicity				χ^2_4=23.69; p<.0001
Hispanic	52.4%	26.0%	56.0%	
White (not of Hispanic origin)	28.4%	54.0%	25.0%	
African-American (not of Hispanic origin)	10.2%	12.0%	10.0%	
Multi-Racial	5.1%	6.0%	5.0%	
Asian/Pacific Islander	3.6%	1.0%	4.0%	
Child's Primary Language				χ^2_1=24.66; p<.0001
English	53.2%	84.0%	49.0%	
Another Language	35.7%	11.0%	39.0%	
Not Reported	11.2%	5.0%	12.0%	
Home Language				χ^2_1=27.86; p<.0001
English	53.7%	81.0%	50.0%	
Another Language	35.7%	11.0%	39.0%	
Not Reported	11.2%	9.0%	7.0%	
DPP Income Tier[2]				χ^2_5=30.96; p<.0001
Tier 1	45.3%	18.0%	49.0%	
Tier 2	24.1%	25.0%	24.0%	
Tier 3	0.6%	5.0%	0.0%	
Tier 4	8.0%	8.0%	8.0%	
Tier 5	15.9%	37.0%	13.0%	
Tier 6—Income Not Reported	6.1%	7.0%	6.0%	
Star Level of Preschool				χ^2_2=17.67; p<.0001
Star 2	3.0%	10.0%	2.0%	
Star 3	61.7%	37.0%	65.0%	
Star 4	35.4%	53.0%	33.0%	
Region of the City				χ^2_4=7.14; n.s.
Central	13.8%	20.0%	13.0%	
Northeast	31.1%	24.0%	32.0%	
Northwest	17.2%	26.0%	16.0%	
Southeast	10.1%	11.0%	10.0%	
Southwest	27.8%	18.0%	29.0%	

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

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

Table A2: DPP Income Tiers

Income Tier 1	
Household Size	*Annual Income Equal to or Less Than*
2	$14,570
3	$18,310
4	$22,050
5	$25,790
6	$29,530
7	$33,270
8	$37,010
If more than 8 family members	Add $3,740 for each additional family member

Income Tier 2		
Household Size	*Annual Income*	
	More Than	*Equal to or Less Than*
2	$14,571	$26,955
3	$18,311	$33,874
4	$22,051	$40,793
5	$25,791	$47,712
6	$29,531	$54,631
7	$33,271	$61,550
8	$37,011	$68,469
If more than 8 family members	Add $6,919 for each additional family member	

Income Tier 3		
Household Size	*Annual Income*	
	More Than	*Equal to or Less Than*
2	$26,956	$32,783
3	$33,875	$41,198
4	$40,794	$49,613
5	$47,713	$58,028
6	$54,632	$66,443
7	$61,551	$74,858
8	$68,470	$83,273
If more than 8 family members	Add $8,415 for each additional family member	

Income Tier 4		
Household Size	*Annual Income*	
	More Than	*Equal to or Less Than*
2	$32,784	$43,710
3	$41,199	$54,930
4	$49,614	$66,150
5	$58,029	$77,370
6	$66,444	$88,590
7	$74,859	$99,810
8	$83,274	$111,030
If more than 8 family members	Add $8,976 for each additional family member	

Income Tier 5	
Household Size	*Annual Income*
	More Than
2	$43,710
3	$54,930
4	$66,150
5	$77,370
6	$88,590
7	$99,810
8	$111,030
If more than 8 family members	Add $11,220 for each additional family member