Denver Preschool Program: Report on Child Outcomes

2010-11 School Year

Prepared for the Denver Preschool Program
By Mary Maguire Klute, Ph.D. and Caroline Ponce
Clayton Early Learning Institute

November 3, 2011

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 coaching and access to 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 Spanish vocabulary and literacy skills 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.

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 fewer than 10% 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 2010-11 school year, nearly all children whose primary language was not English were from the lowest two income tiers as compared with about 40% 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 fairly 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.

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

We followed two groups of DPP graduates who were enrolled in kindergarten and first grade during the 2010-11 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 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 similar to the district as a whole.

Among kindergarteners whose reading was assessed in English, the proportion of DPP graduates who were reading at or above grade level slightly exceeded the proportion of children in the district as a whole who were reading at or above grade level. Among kindergarteners assessed in Spanish, DPP graduates were about 1.5 times more likely to be reading at grade level than the district as a whole.

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 kindergarten and first grade reading skills for these cohorts of children. We did not find any significant 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 quality and child outcomes. One of the domains, Instructional Support, was associated with socioemotional development but there were not

strong associations with academic outcomes. Next year, when this cohort of children is in kindergarten, we will be able to examine this measure of preschool quality in conjunction with reading scores.

SUMMARY AND FUTURE DIRECTIONS

This evaluation described children's progress during the course of their DPP preschool year. 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 greater reading proficiency in kindergarten and first grade than the district as a whole. 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 coaching and access to 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[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 CSAP)?

Since 2010-11 was just the fourth year of the DPP program, we were limited in our ability to address questions 4 and 5. Children who participated in DPP during its first year of operation are expected to be in second grade during the 2010-11 school year. As such, DPP graduates will start taking the CSAP during the 11-12 school year. In addition, while it is the fourth year of DPP's operation, it is important to note that 2010-11 was just the third year of full implementation of the evaluation design. As a result, we will be able to start addressing question 5 after the 12-13 school year, when the first full cohort of DPP children that were studied start to take CSAPs.

[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).

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 will henceforward be 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 is assigned a DPS ID. DPS IDs are 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 Cohorts 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 second grade during the 10-11 school year (see Table 1). We obtained spring 2011 reading assessment scores from DPS for 82 of these children (69% of the total sample, 72% of those for whom we had obtained DPS IDs). Of these, 76 (93%) were enrolled in second grade as expected. The remaining 6 children were enrolled in first grade.

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 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).

---

[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.

[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 1 children were expected to be in the first grade during the 10-11 school year (see Table 1). We obtained spring reading assessment data for 171 children (83% of the whole sample; 86% of those for whom we had obtained DPS IDs). Of these, 96% were in first grade as expected. Three children were in second grade, having apparently skipped a grade. None of these children had reading assessment data for spring 2010, so it is unknown which grade they skipped. Three children were in kindergarten. None of these children had reading assessment data for spring 2010, 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, we applied sampling weights 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 kindergarten during the 10-11 school year (see Table 1). We obtained reading data for 139 children (69% of the sample; 75% of the sample when sampling weights were applied). One child was in first grade; all other children were in kindergarten.

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

	School Year			
	07-08	08-09	09-10	10-11
Cohort 0	Pre-K	Kindergarten	1st Grade	2nd Grade
Cohort 1		Pre-K	Kindergarten	1st Grade
Cohort 2			Pre-K	Kindergarten
Cohort 3				Pre-K

COHORT 3

SAMPLING PLAN

As explained above for Cohort 2, we stratified our sample for Cohort 3 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, we applied sampling weights 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 2010, a sample of 100 children enrolled in community sites was

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

drawn from the group of families that volunteered to participate (henceforth referred to as "volunteers"). In September 2010, 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, we compared those who volunteered to be contacted regarding the evaluation to those who refused to be contacted (henceforth referred to as "non-volunteers") 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. We also examined DPP income tier, which takes into account both family size and income. 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. When testing whether there were significant differences between volunteers and non-volunteers, we also considered whether the groups differed on whether they were missing data on income.

COMMUNITY SITES

In community sites, no significant differences were observed between the 356 families that volunteered to be contacted about the study and the 216 families that refused to be contacted regarding the study. Both volunteers and non-volunteers were similar in terms of sex of the child, ethnicity, DPP income tier, Qualistar rating of the preschool program, home language, child language and region of the city in which the child lives. Since there were no significant differences, no additional stratification of the sample was necessary. A sample of 100 children was drawn from the group of volunteers. This sample is representative of the community site population as a whole in August 2010 with respect to the variables examined.[5] The sample was drawn with replacement; if a selected child was deemed ineligible for the study,[6] a selected family was unable to be contacted to obtain informed consent to participate in the study or if a selected family refused to participate in the study, a replacement child was randomly drawn.

DPS SITES

In DPS sites, significant differences were detected between the 1359 volunteers and 1146 non-volunteers on two variables. First, a significant difference was detected for income tier level.[7] 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 (63% of families in tier 5 volunteered vs. 53% of families in the other income tiers). A significant difference was also detected for region of the city where the child lived.[8] Follow-up analyses revealed that this effect was due to differences between volunteers and non-volunteers in two regions of the city. Parents of children who lived in the northeast region of the city were significantly more likely to volunteer to be contacted about the evaluation than parents of children from other regions of the city (59% of parents of children from the northeast region volunteered vs. 53% of parents of children living in other regions of the city). In contrast, parents of children living in the southwest region of the city were

[5] It is possible that volunteers and non-volunteers differed on some unmeasured characteristics and that the sample may differ from the population as a whole on these characteristics.

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

[7] $\chi^2_5=15.10$, $p<.01$

[8] $\chi^2_4=15.95$, $p<.01$

significantly less likely to volunteer than parents of children living in other areas of the city (49% of parents of children from the southwest region volunteered vs. 57% of parents of children living in other regions of the city). To adjust for these differences, the sampling frame was stratified by income tier (tier 5 vs. other) and region of the city (northeast, southwest, 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 2010 with respect to the variables examined. As with the sample from community sites, the sample was drawn with replacement; if a selected child was deemed ineligible for the study, a selected family was unable to be contacted to obtain informed consent to participate in the study or if a selected family refused to participate in the study, a replacement child was randomly drawn from the same stratum.

SAMPLING WEIGHTS

At the time of sampling, 19.6% of children enrolled in DPP were attending community sites and the remaining 81.4% were attending DPS sites.[9] We drew a sample that included 50% children from community sites and 50% of children from DPS sites. As a result, our sampling design involved oversampling children from community sites. When we analyzed 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.[10] About half of the sample was female. Hispanics represented about half of the sample; the next most common ethnic group was whites. African-Americans made up about 8 percent 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, nearly 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 2007, which is also the cutoff for Free and Reduced Lunch. The next most common income tier was tier 5. About seven percent of families were assigned to the highest tier, tier 6, because they opted out of the requirement to provide their income.

Nearly all (91%) 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 over a quarter of children were enrolled in star 4 preschools. About a quarter of children resided in the northeast, northwest and southwest regions of the city. The smallest proportion of children lived in southeast Denver.

The right hand side of Table 2 presents demographic characteristics by provider type. 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 the distribution of Hispanic, white, and multi-racial children. There was a much larger percentage of Hispanic children in DPS sites, over twice

[9] 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.

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

the magnitude of the proportion of Hispanic children in community sites.[11] In contrast, DPS sites had a smaller proportion of white children enrolled than did community sites.[12] Community sites also had a significantly larger proportion of multi-racial children.[13] 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. Nearly half 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.

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, 5 and 6 across the provider types. DPS sites enrolled a much larger proportion of children from tier 1 than community sites.[14] Forty percent of children enrolled in DPS sites were from tier 1 compared to 21% in community sites. Conversely, community sites enrolled a higher proportion of children from tiers 5 and 6—income not reported than DPS sites.[15] One quarter of children in community sites were from tier 5 compared with just 13% of children from DPS sites. Fourteen percent of children in community sites were from families assigned to tier 6 because their parents opted out of the requirement to provide income, compared with just 5% of children from DPS sites.

The association between provider type and Qualistar Rating was also significant. Follow-up analyses revealed that this was largely due to the enrollment distribution in star 3 and star 4 sites. About two-thirds of children in DPS sites were enrolled in star 3 programs as compared with slightly less than half of children in community sites.[16] Nearly half of children in community sites were enrolled in star 4 sites compared with about one-quarter of children in DPS sites.[17] In sum, the vast majority of children in DPS sites were enrolled in star 3 programs, whereas the distribution of children between star 3 and star 4 programs was more even among children enrolled in community sites.

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.

During the course of the school year, one child transitioned from a DPS site to a community site. In the spring 2011, there were five children that were lost to follow-up for the following reasons: three children did not continue with the study because they moved out of Denver before the spring round, one child left the site and we were unable to contact them, and one child transferred to a non-DPP approved preschool. An alternate from the same strata was selected for four of these children and assessed during the spring round.[18] One child was deemed

---

[11] $\chi^2_1=20.89$; $p<.0001$

[12] $\chi^2_1=13.20$; $p<.001$

[13] $\chi^2_1=4.71$; $p<.05$

[14] $\chi^2_1=8.52$; $p<.01$

[15] Tier 5: $\chi^2_1=4.68$; $p<.05$; Tier 6—Income Not Reported: $\chi^2_1=4.71$; $p<.05$

[16] $\chi^2_1=11.59$, $p<.001$

[17] $\chi^2_1=14.67$, $p<.001$

[18] 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.

no longer eligible for the study at the very end of the data collection window. We were unable to select an alternate for this child before the close of the school year. As a result, the sample size for the spring round is 199 and the total sample size for the 2010-11 school year is 204.

Table 2: Cohort 3 Sample Characteristics Fall 2010[1]

	Entire Sample, weighted[2]	By Provider Type, Unweighted		
Characteristic		Community	DPS	Significance of Difference by Provider Type
Sex				χ^2_1=.02; ns.
Female	50.2%	51.0%	50.0%	
Male	49.8%	49.0%	50.0%	
Ethnicity				χ^2_5=26.70; p<.0001
Hispanic	53.1%	27.0%	59.0%	
White (not of Hispanic origin)	30.7%	51.0%	26.0%	
African-American (not of Hispanic origin)	8.4%	10.0%	8.0%	
Multi-Racial	3.3%	9.0%	2.0%	
Asian/Pacific Islander	3.0%	3.0%	3.0%	
Native	1.6%	0.0%	2.0%	
Child's Primary Language				χ^2_1=31.35; p<.0001
English	55.1%	86.0%	48.0%	
Another Language	38.7%	11.0%	45.0%	
Not Reported	6.3%	3.0%	7.0%	
Home Language				χ^2_1=32.44; p<.0001
English	54.1%	81.0%	48.0%	
Another Language	36.5%	8.0%	43.0%	
Not Reported	9.4%	11.0%	9.0%	
DPP Income Tier[3]				χ^2_5=19.92;p<.01
Tier 1	36.5%	21.0%	40.0%	
Tier 2	26.0%	17.0%	28.0%	
Tier 3	7.5%	14.0%	6.0%	
Tier 4	8.2%	9.0%	8.0%	
Tier 5	15.2%	25.0%	13.0%	
Tier 6—Income Not Reported	6.7%	14.0%	5.0%	
Star Level of Preschool				χ^2_2=13.73; p<.0101
Star 1	0.0%	0.0%	0.0%	
Star 2	8.6%	7.0%	9.0%	
Star 3	63.7%	45.0%	68.0%	
Star 4	27.7%	48.0%	23.0%	
Region of the City				χ^2_4=8.87; n.s.
Central	14.1%	23.0%	12.0%	
Northeast	24.5%	18.0%	26.0%	
Northwest	22.4%	24.0%	22.0%	
Southeast	10.1%	15.0%	9.0%	
Southwest	29.0%	20.0%	31.0%	

[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 2010.
[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.

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 2010-11 school year. Each set of analyses are described in turn below.

FALL 2010

COMMUNITY SAMPLE

Children who were included in the community sample were compared to 472 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.[19] Follow-up analyses revealed that this effect was due to tiers 1 and 3. In the population of DPP children enrolled in community sites, 33% of children were in Tier 1. In the sample, 21% of children were in Tier 1. Conversely, in the population, 8% of children were in Tier 3. In the sample, 14% 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.[20] That is, the community sample was slightly wealthier but was otherwise representative of the population of enrolled children in August 2010.

DPS SAMPLE

Children who were included in the DPS sample were compared to 2405 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.[21] That is, the DPS sample was representative of the population of enrolled children in September 2010.

---

[19] Income Tier: $\chi^2_6=13.76$, $p<.05$

[20] Gender: $\chi^2_1=0.28$, n.s.; ethnicity: $\chi^2_6=9.89$, n.s.; Qualistar rating: $\chi^2_3=1.05$, n.s.; home language: $\chi^2_1=1.05$, n.s.; child primary language: $\chi^2_1=2.26$, n.s.; region of the city: $\chi^2_4=6.01$, n.s.

[21] Gender: $\chi^2_1=0.02$, n.s.; ethnicity: $\chi^2_6=4.11$, n.s.; income tier: $\chi^2_6=2.79$, n.s.; Qualistar rating: $\chi^2_2=2.18$, n.s.; home language: $\chi^2_1=0.71$, n.s.; child primary language: $\chi^2_1=0.76$, n.s.; region of the city: $\chi^2_4=6.34$, n.s.

COMMUNITY SAMPLE

Children who were included in the community sample were compared to 1688 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: child gender, child ethnicity, income tier, Qualistar rating of the child's preschool, home language, child's primary language, and region of the city. Similar to the fall, there was a significant effect for tier level.[22] Follow-up analyses revealed that the sample underrepresented children from tier 1[23] and overrepresented children from tiers 3 and 5, as well as tier 6, who did not provide their income.[24] There were also significant effects for ethnicity, child language, and home language.[25] Follow-up analyses revealed that this was because the sample included more white children and fewer black and Hispanic children than were included in the population at the end of the school year.[26] By the end of the school year, 78% the population of DPP children enrolled in community sites had English as the identified home language. Ninety-two percent of the community sample had English as their identified home language. Similarly, by the end of the school year, 78% of the population of DPP children enrolled in community sites spoke English as their primary language. Ninety percent of the children in the community sample spoke English as their primary language.

These ethnicity 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, 572 children were enrolled. Over the course of the year, 1255 additional children from community sites enrolled. A large proportion of these children were black (17%) and Hispanic (40%), changing the racial/ethnic distribution of enrolled children. About a third of this group of later enrolling children had home languages (31%) and primary languages (34%) 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.[27] In sum, as in the fall, the sample was somewhat wealthier than the population as a whole in summer 2011. Further, because later enrolling children were more likely to be black, Hispanic and non-English speaking, the sample includes more white and English speaking children than the population as a whole in summer 2011.

---

[22] $\chi^2_6=53.12$, $p<.0001$

[23] 20.2% of children in the sample were from tier 1 compared with 48.2% in the population as a whole.

[24] 14.1% of children in the sample were from tier 3 compared with 4.4% in the population as a whole; 26.2% of the in the sample were from tier 5 compared with 16.8% in the population as a whole. 12.2% of children in the sample were from families that chose to opt-out of providing their income compared with 6.8% in the population as a whole.

[25] Ethnicity: $\chi^2_5=18.12$, $p<.01$; home language: $\chi^2_1=9.69$, $p<.01$; child primary language: $\chi^2_1=7.41$. $p<.01$

[26] The sample included 52% white, 9% black, and 26% Hispanic children compared with 34% white, 18% black, and 38% Hispanic at the end of the school year.

[27] Gender: $\chi^2_1=.01$, n.s.; Qualistar rating: $\chi^2_3=7.16$, n.s.; region of the city: $\chi^2_4=4.64$, n.s.

DPS SAMPLE

Children who were included in the DPS sample were compared to 4005 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.[28] 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 teachers at the time of the 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 for a half-day observation. These observations took place between January and June.

Parents were mailed a survey about their children's socio-emotional development in January 2011. Follow-up mailings and phone calls were used to boost response rates. Parents were asked to complete the survey just one time during the course of the school year. A Spanish version of the survey was available for parents and teachers who preferred to complete it in Spanish.

Table 3 presents the total sample sizes for each data collection activity. About a third of the children in the sample spoke Spanish and completed assessments in Spanish as well as English. Response rates for the teacher surveys were excellent, with nearly all teachers completing the survey in the fall and spring. Response rate for the parent survey was also excellent; all parents returned their surveys. In spite of a shortened data collection window for the classroom observations, we were able to conduct observations in the classrooms attended by over three-quarters of the children in the sample.

[28] Gender: χ^2_1=.02, n.s.; Ethnicity: χ^2_6=3.99, n.s.; income tier: χ^2_6=10.29, n.s.; Qualistar rating: χ^2_3=.64, n.s.; home language: χ^2_1=.61, n.s.; child primary language: χ^2_1=.96, n.s.; region of the city: χ^2_4=3.01, n.s.

Table 3: Sample sizes by data collection type, Fall 2010 and Spring 2011

Data Collection Activity	Fall 2010	Spring 2011
Standardized Assessments—English	200	199
Standardized Assessments—Spanish	63	63
DECA—Teacher Report	197 (98%)	196 (98%)
DECA—Parent Report	N/A	199 (100%)
Classroom Observations[1]	158 (77%)	

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

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, and second grade was 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.[29]

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 (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.

[29] Early, D. M., Barbarin, O., Bryant, D. M., Burchinal, M., Chang, F., Clifford, R. M., Crawford, G. M., Howes, C., Ritchie, S., Kraft-Sayre, M. E., Pianta, R. C., Barnett, W. S., & Weaver, W. (2005). Pre-kindergarten in eleven states: NCEDL's Multi-State Study of Pre-Kindergarten and study of State-Wide Early Education Programs (SWEEP): Preliminary descriptive report. Chapel Hill, NC: National Center for Early Development & Learning.

Table 4: Areas of Child Development Assessed

Area Assessed	Name of Assessment	Acronym	Language of Assessment
Receptive Vocabulary	Peabody Picture Vocabulary Test-4[30]	PPVT	English
	Test de Vocabulario en Imagenes Peabody[31]	TVIP	Spanish
Literacy Skills	Woodcock-Johnson III Achievement Battery,[32] Letter-Word Identification Subtest	WJ LWI	English
	Batería III Woodcock-Muñoz,[33] 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[34]	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 teacher-child interactions.[35] Observers visit the classroom and observe for up to 6 30-minute cycles. Each cycle

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

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

[32] 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.

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

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

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

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)[36] and its Spanish language counterpart Evaluación del Desarrollo de la Lectura (EDL2).[37] 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.[38] The assessment yields a reading level for each child. In kindergarten, a reading level of 4 is considered reading on grade level.[39] In first grade, a reading level of 16 is considered on grade level. In second grade, 28 is considered on grade level. In third grade, a reading level of 38 is considered on grade level.

RESULTS: PRESCHOOL YEAR

PRELIMINARY ANALYSES

Table 5 presents descriptive statistics for fall and spring child outcome measures. The PPVT, TVIP, WJ and WM are all scaled such that 100 is an average score, with a standard deviation of 15. Scores within one standard deviation of the mean are considered in the average range (i.e., 85-115). All scores are adjusted for the child's age at the time of assessment. As such, one would expect a child who is developing at an average rate to have the same score over time. In both the fall and the spring, children, on average, scored in the average range for all of the standardized assessments, with the exception of the TVIP in the fall, which was just slightly below the average range. 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.

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

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

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

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

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 and slightly lower scores on Attachment. Once again there was substantial variability in all of the scores.

Table 5: Weighted Descriptive Statistics for Child Outcome Measures

Variable	Fall 2010				Spring 2011			
All Children	N	Mean	Standard Deviation	Range	N	Mean	Standard Deviation	Range
Standardized Assessments								
PPVT Standard Score	200	90.86	26.51	28-145	199	94.53	25.33	20-144
WJ LWI Standard Score	200	99.41	14.64	68-159	199	103.50	13.84	75-167
WJ AP Standard Score	200	104.33	19.47	11-142	199	106.18	11.72	64-136
Teacher-Rated DECA								
Initiative T-Score	197[1]	51.35	7.90	30-61	196	54.72	6.72	31-61
Self-Control T-Score	197	57.27	9.11	33-72	196	59.21	8.93	31-72
Attachment T-Score	197	49.08	8.62	30-72	196	51.95	9.01	28-72
Total Protective Factors T-Score	197	52.37	8.72	31-72	196	55.93	8.57	31-72
Behavioral Concerns T-Score	193	46.93	8.93	31-70	189	47.78	8.67	31-72
Parent-Rated DECA								
Initiative T-Score	--	--	--	--	199	50.33	8.79	30-72
Self-Control T-Score	--	--	--	--	199	54.01	7.99	30-72
Attachment T-Score	--	--	--	--	199	47.07	11.64	28-72
Total Protective Factors T-Score	--	--	--	--	199	49.62	9.00	30-72
Behavioral Concerns T-Score	--	--	--	--	199	52.90	10.54	28-72
Spanish-Speaking Children Only								
Standardized Assessments								
TVIP Standard Score	62	84.35	17.79	59-145	62	89.97	21.69	56-120
WM LWI Standard Score	63	96.47	13.68	69-119	63	101.27	14.04	78-125
WM AP Standard Score	63	94.36	13.71	56-116	63	96.17	11.18	75-115

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

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 (primary language=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 nearly 3 standard deviations higher on the PPVT than their counterparts whose primary language was something other than English. For LWI, children whose primary language was English scored

nearly one standard deviation higher on average than their counterparts whose primary language was something other than English. Children whose primary language was English scored about 1.5 standard deviations higher on AP than children whose primary language was something other than English. 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, over two standard deviations in magnitude. Differences between primary language groups for LWI and AP were slightly smaller than the fall, but still statistically significant. For LWI, the difference between language groups was nearly one standard deviation in magnitude. For AP, the difference between the groups was about 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	134	108.10	15.79	56	66.54	20.57	16.17***
WJ LWI Standard Score	134	104.70	12.32	56	91.43	12.48	7.23***
WJ AP Standard Score	134	114.64	11.34	56	89.50	21.87	11.15***

*p<.05, **p<.01, ***p<.001
[1]Information about the child's primary language was missing for one child 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	111.75	15.93	56	74.32	20.14	12.38***
WJ LWI Standard Score	133	109.49	12.85	56	95.64	11.06	7.04***
WJ AP Standard Score	133	111.56	10.04	56	99.84	10.21	7.30***

*p<.05, **p<.01, ***p<.001
[1]Information about the child's primary language was missing for one child in the sample.

PRESCHOOL QUALITY

The 204 children in the sample were enrolled in 144 classrooms in 104 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.[40]

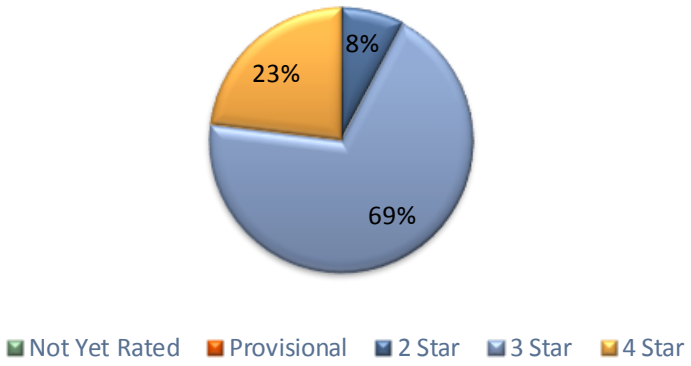
[40] It is important to keep in mind that all of the preschool quality information provided here is based on only a sample of 104 preschools where the children in the sample were enrolled. For information on the quality of all preschool programs participating in DPP during the 10-11 school year, readers are referred to the annual evaluation report prepared by Augenblick, Palaich and Associates.

QUALISTAR RATING AND ACCREDITATION

101 of the 104 preschools were Qualistar rated; 99 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. Three sites became eligible for DPP because they had obtained Accreditation from the National Association for the Education of Young Children (NAEYC). For these sites, the only quality data that is available is the number of stars.[41] Figure 1 presents the breakdown of programs by star level. Over two-thirds of the programs had 3 stars. Nearly another 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.

Table 8 presents descriptive statistics for the five component areas of the Qualistar rating for the 101 sites with a Qualistar rating.[42] 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, five 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 6 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 nearly 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, while others earned all of the possible points. A very small proportion of programs earned the 2 points for having earned an accreditation.

Figure 1: Star Level of Programs Attended by Children in the Sample (n=104 programs).



[41] Providers who were accredited by NAEYC prior to October 2006 received a DPP Quality Rating of 3 stars. Those who were accredited 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.

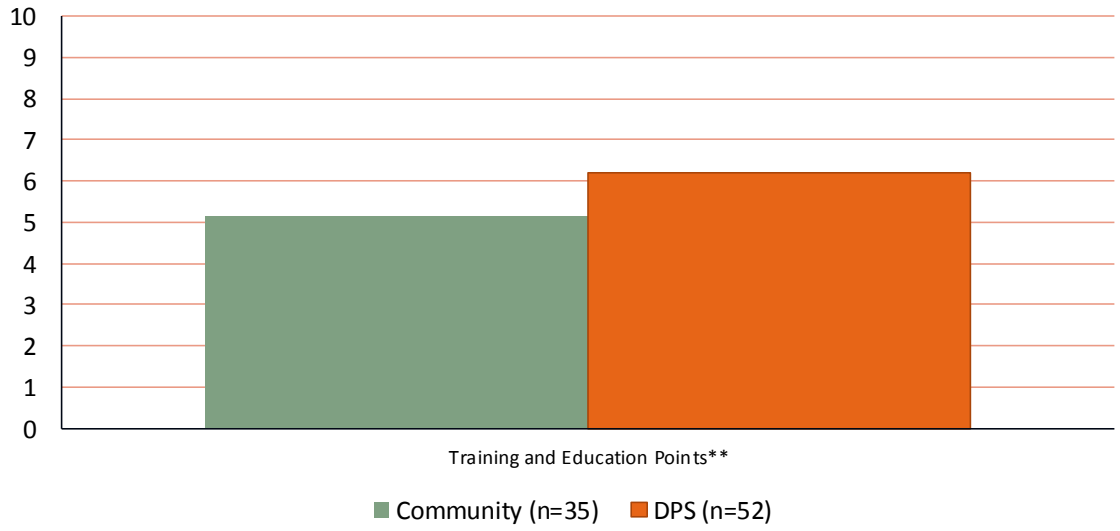
[42] More information about the five component areas of the Qualistar rating is available at: http://www.qualistar.org/qualistar-rating-components.html

Table 8: Qualistar Rating Components for Programs Attended by Children in the Sample (n=101 programs).

Component	Possible Range	Mean	SD	Range
Learning Environment Points	0-10	6.66	1.31	4-10
Family Partnerships Points	0-10	9.13	1.73	0-10
Training and Education Points	0-10	5.75	1.82	1-9
Adult-to-Child Ratios and Group Size Points	0-10	8.90	1.25	6-10
Accreditation Points	0-2	0.06	0.34	0-2

Analyses were conducted to test whether the type of provider (DPS vs. Community) was associated with the components of the Qualistar rating. The two types of programs only differed significantly in one area: training and education points (see Figure 2).[43] 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=98 programs)[1]



** p<.01

[1] Standard deviations: Community=1.92, DPS=1.62

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 income tier, child

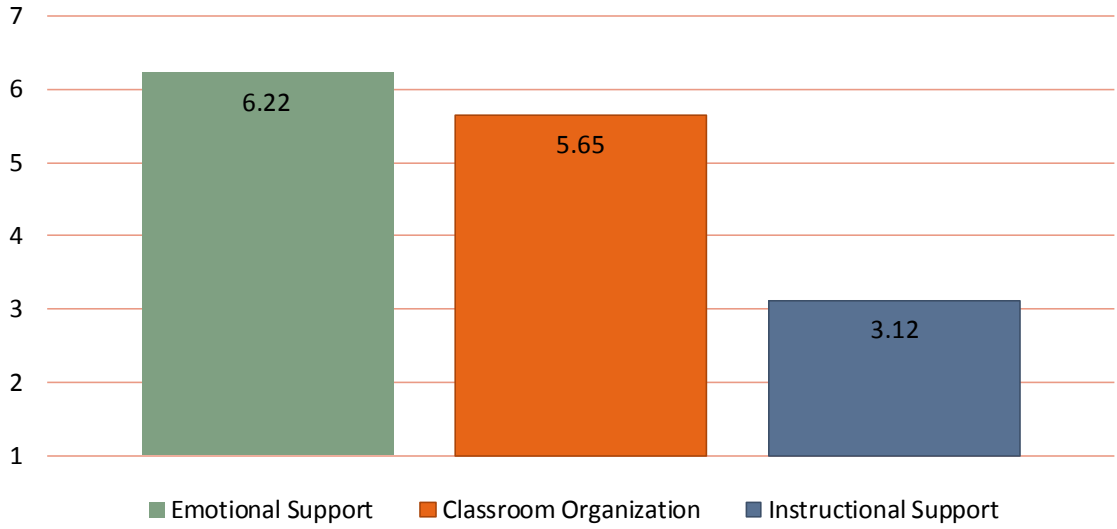
[43] t=3.02, df=99, p<.01

primary language, home language or ethnicity.[44] The only significant association was for region of the city.[45] Follow-up Tukey tests revealed that children residing in the northeast region of the city tended to be in the programs earning the highest number of rating points on average. Children in this region tended to be enrolled in programs that had earned significantly more rating points, on average, than children in the northwest region of the city. This difference was of small magnitude, however; there was just a 2 point difference between average scores for the northeast and northwest regions.[46]

CLASS OBSERVATIONS

Figure 3 displays the mean scores for the 116 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 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.

Figure 3: Average CLASS Domain Scores (n=116 Classrooms)



Figures 4, 5 and 6 provide some information about the variability in these domain scores. For Emotional Support and Classroom Organization, the vast majority of classrooms scored in the high range (scores above 5) and the remainder scored in the middle-range (scores between 3-5). For Instructional Support, very few classrooms

[44] Income tier: F(3,186)=2.12, n.s.; child primary language: F(1,179)=3.32, n.s.; primary home language: F(1,175)=2.64, n.s.; ethnicity: F(5,187)=1.61, n.s.

[45] F(4,186)=2.88, p<.05

[46] 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

scored in the high range, and the remainder were split fairly evenly between the low (scores below 3) and middle-range.

Figure 4: Distribution of Scores for Emotional Support (n=116 Classrooms)

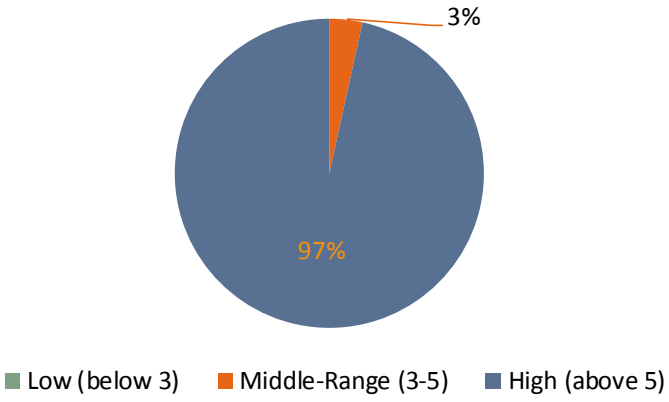


Figure 5: Distribution of Scores for Classroom Organization (n=116 Classrooms)

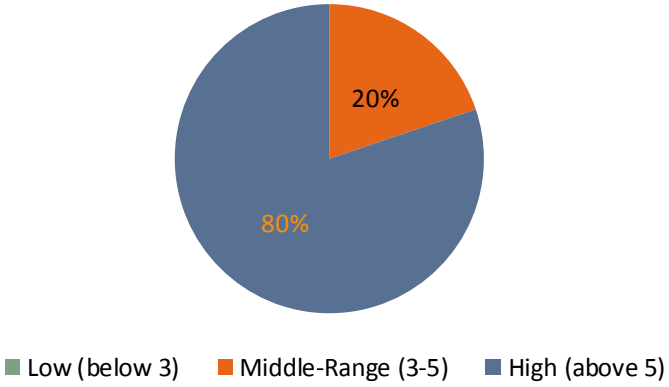
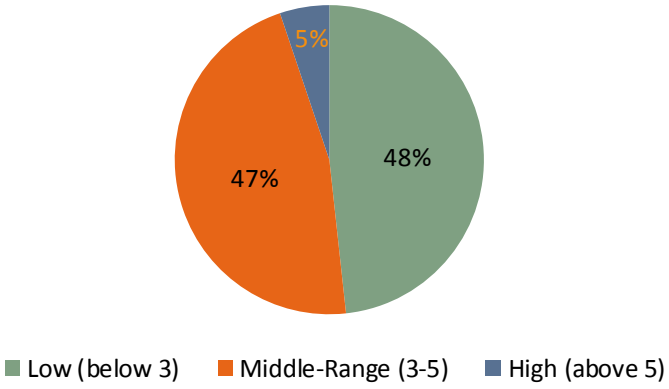
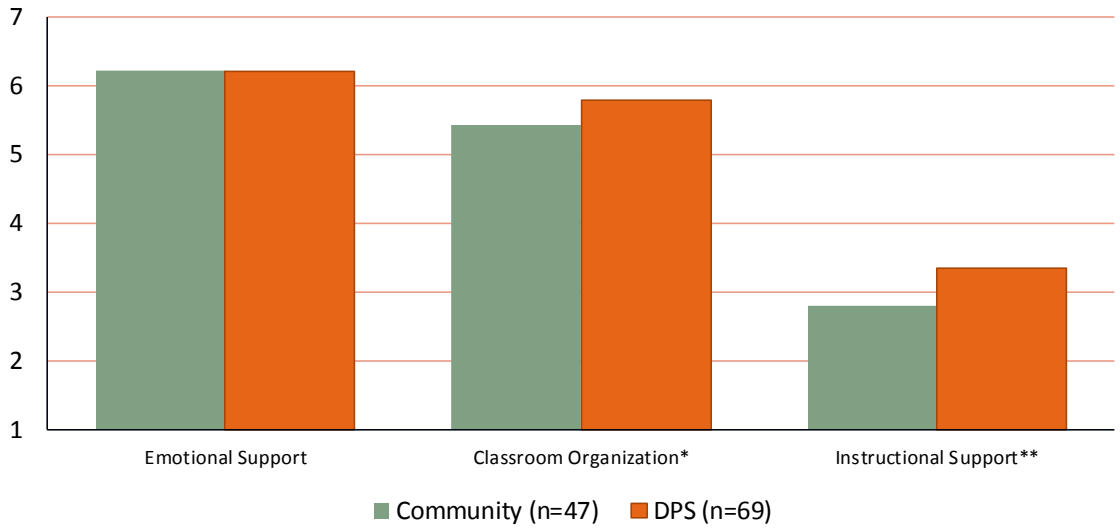


Figure 6: Distribution of Scores for Instructional Support (n=116 Classrooms)



We also conducted analyses to test for differences in CLASS domain scores by provider type. The results of these analyses are presented in Figure 7. Scores for Emotional Support were, on average, similar for the two provider types. Scores differed significantly for the other two domain scores, however. On average, DPS preschools scored about a third of a point higher than community preschools on Classroom Organization. DPS preschools scored about a half of a point higher than community preschools on Instructional Support.

Figure 7: CLASS Domain Scores by Provider Type



*p<.05; **p<.01

[1]Standard deviations: Emotional Support—Community=0.55, DPS=0.51; Classroom Organization—Community=0.75, DPS=0.74; Instructional Support—Community=1.06, DPS=1.02

As with the Qualistar Rating, we conducted analyses to test for associations between CLASS domain scores and child and family background characteristics. None of the domain scores was significantly associated with income tier or ethnicity.[47] Both child primary language and home language were significantly associated with Classroom Organization, but not Emotional Support or Instructional Support.[48] Children who did not speak English as their primary language tended to be enrolled in classrooms with higher levels of Classroom Organization than children who did speak English as their primary language. A similar pattern was observed for home language (see Figure 8). There were also significant associations between region of the city and both Emotional Support and Classroom Organization, but not Instructional Support.[49] Follow-up Tukey tests revealed that children residing in the northeast region of the city tended to be in the programs earning the highest number of points on Emotional

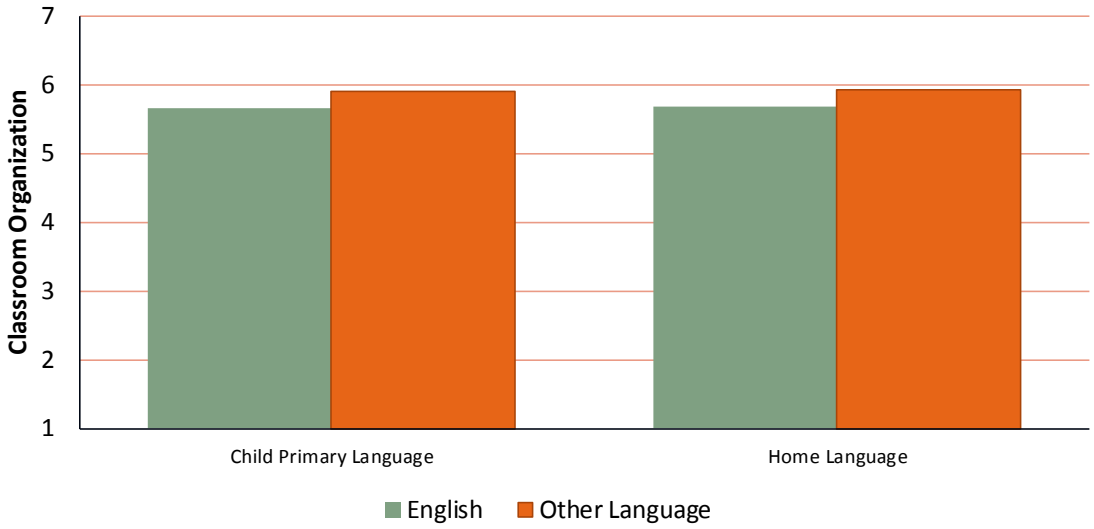
[47] Income tier: Emotional Support—F(3,157)=1.24, n.s.; Classroom Organization—F(3,157)=0.42 n.s.; Instructional Support—F(3,157)=0.58, n.s.; ethnicity: Emotional Support—F(3,157)=0.97, n.s.; Classroom Organization—F(3,157)=1.00, n.s.; Instructional Support—F(3,157)=0.29, n.s

[48] Child Primary Language: Emotional Support—F(1,152)=2.73, n.s.; Classroom Organization—F(1,152)=4.05, p<.05; Instructional Support—F(1,152)=0.15, n.s.; Home Language: Emotional Support—F(1,144)=1.41, n.s.; Classroom Organization—F(1,144)=4.06, p<.05; Instructional Support—F(1,144)=1.69, n.s.

[49] Region of the City: Emotional Support—F(4,157)=5.09, p<.001; Classroom Organization—F(4,157)=4.57, p<.01; Instructional Support—F(4,157)=2.04, n.s.

Support and Classroom Organization, on average. These children tended to be enrolled in classrooms that had earned significantly more Emotional Support points, on average, than children in the southwest, central and southeast regions of the city. The largest difference was between the southeast region and northeast region, a difference of nearly half of a point.[50] Similarly, children residing in the northeast region tended to be enrolled in classrooms that earned significantly more Classroom Organization points than children living in the southwest and southeast regions of the city. The largest difference here was between the northeast and southeast regions, a difference of about two-thirds of a point.[51]

Figure 8: Classroom Organization Scores, by Child and Home Language



KINDERGARTEN READINESS

STANDARDIZED ASSESSMENTS

Analyses were conducted to determine how ready 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 9 presents the percent of children scoring

[50] Emotional Support: Central Region-mean=6.04, sd=0.63; Northeast Region-mean=6.47, sd=0.35; Northwest Region-mean=6.31. sd=0.44; Southeast Region-mean=6.01, sd=0.51; Southwest Region-mean=6.12, sd=0.55

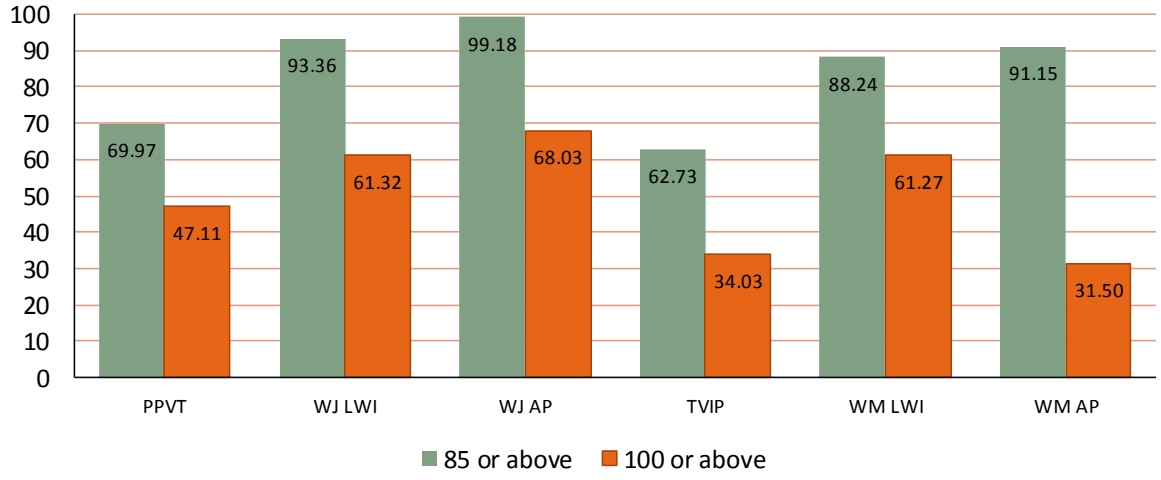
[51] Classroom Organization: Central Region-mean=5.71, sd=0.65; Northeast Region-mean=6.12, sd=0.53; Northwest Region-mean=5.78. sd=0.69; Southeast Region-mean=5.45, sd=0.77; Southwest Region-mean=5.56, sd=0.90

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. Over half of children scored 100 or above on WJ LWI and over two-thirds of children scored 100 or above on WJ Applied 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 (96%) scored 85 or above on the PPVT as compared with a relatively small proportion of children whose primary language was not English (31%).[52] 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 85% of children whose primary language was not English.[53] For WJ Applied Problems, an accurate statistical test could not be computed because all children whose primary language was English scored 85 or above. This is compared with 98% of children whose primary language was not English.

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



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 9% of children with another primary language.[54] For WJ LWI, 80% of children whose primary language was

[52] χ^2_1=88.69, p<.0001

[53] χ^2_1=13.82, p<.001

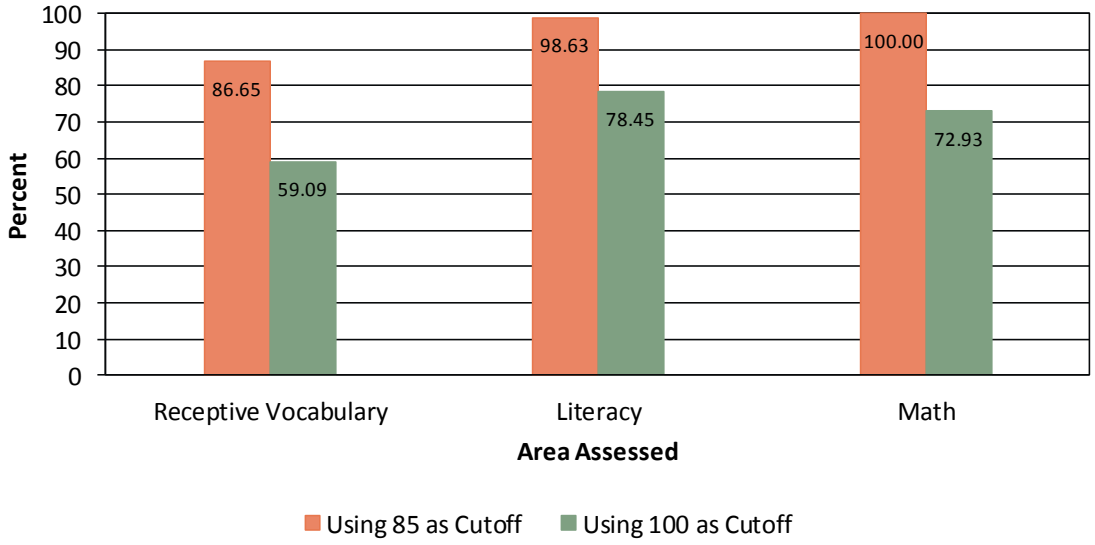
[54] χ^2_1=80.59, p<.0001

English scored 100 or greater as compared with 36% of children with another primary language.[55] Finally, for WJ Applied Problems, 87% of children whose primary language was English earned scores of 100 or above compared with 41% of children whose primary language was something other than English.[56]

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 slightly less than 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 about 30% 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 10. 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 all children did on the math assessment (WJ-AP or WM-AP). When a score of 100 was used as a cutoff, over half of children met or exceeded this benchmark for vocabulary; approximately three-quarters met or exceeded this benchmark for both literacy and math.

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



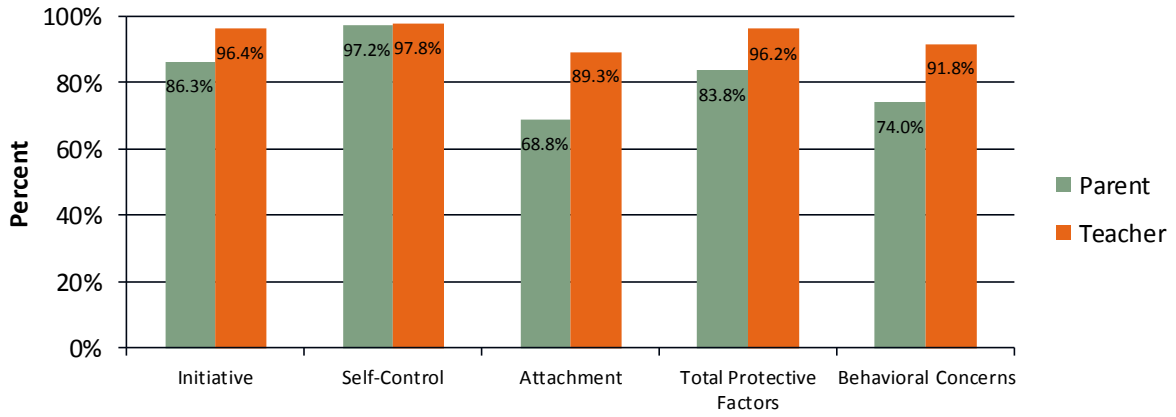
[55] χ^2_1=38.84, p<.0001

[56] χ^2_1=45.29, p<.0001

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. 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 over two-thirds of children in the typical or strength range for Attachment and nearly three-quarters of children in the typical range for Behavioral Concerns. Teachers rated over 90% of children in the typical or strength range for Initiative, Self-Control, Total Protective Factors, and Behavioral Concerns. They also rated nearly 90% of children in the typical range for Attachment.

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



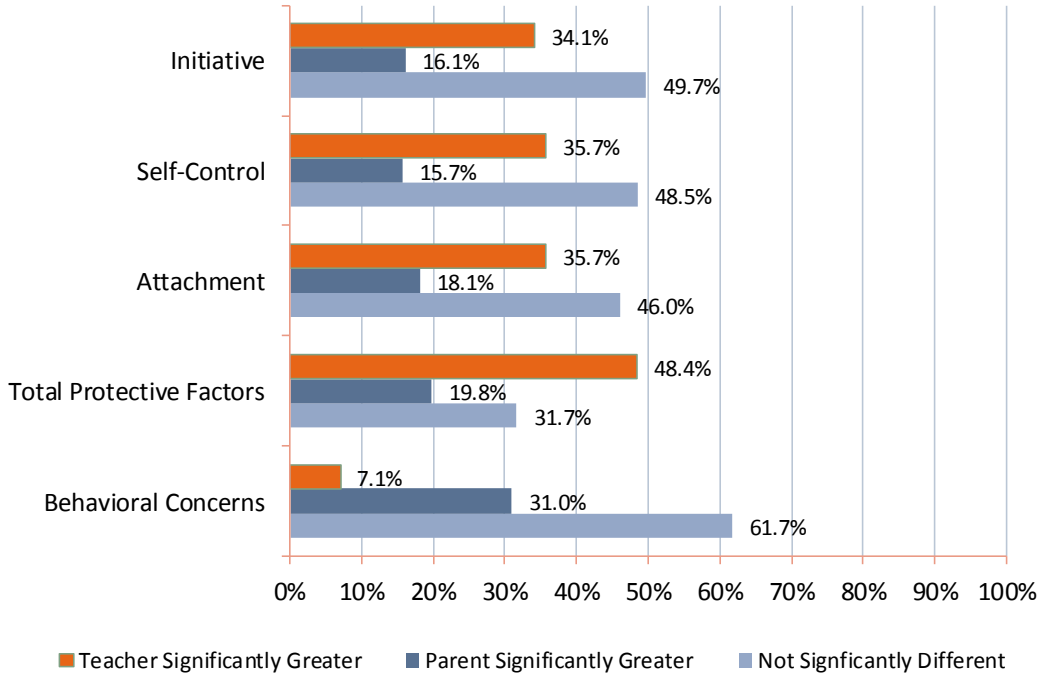
We examined the differences between teachers' and parents' ratings using guidelines from the authors of the DECA. The authors developed these guidelines to help users distinguish between differences in scores due to measurement error and differences that are likely due to a meaningful difference between scores. For Initiative, a difference of 10 is needed to conclude that there is a significant difference between the parent and teacher rating. The average difference between teachers' and parents' reports, 4.4 (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 half of children, teachers' and parents' reports did not significantly differ. For slightly over 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, 5.2 (sd=11.1), did not exceed this threshold, indicating that, on average, teachers' and parents' ratings did not differ. As displayed in Figure 12, for nearly half of children, teachers' and parents' reports did not significantly differ. For about a third of the sample, 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, 4.8 (sd=14.3), did not exceed this threshold. Once again, for about 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 6.3 (sd=12.0), which was below that threshold. However, for nearly half of children, the teacher's rating was significantly greater than the parent's rating (see Figure 12). For slightly less than a third 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 4.9 (sd=12.7), 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 twice as 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 were 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 4 points, or nearly one-third 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. The pattern for assessments administered in Spanish was similar, with significant and large increases observed for TVIP and WM LWI. The magnitude of the increase for both of these assessments was about one-third of a standard deviation. Children's scores remained relatively constant on the WM Applied Problems.

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

Variable	N	Fall Mean (Standard Deviation)	Spring Mean (Standard Deviation)	t
Standardized Assessments—English				
PPVT	195	90.55 (26.67)	94.57 (25.41)	4.84***
WJ-LWI	195	99.48 (14.78)	103.69 (13.78)	7.35***
WJ-AP	195	104.31 (19.66)	106.30 (11.78)	1.93
Standardized Assessments—Spanish				
TVIP	59	85.13 (17.50)	90.35 (21.47)	3.57***
WM-LWI	61	96.43 (13.90)	101.10 (14.21)	3.15**
WM-AP	61	95.09 (12.39)	96.20 (11.35)	0.99
Teacher Survey				
Initiative T-Score[1]	189	51.15 (7.95)	54.86 (6.58)	5.91***
Self-Control T-Score	189	57.18 (9.21)	59.37 (8.94)	2.64**
Attachment T-Score	189	49.03 (8.78)	51.96 (8.83)	4.10***
Total Protective Factors T-Score	189	52.22 (8.82)	56.09 (8.57)	5.12***
Behavioral Concerns T-Score	179	47.33 (8.93)	47.67 (8.75)	0.59

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

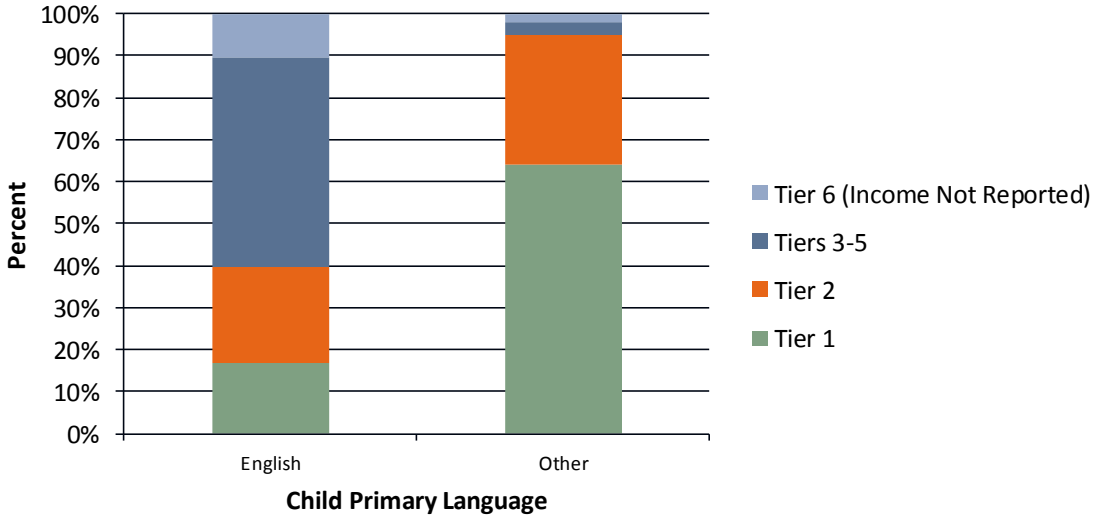
[1]Some teachers and parents left items blank on the DECA. Scores were only calculated if at least 75% of the items were present. This resulted in some missing data for the DECA.

Significant improvements were also observed in many 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), but there was not a significant change over time in Behavioral Concerns. Of particular note are the increases in Initiative and Total Protective Factors (about one-third of a standard deviation) and Attachment (over a quarter 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 tier 6).[57] It is important to note that these two background characteristics, income tier and child's primary language, are strongly associated (see Figure 13).[58] Nearly all children whose primary language is something other than English are from tiers 1 or 2 whereas only about 40% 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



[57] For analyses of assessments administered in Spanish, a three-level income tier group variable was used because only one family with a child assessed in Spanish opted out of reporting income. This family was dropped from the analyses.

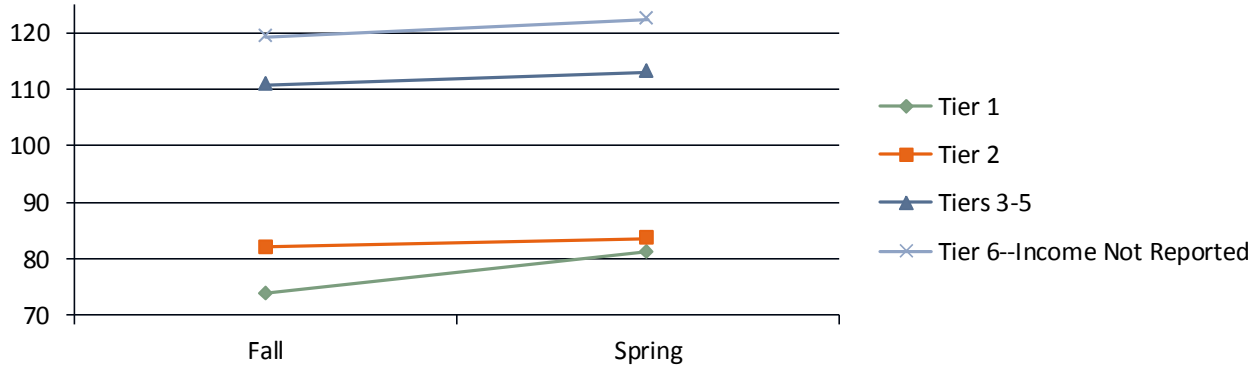
[58] χ^2_3=67.58, p<.0001

INCOME TIER

A series of Repeated Measures ANOVAs[59] was conducted with this income tier group predicting scores over time on assessments administered in English and Spanish as well as teacher-rated DECA. There were significant interactions between income tier group and time for PPVT,[60] WJ LWI,[61] WJ AP,[62] and DECA Attachment.[63] Results of these analyses are depicted in Figures 14-17.

Figure 14 shows average PPVT scores over time, by income tier group. The difference in average scores by tier groups is striking, with children in the lower income tiers scoring much lower on average than children in the higher tier groups. Follow-up Tukey[64] tests revealed that children from tier 1 increased significantly more over time than children in tier 2. Pairwise comparisons of the other groups were not significant.

Figure 14: Weighted PPVT Standard Scores over Time, by Income Tier Group[1]

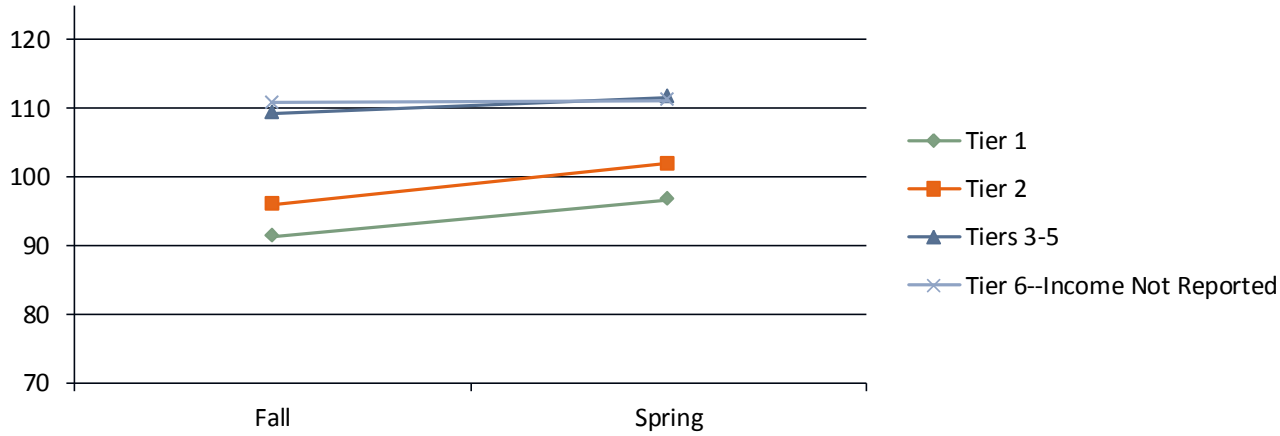


[1]Standard Deviations: Tier 1: Fall=24.43, Spring=23.84; Tier 2: Fall=24.40, Spring=26.13; Tiers 3-5: Fall=15.33, Spring=12.32; Tier 6: Fall=9.57, Spring=11.83.

[59] 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.

[60] $F(3,191)=3.25$, $p<.05$
[61] $F(3,191)=3.45$, $p<.05$
[62] $F(3,191)=12.08$, $p<.0001$
[63] $F(3,185)=3.90$, $p<.01$
[64] Results of ANOVA simply tell you that there is a difference between the specified groups on the outcome variable. When more than two groups are specified, follow-up tests are required to determine which pairs of groups are significantly different. Tukey tests are one particularly conservative type of follow-up test.

Figure 15: Weighted WJ Letter-Word Identification Standard Scores over Time, by Income Tier Group [1]

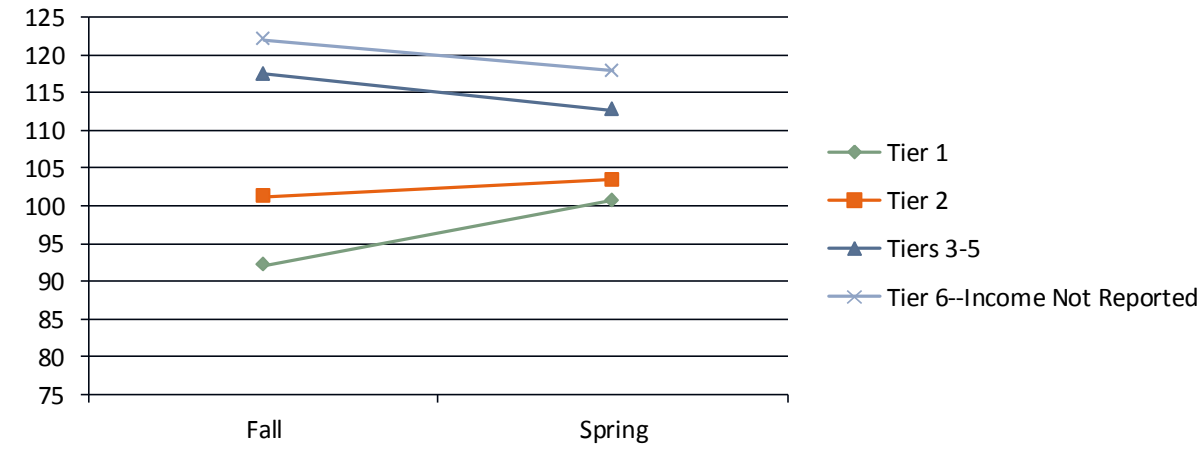


[1]Standard Deviations: Tier 1: Fall=12.70, Spring=12.71; Tier 2: Fall=14.67, Spring=14.94; Tiers 3-6: Fall=11.43, Spring=10.72; Tier 6: Fall=9.09, Spring=8.76.

Results of the analysis of WJ LWI are presented in Figure 15. Follow-up Tukey tests revealed a trend toward significant differences in change over time between tier 2 and both tiers 3-5 and tier 6, the group that did not report income. Children in tier 2 increased significantly more over the course of the school year than children in the higher tier groups.

Results of the analysis of WJ Applied Problems are presented in Figure 16. Follow-up Tukey tests revealed that tier 1 was significantly different than the tiers 3-5 group and tier 6, who did not report their income. Children in tier 1, on average, increased over time, while the tiers 3-5 and tier 6 groups actually decreased over time, on average. It is important to note that both of these groups started out the year with average scores that were extremely high, over one standard deviation above the mean. Because their scores were so extremely high, it is not surprising that scores decreased over time. Average scores in the spring were still extremely high.

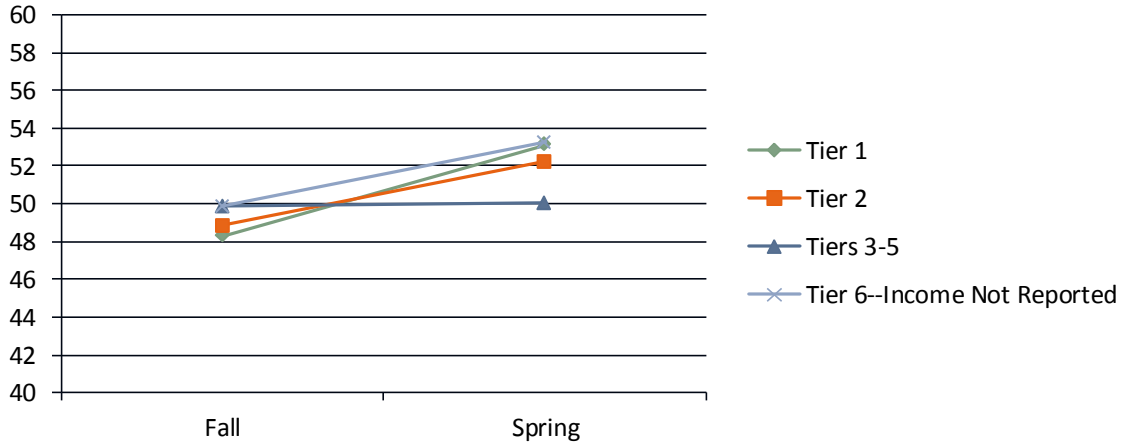
Figure 16: Woodcock-Johnson Applied Problems Scores over Time, by Income Tier Group [1]



[1]Standard Deviations: Tier 1: Fall=20.74, Spring=12.35; Tier 2: Fall=19.13, Spring=10.57; Tiers 3-5: Fall=10.39, Spring=8.40; Tier 6: Fall=6.52, Spring=8.02.

Results for teachers' ratings on DECA Attachment over time are presented in Figure 17. Follow-up Tukey tests revealed a significant difference between tier 1 and the tier 3-5 group. Children in tier 1 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=9.91, Spring=9.58; Tier 2: Fall=10.06, Spring=9.56; Tiers 3-5: Fall=7.77, Spring=8.21; Tier 6: Fall=5.49, Spring=6.23.

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.[65] There were significant interactions between primary language group and time for PPVT,[66] WJ AP,[67] DECA Initiative,[68] DECA Attachment,[69] and DECA Total Protective Factors.[70] Results of these analyses are presented in Figures 18-22.

[65] 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.

[66] $F(1,183)=5.49$, $p<.02$

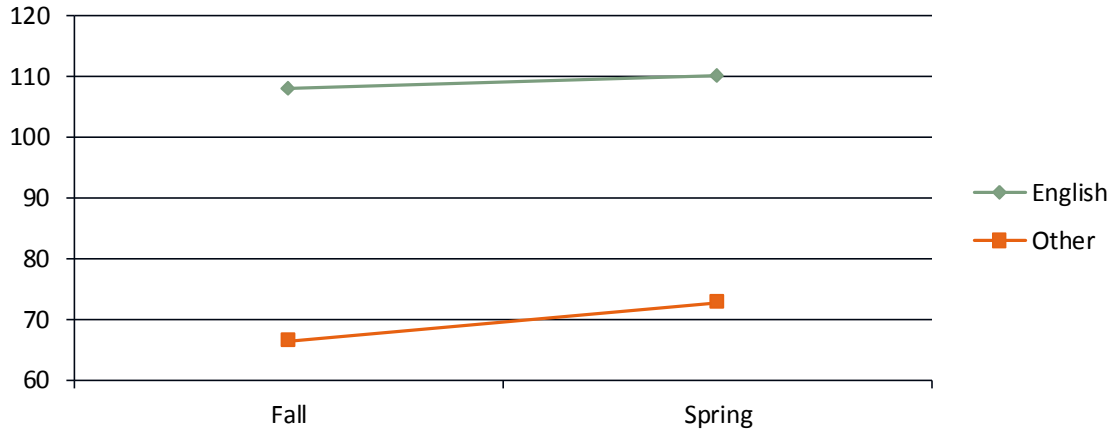
[67] $F(1,183)=46.39$, $p<.0001$

[68] $F(1,178)=4.52$, $p<.05$

[69] $F(1,178)=5.19$, $p<.05$

[70] $F(1,178)=3.98$, $p<.05$

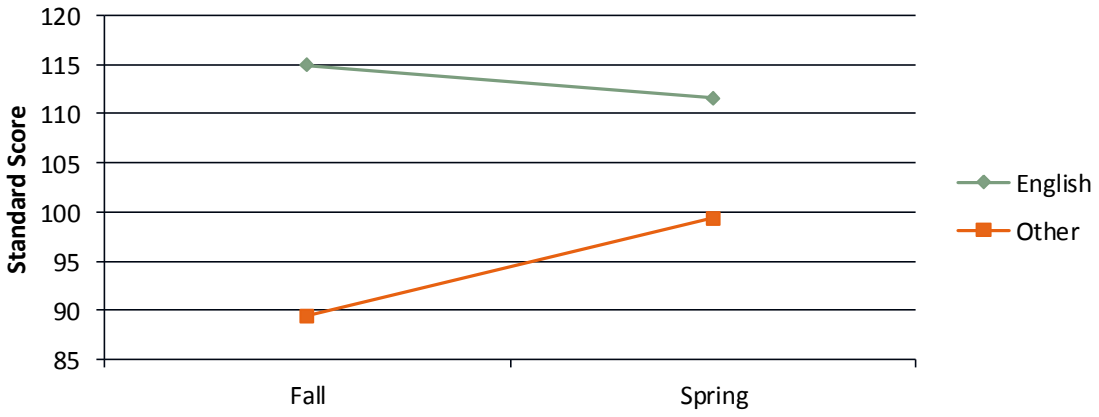
Figure 18: Weighted PPVT Scores over Time, by Child Primary Language[1]



[1]Standard Deviations: English: Fall=15.95, Spring=14.02; Other: Fall=20.74, Spring=24.62.

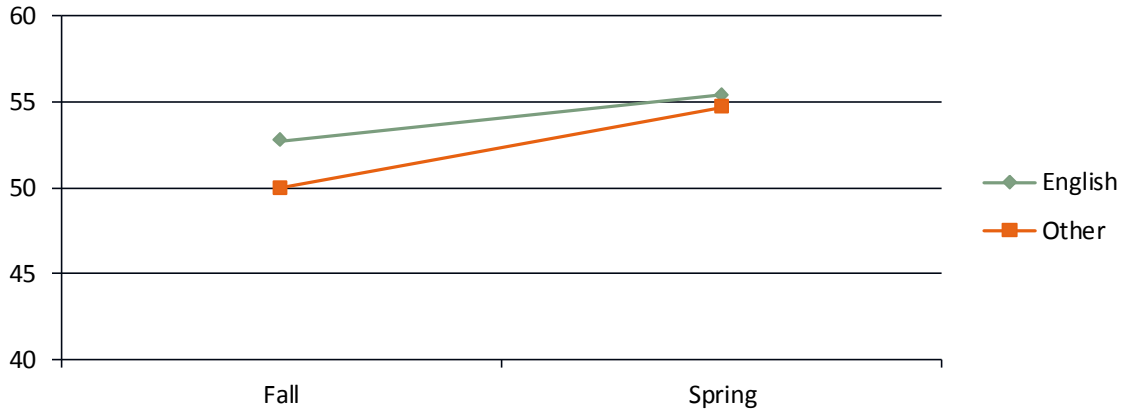
For all five assessments, the general pattern was that children whose primary language was not English tended to show larger increases in scores from fall to spring than their counterparts who spoke English as their primary language. For PPVT (see Figure 18), children whose primary language was English increased only slightly, while their counterparts with another primary language increased an average of over one-third of a standard deviation. For WJ AP (see Figure 19), 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 nearly two-thirds of a standard deviation on average. For Initiative (see Figure 20), children whose primary language was not English started lower and showed a steeper increase over the course of the year. At the end of the year, the two groups had similar scores. For Attachment (see Figure 21) and Total Protective Factors (see Figure 22), children whose primary language was not English increased over the course of the year, while children whose primary language was English remained relatively constant over time. For LWI, self-control, and behavioral concerns, the child primary language by time interaction was non-significant, indicating that children progressed similarly in these areas over the course of their preschool year, regardless of their primary language.

Figure 19: Weighted Woodcock-Johnson AP Scores over Time, by Child Primary Language[1]



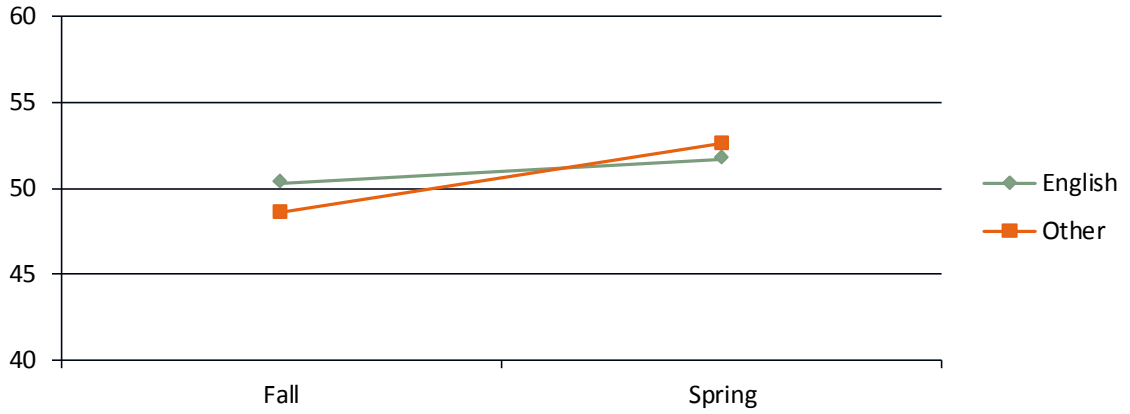
[1]Standard Deviations: English: Fall=11.298, Spring=9.04; Other: Fall=22.03, Spring=12.36.

Figure 20: Weighted Teacher-Rated Initiative Scores over Time, by Child Primary Language[1]



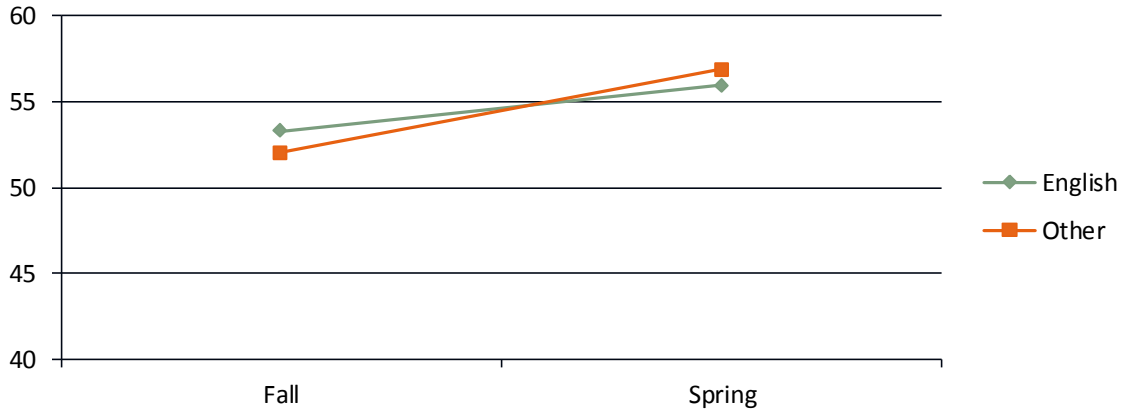
[1]Standard Deviations: English: Fall=6.99, Spring=6.06; Other: Fall=9.41, Spring=7.22.

Figure 21: Weighted Teacher-Rated Attachment Scores over Time, by Child Primary Language[1]



[1]Standard Deviations: English: Fall=7.94, Spring=8.16; Other: Fall=9.70, Spring=9.89.

Figure 22: Weighted Teacher-Rated Total Protective Factors over Time, by Child Primary Language [1]



[1] Standard Deviations: English: Fall=8.01, Spring=8.26; Other: Fall=9.94, Spring=8.93.

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.

There was not a strong pattern of significant associations in these analyses. For all of the English academic assessments (vocabulary, literacy and math), the correlations were non-significant. For the Spanish academic assessments, just one of the 6 correlations was significant and in an unexpected direction. Classroom Organization was significantly and negatively associated with Spanish Applied Problems (r=-.36, p<.05). That is, after taking into account children's math skills in the fall (assessed in Spanish), higher levels of Classroom Organization were associated with lower math skills as assessed in Spanish in the spring.

The pattern of results for the teacher DECA was more consistent. There were significant correlations, in the expected direction, between Instructional Support (but not Classroom Organization) and both protective factors (r=.18, p<.05) and behavioral concerns (r=-.18, p<.05). That is, after taking into account teachers' ratings of children's behavior in the fall, children in classrooms with higher levels of Instructional Support tended to demonstrate more protective factors and fewer behavior problems in the spring than did children in classrooms with lower levels of Instructional Support.

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

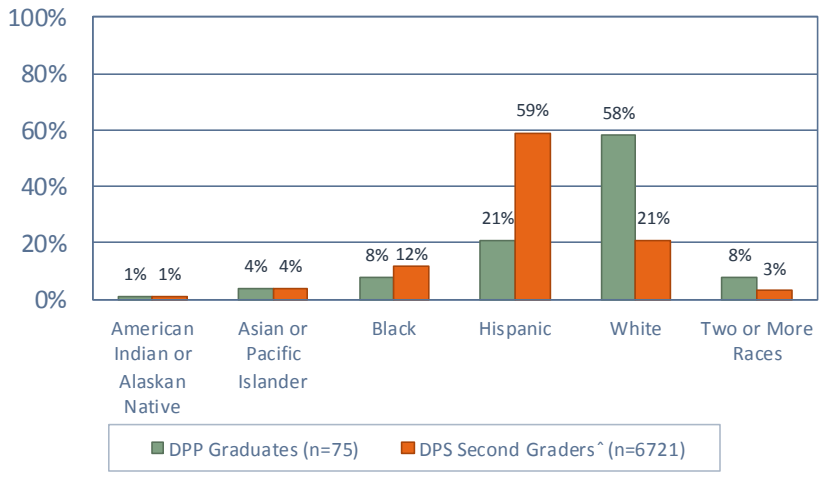
Prior to making comparisons between reading scores for DPP graduates and DPS as a whole, it is important to consider whether the samples of DPP graduates are similar demographically to the district as a whole. The Colorado Department of Education (CDE) provides demographic data on school districts in Colorado in the fall of each school year.[71] 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 23 displays a comparison of the racial/ethnic composition of the sample of Cohort 0 DPP graduates with reading assessment data for school year 10-11 and the population of children enrolled in second grade as of fall 2010. 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 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.

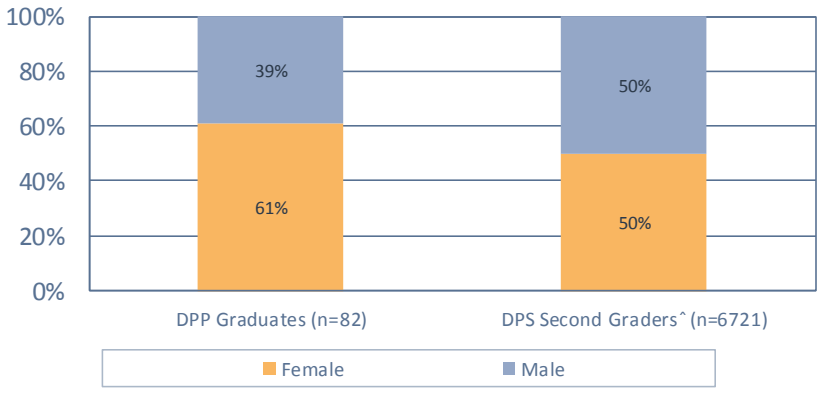
Figure 24 displays the gender composition for these same groups of children. Again, we see that, three years after DPP, Cohort 0 children differ markedly from the district's second graders as a whole. While the group of second graders

Figure 23: Race/Ethnicity of Cohort 0 DPP Graduates and All Second Graders Enrolled in DPS in Fall 2010*



*This figure includes only DPP Graduates for whom reading data were available for the 10-11 school year and omits 5 children 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.

Figure 24: Gender of Cohort 0 DPP Graduates and All Second Graders Enrolled in DPS in Fall 2010*



*This chart includes only Cohort 0 DPP Graduates for whom reading assessment data were available for the 10-11 school year.
^The group of DPS Second Graders includes the DPP graduates.

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

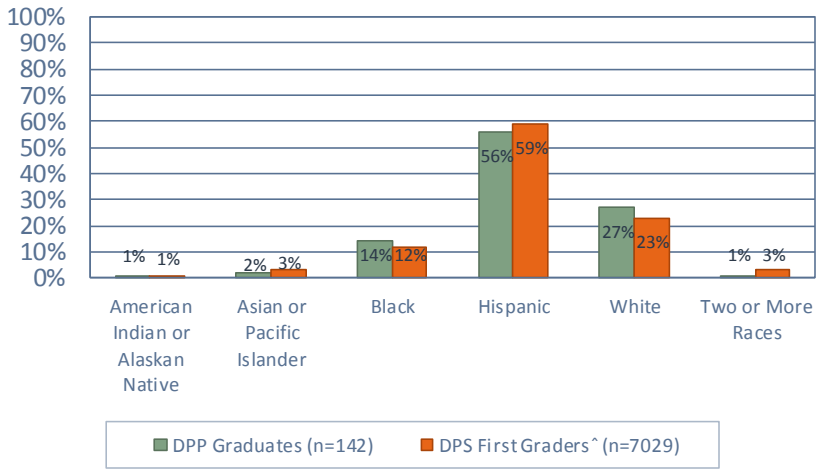
enrolled in DPS in fall 2010 were evenly divided between boys and girls. Cohort 0 children with reading assessment data in spring 2011 were comprised of nearly two-thirds girls and more than one-third boys.

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 25 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 first grade in DPS as of fall 2010. The racial and ethnic compositions of the two groups are remarkably similar.

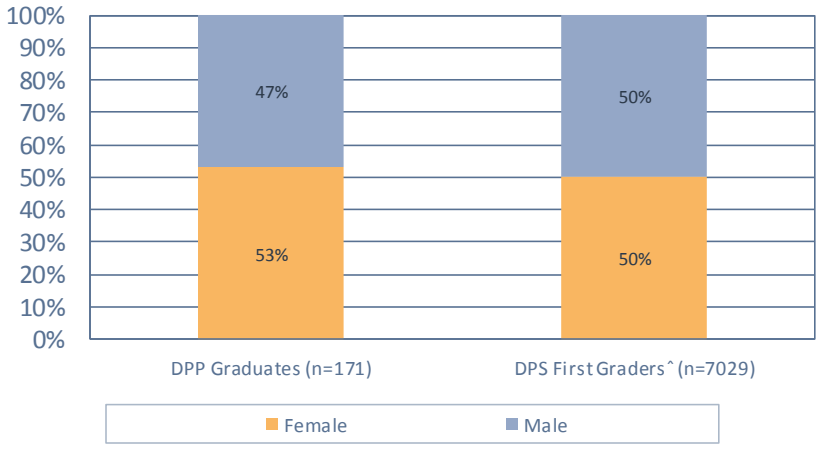
Figure 25: Race/Ethnicity of Cohort 1 DPP Graduates and All First Graders Enrolled in DPS in Fall 2010*



*This figure includes only Cohort 1 DPP Graduates for whom reading assessment data were available for the 10-11 school year and omits 29 DPP graduates with missing or ambiguous race/ethnicity data in the database of DPP children provided to Clayton by ACS. Perentages sum to 101 because of rounding error.
^The group of DPS First Graders includes the DPP graduates. Percentages sum to 101 because of rounding error.

Figure 26 presents the gender composition of Cohort 1 DPP graduates with reading assessment data and all first graders in DPS as of fall 2010. The district as a whole was split evenly between the genders. Cohort 1 DPP Graduates with spring 2011 reading assessment data were split a bit less evenly, with slightly more girls than boys, but still rather close to an even split.

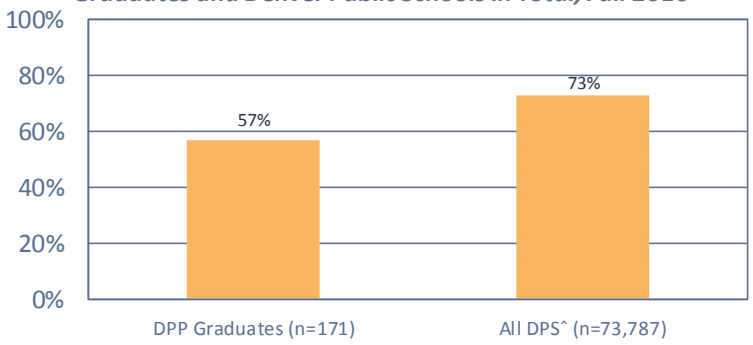
Figure 26: Gender of Cohort 1 DPP Graduates and All First Graders Enrolled in DPS in Fall 2010*



*This figure includes only Cohort 1 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.

Finally, Figure 27 presents the proportion of children eligible for free and reduced lunch for the district as a whole and for the sample of DPP graduates.[72] 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 27: Free/Reduced Lunch Status of Cohort 1 DPP Graduates and Denver Public Schools in Total, Fall 2010



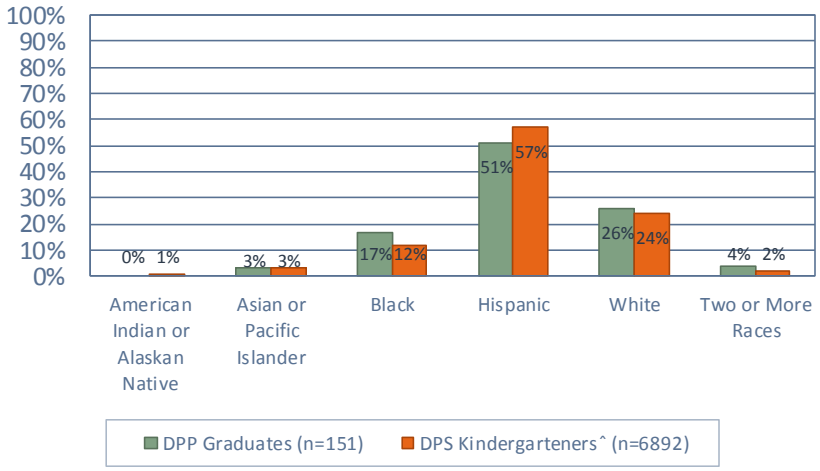
^The group All DPS includes the DPP graduates.

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.

COHORT 2

Figure 28 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 kindergarten in DPS as of fall 2010. The racial and ethnic compositions of the two groups are remarkably similar.

Figure 28: Race/Ethnicity of Cohort 2 DPP Graduates and All Kindergarteners Enrolled in DPS in Fall 2010*



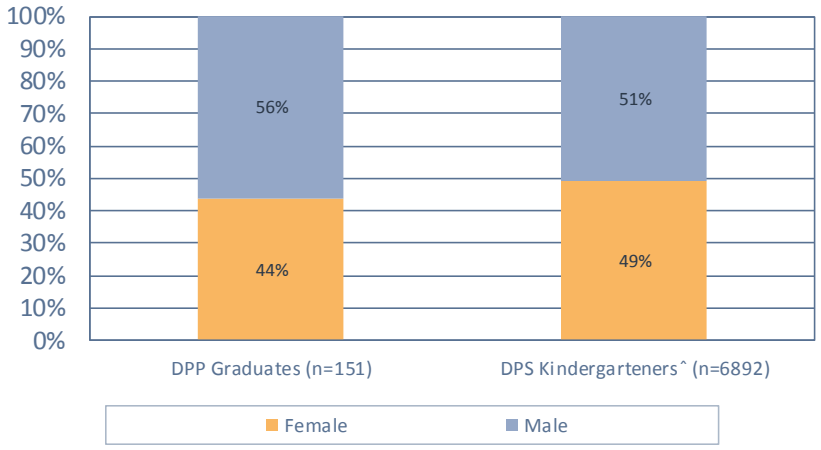
*This figure includes only Cohort 2 DPP Graduates for whom reading assessment data were available for the 10-11 school year. Percentages for DPP Graduates are weighted and sum to 101 because of rounding error.
^The group of DPS Kindergarteners includes the DPP graduates.

Figure 29 presents the gender composition of Cohort 2 DPP graduates with reading assessment data and all kindergarteners in DPS as of fall 2010. The district as a whole was split very close to evenly between

[72] 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.

the genders. Cohort 2 DPP Graduates with spring 2011 reading assessment data were split a bit less evenly, with slightly fewer girls than boys, but still rather close to an even split.

*Figure 29: Gender of Cohort 2 DPP Graduates and All Kindergarteners Enrolled in DPS in Fall 2010**

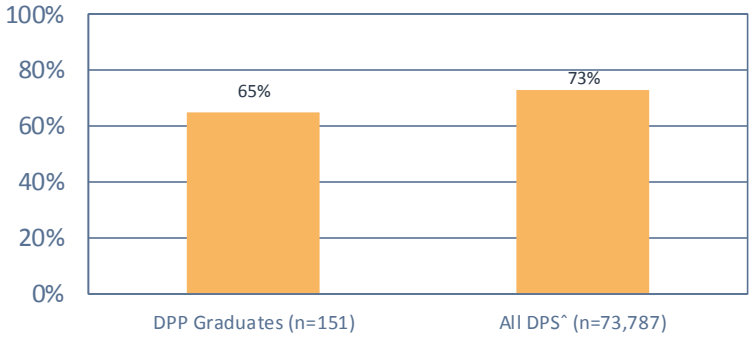


*This figure includes only Cohort 2 DPP Graduates for whom reading assessment data were available for the 10-11 school year.
^The group of DPS Kindergarteners includes the DPP graduates.

Finally, Figure 30 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 2011 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 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.

Figure 30: Free/Reduced Lunch Status of Cohort 2 DPP Graduates and Denver Public Schools in Total, Fall 2010



^The group All DPS includes the DPP graduates.

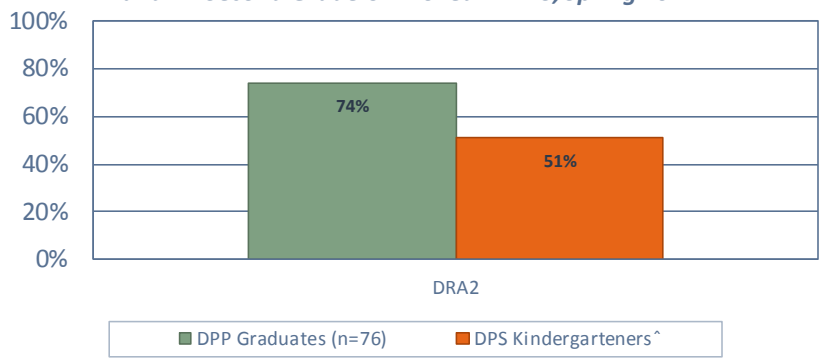
WHAT IS THE OVERALL READING PROFICIENCY OF DPP GRADUATES IN THE EARLY ELEMENTARY YEARS? HOW DOES THIS COMPARE TO THE DISTRICT AS A WHOLE?

To address this research question, we examined the proportion of DPP graduates who were reading on grade level as measured by the DRA2 and EDL2 alongside statistics for the district as a whole. We focused our analyses on children who were enrolled in the expected grades (i.e., second grade for Cohort 0, first grade for Cohort 1, and kindergarten for Cohort 2). 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 31: Proportion of Children Reading On Grade Level in the Spring of the Second Grade Year, Cohort 0 DPP Graduates and All Second Graders Enrolled in DPS, Spring 2011**



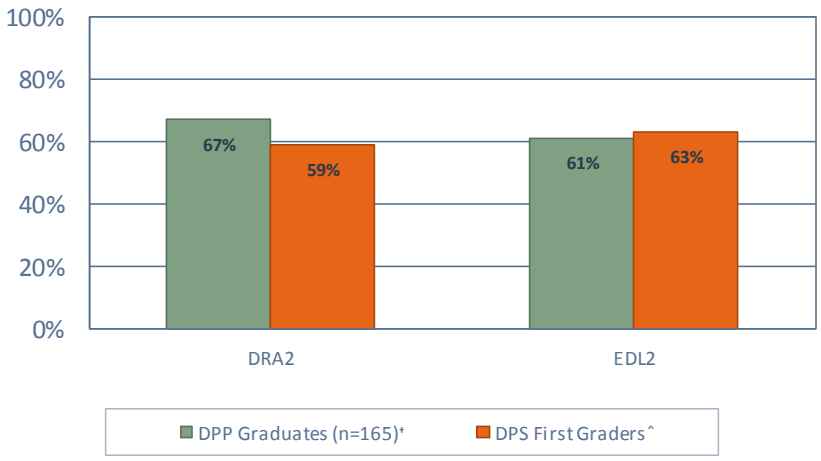
*A score of 28 is considered reading "on grade level" for the end of second grade.
^The group of DPS second graders includes the DPP graduates.

Figure 31 displays the proportion of Cohort 0 DPP graduates whose reading level was at or above grade level as assessed by the DRA2. No children in this cohort were assessed using the EDL2 in spring 2011. Three-quarters of Cohort 0 DPP graduates were reading on grade level in the spring of the kindergarten year. This is compared to about half of all second graders in the district overall. 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.

COHORT 1

Figure 32 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 first graders in the district as a whole in spring 2011. About two-thirds of DPP graduates assessed in English with the DRA2 were reading at or above grade level at the end of kindergarten. This slightly exceeded the proportion of children reading on grade level in the district as a whole. The proportions of children reading on grade level or above for children assessed in the Spanish using the EDL2 were similar for DPP Graduates and the district as a whole. In both cases, nearly two-thirds of children were reading at or above grade level.

*Figure 32: Proportion of Children Reading On Grade Level in the Spring of the First Grade Year, Cohort 1 DPP Graduates and All First Graders Enrolled in DPS, Spring 2011**

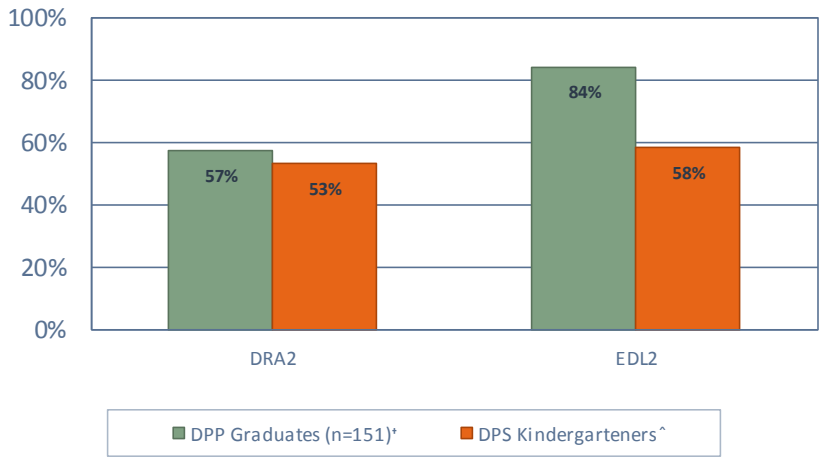


*A score of 16 is considered reading "on grade level" for the end of first grade.
†The group of DPP Graduates includes 137 children assessed with the DRA2 and 28 assessed with the EDL2.
^The group of DPS first graders includes the DPP graduates.

COHORT 2

Figure 33 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 kindergarteners in the district as a whole in spring 2011. Fifty-seven percent of DPP graduates assessed in English with the DRA2 were reading at or above grade level at the end of kindergarten.[73] This is slightly higher than the proportion reading at or above grade level in the district as a whole (53%). Eighty-four percent of DPP graduates assessed in Spanish using the EDL2 were reading at or above grade level at the end of kindergarten. In contrast, 58% of children in the district as a whole were reading at or above grade level as assessed by the EDL2.

Figure 33: Proportion of Children Reading On Grade Level in the Spring of Kindergarten Year, Cohort 2 DPP Graduates and All Kindergarteners Enrolled in DPS, Spring 2011*



*A score of 4 is considered "on grade level" for the end of kindergarten.
†The group of DPP Graduates includes 126 children assessed with the DRA2 and 25 assessed with the EDL2. Analyses are weighted.

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.

[73] When Cohort 1 was in kindergarten, 92% of DPP graduates assessed with the DRA2 were reading on grade level and 85% of DPP graduates assessed with the EDL2 were reading on grade level (reported in Klute, M. M. (2011). *Denver Preschool Program Annual Evaluation Report Addendum: Children's Performance After Leaving DPP, 2008-09 and 2009-10 School Years.* Unpublished Report. Denver: Clayton Early Learning Institute.). There appears to be a dramatic difference between results for Cohort 1 and the results for Cohort 2 reported in Figure 33. Part of the difference between the two cohorts is due to an increase in expectations for reading that started with the 2010-11 school year. Because of the increase in the proportion of children enrolled in full-day kindergarten, DPS raised its expectations for reading on grade level at the end of kindergarten from a score of 3 on the DRA to a score of 4 on the DRA. For the purposes of comparison across cohorts, we computed the percent of Cohort 2 DPP graduates reading on grade level using the old cutoff of 3. When we did so, 81% of Cohort 2 DPP graduates assessed with the DRA2 and 98% of Cohort 2 DPP graduates assessed with the EDL2 children met the benchmark. These figures are more similar to what was observed with Cohort 1.

COHORT 0

As with the previous question, our analysis focused only on the children who were in the expected grades. In spring 2011, 76 Cohort 0 children were enrolled in second grade and had reading assessment data. Of these, 66 were assessed only in English during the preschool year. All of these children were assessed in English at the end of the second grade year. Associations between the language and literacy preschool assessments and second grade DRA2 scores were strong and significant. PPVT-4 scores in preschool were correlated with second grade DRA2 scores at .49 ($p<.001$). Letter-Word Identification was correlated with DRA2 at .54 ($p<.0001$). Applied Problems was associated with DRA2 scores at .24 ($p<.10$). The magnitude of the associations of DRA2 with PPVT-4 and Letter-Word Identification were similar to those observed at the end of first grade year.[74] The association between DRA2 and Applied Problems was of smaller magnitude than in first grade.

Ten children with second 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 in English using the DRA2 at the end of first 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 reading scores at the end of second grade. The association was strongest for the preschool literacy assessment, which measures skills that are similar to what the DRA2 measures. The subgroup of children identified as Spanish speakers in this cohort was not large enough to examine.

COHORT 1

In spring 2011, 165 children were enrolled in first grade and had reading assessment data. Ninety-seven of these children were assessed only in English in preschool and were assessed with the DRA2 in the first grade year. Associations between the preschool assessments and first grade DRA2 scores were strong and significant. PPVT-4 scores in preschool were correlated with first grade DRA2 scores at .57 ($p<.0001$). Letter-Word Identification was correlated with DRA2 at .55 ($p<.0001$). Applied Problems was associated with DRA2 scores at .50 ($p<.001$). The magnitude of the associations of DRA2 with PPVT-4 and Applied Problems was larger than what was observed in kindergarten.[75] The association between first grade DRA2 and Letter-Word Identification was similar to what was observed in kindergarten.

Sixty-eight Cohort 1 children with first grade reading assessment data in spring 2011 had been assessed in both English and Spanish during their preschool year. Forty of these children were assessed in English using the DRA2 at the end of first grade. Twenty-eight 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.

---

[74] For more information about the analyses of Cohort 0 for the kindergarten and first grade years, readers are referred to a previously prepared report. Klute, M. M. (2011). *Denver Preschool Program Annual Evaluation Report Addendum: Children's Performance After Leaving DPP, 2008-09 and 2009-10 School Years.* Unpublished Report. Denver: Clayton Early Learning Institute.

[75] For more information about the analyses of Cohort 0 for the kindergarten and first grade years, readers are referred to a previously prepared report. Klute, M. M. (2011). *Denver Preschool Program Annual Evaluation Report Addendum: Children's Performance After Leaving DPP, 2008-09 and 2009-10 School Years.* Unpublished Report. Denver: Clayton Early Learning Institute.

For children assessed in English in first grade, there were no significant associations between assessments in preschool in English and first grade DRA2 scores. For Spanish preschool assessments, there was a significant association between first grade DRA2 and Applied Problems (r=.38, p<.05). There was also a trend toward significance for the correlation between Letter-Word Identification scores and DRA2 scores (r=.28, p<.10). This is a different pattern of results than was observed when these children were in kindergarten. Last year, there were no significant associations between DRA2 scores and preschool assessments.

For children assessed in Spanish in first grade, EDL2 scores were significantly associated with nearly all of the preschool assessments, in both English and Spanish. For the English preschool assessments, PPVT-4 was correlated .39 with EDL2 (p<.05); Letter-Word Identification was correlated with EDL2 at .49 (p<.01) and Applied Problems was correlated with EDL2 at .44 (p<.05). For the Spanish preschool assessments, TVIP was correlated with EDL2 at .38 (p<.05) and Applied Problems was correlated with EDL2 at .53 (p<.01). There was also a trend toward a significant association between EDL2 scores and Letter-Word Identification (r=.32, p<.10).

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 first 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 first grade year, presumably children who were judged by their first grade teachers to have stronger English skills, there was not a strong pattern of associations between kindergarten readiness, assessed in English and children's reading scores at the end of first grade. However, their preschool math abilities, and to some extent, their preschool literacy skills assessed in Spanish were associated with first grade reading assessed in English. For children who were assessed in Spanish at the end of the first grade year, there was evidence for an association between kindergarten readiness assessed in both Spanish and English and the reading skills assessed in Spanish at the end of the first grade year.

COHORT 2

In spring 2011, 151 children were enrolled in kindergarten 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 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 .58 (p<.0001). Letter-Word Identification was correlated with DRA2 at .74 (p<.0001). Applied Problems was associated with DRA2 scores at .52 (p<.001). The magnitude of these associations was larger than associations observed for Cohort 1 in the kindergarten year.

Forty percent of Cohort 2 children with kindergarten reading assessment data in spring 2011 had been assessed in both English and Spanish during their preschool year. Fifty-nine percent of these children were assessed in English using the DRA2 at the end of kindergarten. Forty-one 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 only one significant association between assessments in preschool and kindergarten DRA2 scores. Letter-Word Identification, assessed in Spanish, was associated with kindergarten DRA2 at .50 (p<.01). For children assessed in Spanish in kindergarten there were no significant associations.

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 not a strong pattern of associations between kindergarten readiness and reading skill at the end of kindergarten as assessed in either Spanish or English.

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

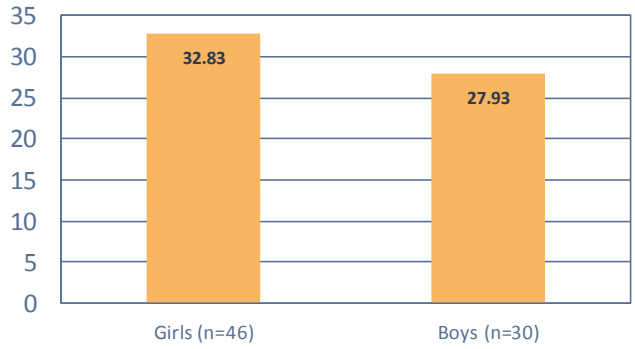
To address this question, we compared mean DRA2 and EDL2 reading levels for children belonging to different demographic subgroups. For Cohort 0, we examined child gender, race/ethnicity, home language and child primary language. We were unable to examine family income because income information for Cohort 0 was unreliable. For Cohorts 1 and 2, we examined child gender, race/ethnicity, home language, child primary language, income tier and region of the city.

COHORT 0

ENGLISH READING ASSESSMENT (DRA2)

In spring 2011, among Cohort 0 DPP graduates, there was a significant difference between boys and girls in second grade DRA2 scores (see Figure 34).[76] Girls, on average, exceeded grade level expectations for reading at the end of second grade,[77] while boys, were on average, reading right at grade level. This is similar to the gender differences observed for this cohort in first grade.

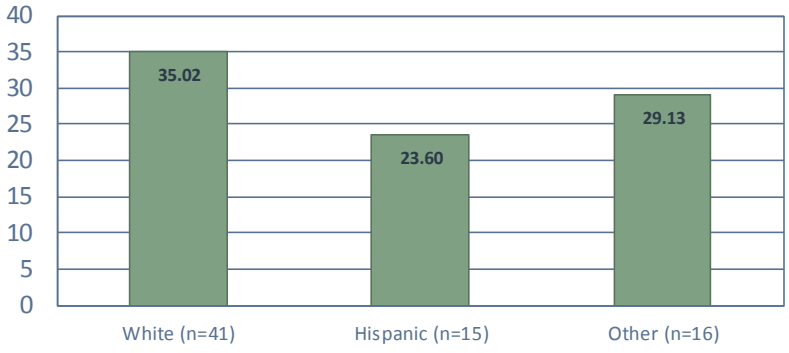
Figure 34: Second Grade Reading Level (DRA2) by Child Gender, Cohort 0, Spring 2011*



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

There were also significant differences by race/ethnicity (see Figure 35).[78] Follow-up Tukey tests revealed that white children had significantly higher scores than Hispanic children. White children had average DRA2 scores that approached the benchmark for grade 3 reading.[79]

Figure 35: Second Grade Reading Level (DRA2) by Race/Ethncity, Cohort 0, Spring 2011*



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

[76] t=2.19, df=74, p<.05

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

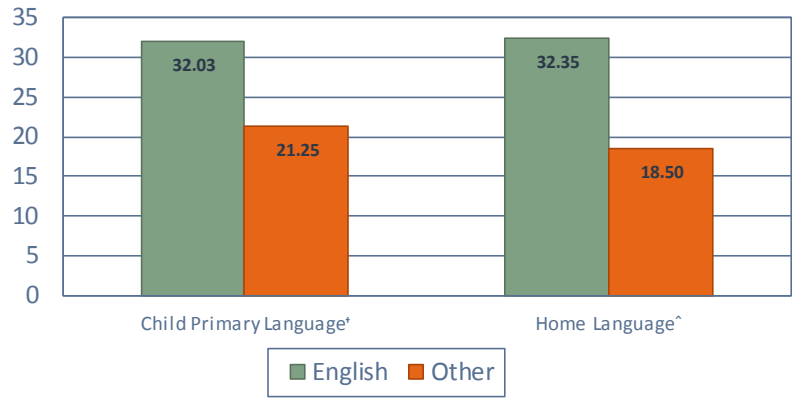
[78] F(2,71)=10.17, p<.0001.

[79] A DRA2 score of 38 is considered "on grade level" for the end of third grade.

Hispanic children's average score fell short of grade level expectations. Children in the "other" category were reading at grade level, on average.

When we examined child primary language and home language, we found that both of these variables were significantly associated with second grade DRA2 scores (see Figure 36).[80] Children who spoke English as their primary language scored about 11 points higher, on average than their counterparts that had another primary language. A similar pattern was observed for home lanugage, but with a larger gap of 14 points. These differences are large, considering that the increase in grade level expectations from year to year is 10-12 points. 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 36: Second Grade Reading Level (DRA2) by Child Primary Language and Home Language, Cohort 0, Spring 2011*



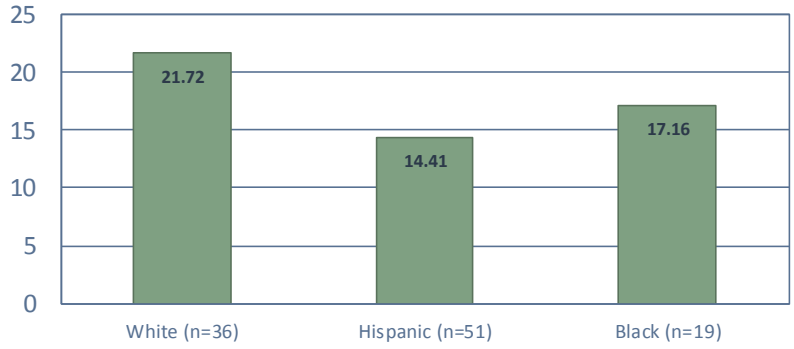
*A score of 28 is considered reading "on grade level" for the end of second grade.
*68 children had English identified as their primary language, 8 had another language.
^68 children had English identified as the home language, 8 had another language.

COHORT 1

ENGLISH READING ASSESSMENT (DRA2)

In spring 2010, there was not a significant difference between boys and girls on the DRA2.[81] There was, however, a difference in DRA2 scores by race/ethnicity (see Figure 37).[82] Follow-up Tukey tests revealed that white children, whose scores, on average, exceeded grade level expectations, scored

Figure 37: First Grade Reading Level (DRA2) by Race/Ethnicity, Cohort 1, Spring 2011*



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

[80] Child Primary Language: t=3.13, df=74, p<.01; Home Language: t=4.20, df=74, p<.0001.

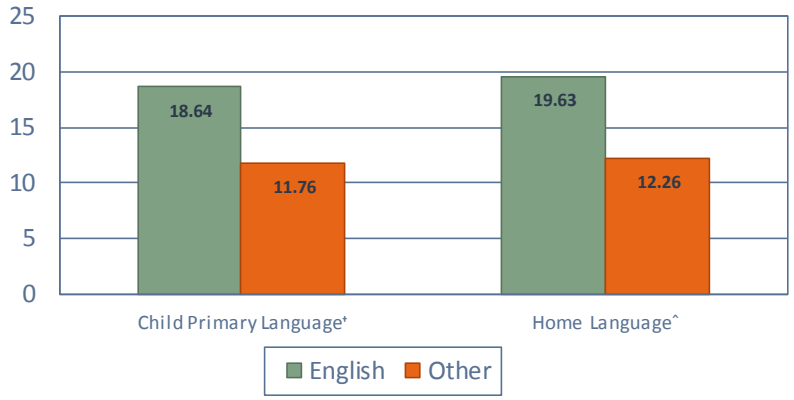
[81] t=1.12, df=135, n.s.

[82] This analysis omitted 3 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. $F(2,105)=10.68$, p<.0001

significantly higher than Hispanic children, whose average score fell short of grade level expectations.[83] Black children's scores, on average, slightly exceeded grade level expectations.

DRA2 scores also differed by children's primary language and home language (see Figure 38).[84] Children whose primary language was English scored significantly higher than children with another primary language. A similar pattern was observed for home language.

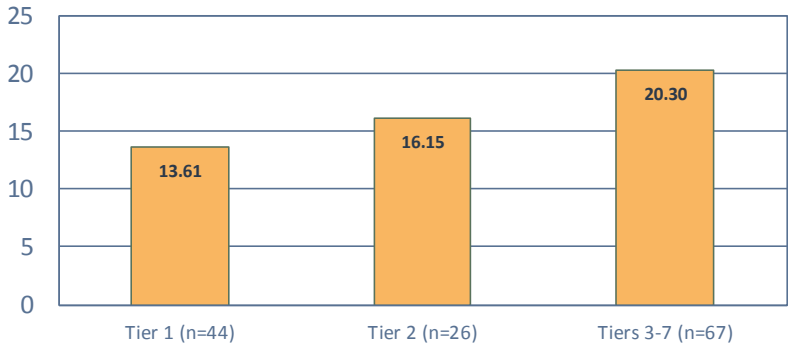
*Figure 38: First Grade Reading Level (DRA2) by Child Primary Language and Home Language, Cohort 1, Spring 2011**



*A score of 16 is considered reading "on grade level" for the end of first grade.
†111 children had English identified as their primary language, 25 had another language.
^87 children had English identified as the home language, 27 had another language.

There was also a significant difference by income tier (see Figure 39).[85] 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 both of the other income tier groups. The difference between Tier 1 and Tier 2 was not statistically significant.

*Figure 39: First Grade Reading Level (DRA2) by Income Tier, Cohort 1, Spring 2011**



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

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,[86] 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

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

[84] Child Primary Language: t=4.68, df=112, p<.0001; Home Language: t=4.29, df=134, p<.0001.

[85] 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. $F(2,136)$=12.25, p<.0001.

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

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.

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.[87]

SPANISH READING ASSESSMENT (EDL2)

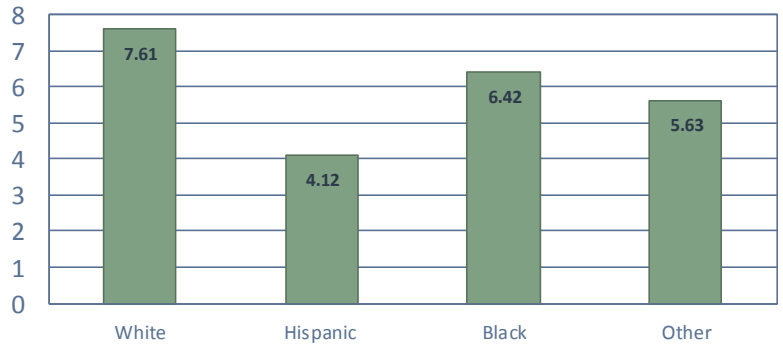
As with the English reading assessment, there was not a significant difference in EDL2 scores by child gender.[88] We were unable to test for differences by race/ethnicity because virtually all of the children assessed in Spanish using the EDL2 were Hispanic. Similarly, we were unable to test for differences 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.[89]

COHORT 2

ENGLISH READING ASSESSMENT (DRA2)

In spring 2011, there was not a significant difference between boys and girls on the DRA2.[90] There was, however, a difference in DRA2 scores by race/ethnicity (see Figure 40).[91] All four groups were reading, on average, above grade level.[92] However, white children had scores that were significantly higher than Hispanic children, on average.

Figure 40: Kindergarten Reading Level (DRA2) by Race/Ethnicity, Cohort 2, Spring 2011*



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

[87] F(4,135)=1.12, n.s.

[88] t=0.52, df=26, n.s.

[89] Income tier: F(2,27)=0.40, n.s.; region of the city: F(4,27)=2.22, n.s.

[90] t=0.98, df=114, n.s.

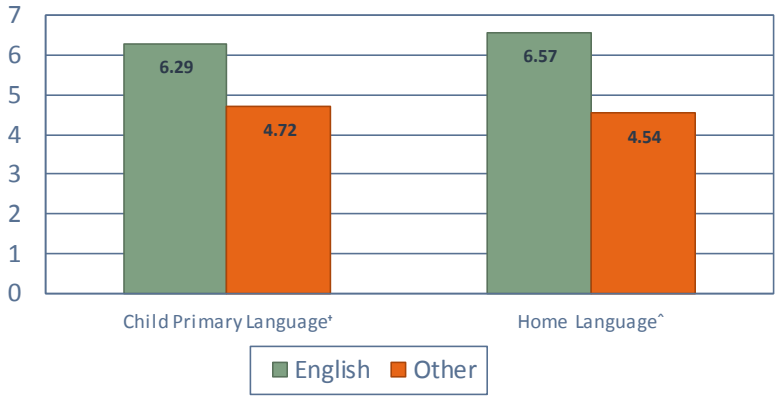
[91] F(3,115)=3.85, p<.05

[92] Starting with the 2010-11 school year, a DRA2 score of 4 is considered "on grade level" for the end of kindergarten.

DRA2 scores differed by children's home language, but not their primary language (see Figure 41).[93] Both home language groups were reading at or above grade level on average. However, children whose home language was English scored significantly higher than children with another home language.

There was also a significant difference by income tier (see Figure 42).[94] While all three 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 3-7 had significantly higher reading levels than children in Tier 1. The other pairwise comparisons were not statistically significant.

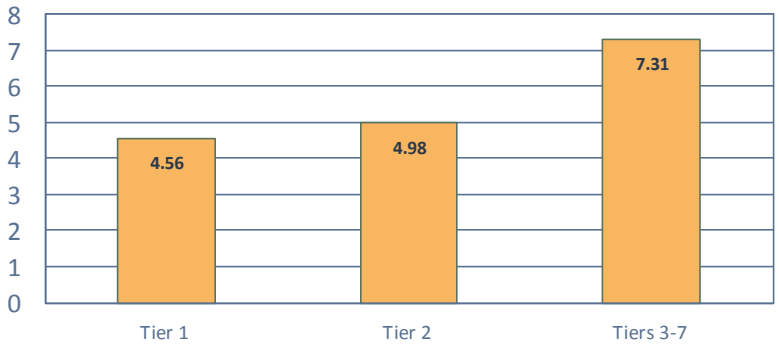
Figure 41: Kindergarten Reading Level (DRA2) by Child Primary Language and Home Language, Cohort 2, Spring 2011*



*A score of 4 is considered "on grade level" for the end of kindergarten.
†85 children had English identified as their primary language, 26 had another language.
^82 children had English identified as the home language, 29 had another language.

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.[95] 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.

Figure 42: Kindergarten Reading Level (DRA2) by Income Tier, Cohort 2, Spring 2011*



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

[93] Child Primary Language: t=1.48, df=109, n.s.; Home Language: t=2.03, df=109, p<.05.

[94] 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. $F(2,115)=4.23$, p<.05.

[95] 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.

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.[96]

SPANISH READING ASSESSMENT (EDL2)

As with the English reading assessment, there was not a significant difference in EDL2 scores by child gender.[97] 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.[98]

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=45) with children who had been enrolled in DPS preschools (n=31) on second grade reading assessments in English (DRA2) administered during the spring of 2011. There was not a significant difference between these two groups.[99]

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.[100]

Twenty-eight Cohort 1 DPP graduates were assessed in Spanish with the EDL2 at the end of kindergarten. 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.

---

[96] $F(4,115)=1.12$, n.s.

[97] $t=0.50$, df=16, n.s.

[98] Income tier: $F(2,17)=0.02$, n.s.; region of the city: $F(4,17)=0.57$, n.s.

[99] $t=0.10$, df=74, n.s.

[100] $t=0.78$, df=121, n.s.

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=65) on kindergarten reading assessments in English (DRA2) administered during the spring of 2011. There was not a significant difference between these two groups.[101] 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=3).

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 and 2, 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 and 09-10,[102] there was very limited variability in the star rating of the preschools attended by Cohort 1 and Cohort 2 DPP children. 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). 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.

COHORT 1

When we examined the distribution of star rating for children who had first grade DRA2 data in spring 2011, we found that, similar to the preschool year, only 4% had been enrolled in preschools with less than a star 3 rating. Seventy-one percent of children had been enrolled in star 3 preschools and 25% had been enrolled in star 4 preschools. There was no difference between these groups on their first grade DRA2 scores.[103] When we examined the distribution of star rating for children with first grade EDL2 data, we found that 75% of them had been enrolled in star 3 preschools. 22% 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 first 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.[104]

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. In all cases, the correlation between the measure of quality and the reading assessment score was non-significant.

---

[101] t=0.39, df=114, n.s.

[102] 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.

[103] $F(2, 135)$=0.96, n.s.

[104] $F(1, 26)$=0.64, n.s.

COHORT 2

When we examined the distribution of star rating for children who had first grade DRA2 data in spring 2011, we found that, similar to the preschool year, only 8% had been enrolled in preschools with less than a star 3 rating. Seventy percent of children had been enrolled in star 3 preschools and 22% had been enrolled in star 4 preschools. There was no difference between these groups on their first grade DRA2 scores.[105] When we examined the distribution of star rating for children with first grade EDL2 data, we found that 81% of them had been enrolled in star 3 preschools. 13% 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 first 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.[106]

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. In all cases, the correlation between the measure of quality and the reading assessment score was non-significant.

[105] $F(2, 104)=3.76$, n.s.

[106] $F(1, 16)=1.62$, n.s.

CONCLUSIONS

Clayton Early Learning Institute's 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 CSAP)?

As described above, none of the children who participated in DPP are old enough to take the CSAP, which begins in third grade. Children who participated in the program during DPP's first year of operation were in second grade during the 2010-11 school year. Because of this, our analyses for questions 4 and 5 focused on reading assessments administered in kindergarten, first grade and second grade.

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 Spanish vocabulary and literacy skills 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.

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 for 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: about 60% for vocabulary and about three-quarters for literacy and math. When teachers rated children's behaviors, their ratings of protective factors were high for most children. Protective factors were rated as an area of concern by teachers for 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 fewer than 10% of children. Parents identified protective factors as an area of concern for about one-sixth of children and behavioral concerns as an area of concern for about a quarter 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 are right in line with that threshold. Parents identified behavioral concerns as an area of concern for about 10% 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 40% 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 40% 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 all three academic assessments in English, but there was no association for the assessments in Spanish. The significant contrasts varied somewhat between the assessments, but the general pattern was that children from Tiers 1 and/or 2 tended to increase more rapidly than children in tiers 3-6. Finally, with respect to teachers' ratings of Attachment, children in Tier 1 increased significantly more than children in Tiers 3-6, who are likely to be from higher income families. It is noteworthy that income tier was not associated with the other 4 socio-emotional scores, resulting in a fairly weak pattern of results for this area overall.

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 academic assessments. There was also a fairly consistent pattern of effects for the socio-emotional scores. For three out of the four positive socioemotional variables, children whose primary language was English started the

year lower and increased more over time than their primarily English-speaking counterparts. This difference was not observed for teachers' reports of negative behaviors.

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 and 2 were demographically similar to the populations of children in 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.

Cohort 1 children were compared to the population of first graders in DPP. Among children whose reading was assessed in English in first 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 similar to the district as a whole.

A different pattern was observed for the group of DPP graduates who were enrolled in kindergarten during the 10-11 school year. Among children whose reading was assessed in English in kindergarten, the proportion of DPP graduates who were reading at or above grade level slightly exceeded 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 about 1.5 times more likely to be reading at grade level than the district as a whole.

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

There was not strong evidence for an association between classroom quality and children's kindergarten readiness in the academic domains. For socioemotional development, however, we found that children tend to leave for kindergarten displaying more positive behaviors and fewer negative behaviors when they had been enrolled in classrooms with higher levels of Instructional Support.

Our ability to examine quality in conjunction with later academic success 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 any evidence for an association between preschool quality and reading skill in kindergarten or first grade.

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 Spanish at a rate which exceeded what would be expected simply because of maturation. Their growth in vocabulary progressed at a rate that was similar to the average growth in the population at large. Children demonstrated positive changes in their socio-emotional functioning over time; teachers reported that children demonstrated more positive behaviors at the end of the school year than at the beginning. Growth in academic domains was most pronounced among children who came from lower income tiers (defined by family income and family size) and those whose primary language was not English. Children in these groups tended to start off their preschool year with lower scores, but make

larger gains over time, making progress toward closing the achievement gap. The similarity of findings for these two subgroups is largely due to the fact that income tier and primary language are strongly associated in this sample, making it impossible to disentangle the effects of these two variables. In terms of socio-emotional functioning over time, the growth that was observed was most pronounced among children whose primary language was not English.

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 kindergarten and first grade reading skills for these cohorts of children. We did not find any significant 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 quality and child outcomes. One of the domains, Instructional Support, was associated with socioemotional development but there were not strong associations with academic outcomes. Next year, when this cohort of children is in kindergarten, we will be able to examine this measure of preschool quality in conjunction with reading scores.

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 greater reading proficiency in kindergarten and first grade than the district as a whole. 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 2011

	Entire Sample, weighted[1]	By Provider Type, Unweighted		
Characteristic		Community	DPS	Significance of Difference by Provider Type
Sex				χ^2_1=.01; n.s.
Female	50.9%	50.5%	51.0%	
Male	49.1%	49.5%	49.0%	
Ethnicity				χ^2_5=28.99; p<.0001
Hispanic	53.8%	26.3%	60.0%	
White (not of Hispanic origin)	29.9%	51.5%	25.0%	
African-American (not of Hispanic origin)	8.2%	9.1%	8.0%	
Asian/Pacific Islander	3.2%	4.0%	3.0%	
Multi-Racial	3.3%	9.1%	2.0%	
Native	1.6%	0.0%	2.0%	
Child's Primary Language				χ^2_1=34.54; p<.0001
English	54.4%	86.9%	47.0%	
Another Language	39.4%	10.1%	46.0%	
Not Reported	6.3%	3.0%	7.0%	
Home Language				χ^2_1=35.35; p<.0001
English	53.2%	80.8%	47.0%	
Another Language	37.2%	7.1%	44.0%	
Not Reported	9.6%	12.1%	9.0%	
DPP Income Tier[2]				χ^2_5=22.72; p<.001
Tier 1	36.4%	20.2%	40.0%	
Tier 2	25.6%	15.2%	28.0%	
Tier 3	7.7%	15.2%	6.0%	
Tier 4	8.4%	10.1%	8.0%	
Tier 5	15.3%	25.3%	13.0%	
Tier 6—Income Not Reported	6.7%	14.1%	5.0%	
Star Level of Preschool				χ^2_2=13.84; p<.01
Star 2	4.2%	5.1%	4.0%	
Star 3	67.3%	46.5%	72.0%	
Star 4	28.5%	48.5%	24.0%	
Region of the City				χ^2_4=8.60; n.s.
Central	14.9%	23.2%	13.0%	
Northeast	24.9%	20.2%	26.0%	
Northwest	21.4%	23.2%	21.0%	
Southeast	10.1%	15.2%	9.0%	
Southwest	28.6%	18.2%	31.0%	

[1]The weighted sample results are representative of the population of children enrolled in DPP in Fall 2010.

[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
9	$40,750
If more than 9 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
9	$40,751	$75,388
If more than 9 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
9	$75,389	$91,688
If more than 9 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
9	$91,689	$122,250
If more than 9 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
9	$122,250
If more than 9 family members	Add $11,220 for each additional family member