



Denver Preschool Program: Report on Child Outcomes

2009-10 School Year

Prepared for the Denver Preschool Program
By Mary Maguire Klute, Ph.D.
Clayton Early Learning Institute
November 29, 2010

CONTENTS

Executive Summary	3
Introduction	6
Methods	7
Sample	7
Procedures	14
Measures	14
Results	16
Preliminary Analyses.....	16
Kindergarten Readiness	21
Change in Assessments over the preschool year.....	24
Conclusions	36
Question 1: Do children make progress in their development while in DPP early childhood environments?	36
Question 2: To what extent and in what areas are children enrolled in DPP ready for kindergarten?.....	36
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?	37
Question 5: Is attendance at higher-rated preschool programs associated with greater kindergarten readiness?.....	38
Summary and Future Directions	38
Appendix.....	40

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 in 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, literacy and math skills. Spanish-speaking children also made progress in their Spanish literacy and math skills (assessed in Spanish) over the course of their preschool year. The gains observed were above and beyond what would be expected based on normal development. Progress was observed in socio-emotional development as well. Over the course of the preschool year, teachers reported that children demonstrated significantly more protective factors (i.e., initiative, self-control, and attachment) and fewer behavioral concerns.

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

Results of the evaluation suggest that the vast majority of children are ready for school, both academically and socio-emotionally. When considering skills assessed in English and Spanish, where appropriate, we concluded that relatively few children had scores in the risk range on assessments of their vocabulary, literacy and math skills. Further, more children than would be expected scored at or above the average on these assessments. Results were particularly striking for the literacy and math assessments, where nearly 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 10% 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 15% of children.

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, over 80% of children whose primary language was not English were from the lowest two income tiers as compared with about half of children whose primary language is English. As a result, it is impossible to disentangle the effects of income and primary language. Any associations that are observed are likely associated with the co-occurrence of these two factors.

Children 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. The weak pattern of findings for our socio-emotional assessment supports the conclusion that children are making similar progress during their DPP year, regardless of their primary language or families' income.

IS ATTENDANCE AT HIGHER-RATED PRESCHOOL PROGRAMS ASSOCIATED WITH GREATER KINDERGARTEN READINESS?

There was not strong evidence that the star rating was associated with children's progress over time. However, the vast majority of children in the evaluation sample were enrolled in 3 star or 4 star preschools. Only 7% of children were enrolled in lower quality preschools. This greatly limits our ability to address this question using the star rating, as it effectively reduces the question to one of whether outstanding preschools are associated with greater kindergarten readiness than very good preschools. However, the Qualistar rating is comprised of 5 component areas, three of which had greater variability than the star rating categories: ratio/group size points, training and education points and learning environment. In these analyses, an unexpected, but fairly consistent pattern emerged. Being in a higher-quality program tended to be associated with lower scores on assessments administered in Spanish. This finding echoes an isolated finding from last year that higher learning environment scores were associated with lower gains in Spanish vocabulary scores. It may be the case that some of the higher scoring preschools may do a good job of teaching children in English, but have less of an emphasis on supporting children's Spanish language skills.

SUMMARY AND FUTURE DIRECTIONS

This evaluation described children's progress during the course of their DPP preschool year. In general, children progressed in their academic skills (vocabulary, literacy and math skills) as assessed in English at a rate which exceeded what would be expected simply because of maturation. Spanish-speaking children made significant gains in their literacy skills and math skills (assessed in Spanish) but 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 and fewer behavior problems at the end of the school year than at the beginning.

When change over time was examined for subgroups of children, we observed that growth in vocabulary, literacy and math skills was most pronounced among children from lower income tiers and those whose primary

language was not English. Children in these groups tended to start off the year lower than their counterparts in other groups, but make greater gains over the course of the year. The similarity of findings for income tier and child's primary language is not surprising, as these two variables were strongly associated in this sample. That is, having a primary language other than English tended to co-occur with poverty in this sample, making it impossible to disentangle the effects of these two factors. In terms of socio-emotional functioning over time, the growth that was observed was fairly consistent across subgroups. In addition, nearly all preschools attended by children in this sample were of relatively high quality (i.e., star 3 and 4) greatly limiting our ability to adequately test the association between preschool quality and child outcomes over time. However, there was an unexpected pattern of results suggesting that Spanish-speaking children have fewer gains over time in their academic skills, assessed in Spanish, in higher-quality programs as compared with lower-quality programs. We have proposed additional data collection on DPP classrooms for the upcoming school year, which will hopefully shed light on the reasons for this unexpected pattern of results.

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. Children from higher-risk groups tended to make progress toward closing the achievement gap that was present at the beginning of the preschool year. These early results are promising. Results from future years of this annual evaluation will provide the opportunity to replicate these findings as well as follow children from this evaluation cohort into 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 in 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 beyond¹. This portion of the evaluation was designed to address five questions relevant to children's development while enrolled in DPP and beyond:

1. Do children make progress in their development while in DPP early childhood environments (i.e., language, literacy, mathematics, and social-emotional development)?
2. To what extent and in what areas are children enrolled in DPP ready for kindergarten?
3. Do children from different income levels and with different primary languages make similar progress in their development while in DPP early childhood environments?
4. Do children who received DPP tuition credits compare favorably with their demographic counterparts who did not receive DPP tuition credits on assessments administered by Denver Public Schools (DPS) in kindergarten?
5. Is attendance at higher-rated preschool programs associated with greater kindergarten readiness and long-term academic success (as measured by CSAP)?

Since this is just the third year of the DPP program, we are limited in our ability to address questions 4 and 5. This is the second year of full implementation of the evaluation design. The sample of children studied last year (enrolled in DPP during the 08-09 school year), has gone on to kindergarten. We are in the process of obtaining kindergarten assessment data from DPS. Once we receive the data, we will begin analyzing it and prepare a separate report that will begin to address question 4. With respect to question 5, this year we are able to provide results about how preschool program quality ratings are associated with kindergarten readiness, but not yet with long-term academic success.

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

SAMPLING PLAN

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 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.² In August 2009, a sample of 100 children enrolled in community sites was drawn from the group of families that volunteered to participate (henceforth referred to as “volunteers”). In September 2009, 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, DPP income tier, Qualistar rating of the preschool program, home language, child language, and region of the city in which the child lives. 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, significant differences were detected between volunteers and non-volunteers on three variables. First, a significant difference was detected for ethnicity.³ Follow-up analyses revealed that this effect was due to differences between volunteers and non-volunteers in the Hispanic ethnicity subgroup. Parents who identified their child’s ethnicity as Hispanic were significantly less likely to volunteer to be contacted about the evaluation than parents of children from other ethnic groups (57% of parents of Hispanic children volunteered vs. 68% of parents of children of other ethnic groups).⁴ Analyses also revealed significant differences between volunteers and non-volunteers on both home language and child primary language. Parents of children from homes where the primary language spoken was English were more likely to volunteer than parents of children who lived in homes where the primary language was something other than English (where home language was English,

² Information about the evaluation was provided on the DPP application, which was available in both English and Spanish.

³ $\chi^2_6=13.91, p<.05$

⁴ $\chi^2_1=9.68, p<.01$

67% volunteered to be contacted vs. other home language, 55% volunteered to be contacted).⁵ A similar pattern was observed for child's primary language (67% of parents of children with English primary language volunteered to be contacted vs. 55% of parents of children with another primary language).⁶ The similar pattern of findings for home language and child primary language was not surprising; for 98% of the sample, child primary language was the same as home language. To adjust for these differences, the sampling frame was stratified by Hispanic ethnicity and home language.⁷ The proportion of children drawn from each stratum was adjusted to match the proportions in the population of children enrolled at the time of sampling. The result was a sample of 100 that was representative of the community site population as a whole in August 2009 with respect to the variables examined⁸. The sample was drawn with replacement; if a selected child was deemed ineligible for the study,⁹ a selected family was unable to be contacted to obtain informed consent to participate in the study or if a selected family refused to participate in the study, a replacement child was randomly drawn from the same stratum.

DPS SITES

In DPS sites, significant differences were detected between volunteers and non-volunteers on only one variable. First, a significant difference was detected for region of the city where the child lived.¹⁰ Follow-up analyses revealed that this effect was due to differences between volunteers and non-volunteers in the southwest region of the city. Parents of children who lived in the southwest region of the city were significantly less likely to volunteer to be contacted about the evaluation than parents of children from other regions of the city (49% of parents of children from the southwest region volunteered vs. 55% of parents of children living in other regions of the city). To adjust for this difference, the sampling frame was stratified by whether or not the child lived in the southwest region. 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 2009 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.

⁵ $\chi^2_1=6.81, p<.01$

⁶ $\chi^2_1=6.44, p<.05$

⁷ It was not possible to stratify by both child primary language and home language because the association between these two variables was so strong. These variables were identical for 98% of children. We chose to stratify by home language because the strength of the association with volunteering to be contacted was stronger than it was for child primary language.

⁸ 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.

⁹ Typically children were deemed ineligible because they were no longer enrolled in a DPP preschool at the time the family was contacted for participation.

¹⁰ $\chi^2_4=10.54, p<.05$

SAMPLING WEIGHTS

A sample of 200 children was drawn during fall 2009 using the sampling procedure described above. At the time of sampling, 21.1% of children enrolled in DPP were attending community sites and the remaining 78.9% were attending DPS sites.¹¹ We drew a sample that included 50% children from community sites and 50% of children from DPS sites. After sampling was complete but prior to assessment, one child moved from a community site to a DPS site. As a result, the final sample included 99 children from community sites (49.5%) and 101 children from DPS sites (50.5%). 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 groups' 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 1.¹² 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 16 percent of the sample. Nearly two-thirds 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 2007. 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 7—income reported. About a tenth of families were assigned to the highest tier, tier 7, because they opted out of the requirement to provide their income.

Nearly all (93%) of the children were enrolled in preschools with a 3 or 4 star rating. Over two-thirds of children were enrolled in star 3 preschools and nearly a quarter of children were enrolled in star 4 preschools. About a quarter of children resided in the northeast and southwest regions of the city, about one-fifth lived in the northwest region. The smallest proportion of children lived in southeast Denver.

The right hand side of Table 1 presents demographic characteristics by provider type. There was a trend toward a significant difference in the distribution of the sex of the child in the two provider types. The sample of children from community sites included a slightly larger proportion of girls than the sample of children from DPS sites. 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 and white children. There was a much larger percentage of Hispanic children in DPS sites, nearly twice the magnitude of the proportion of Hispanic children in community sites.¹³ In contrast, DPS sites had a smaller proportion of white children enrolled than did community sites.¹⁴ Provider type was also significantly associated with both child primary language and home language.¹⁵ DPP

¹¹ 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.

¹² Sample characteristics for the spring sample, which were nearly identical, are presented in the appendix.

¹³ $\chi^2_1=14.61$; $p<.001$

¹⁴ $\chi^2_1=5.39$; $p<.05$

preschools tend to serve a population of children that is more diverse in terms of language. Forty percent of children in DPS sites have a primary language other than English compared with less than a quarter of children in community sites. Not surprisingly, a similar pattern was observed for home language.

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. Three-quarters of children in DPS sites were enrolled in star 3 programs as compared with slightly over half of children in community sites.¹⁶ About a third of children in community sites were enrolled in star 4 sites compared with about one-fifth of children in DPS sites.¹⁷ In sum, the vast majority of children in DPS sites were enrolled in star 3 programs, whereas there was more diversity in the quality rating of the programs in which children in community sites were enrolled, including more children in sites with the highest quality rating.

Finally, there was a significant association between provider type and region of the city. Follow-up analyses revealed that this was largely due to the distribution of children living in the southwest region of the city. Nearly a third of children enrolled in DPS sites lived in this region of the city compared with only 13% of children enrolled in community sites.¹⁸

We attempted to follow up with each of these children again in spring 2010. One child did not continue with the study because the child left the DPP preschool between the fall and spring rounds. This child was no longer enrolled in any DPP preschool, and was thus no longer eligible for the study. An alternate from the same strata as the child lost to follow-up was selected and assessed during the spring round.¹⁹ As a result, the total sample size for the 09-10 school year is 201.

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 2009-10 school year. Each set of analyses are described in turn below.

¹⁵ Child Language (English vs. Other Language): $\chi^2_1=9.80$; $p<.01$; Home Language (English vs. Other Language): $\chi^2_1=11.95$, $p<.001$

¹⁶ $\chi^2_1=8.82$, $p<.01$

¹⁷ $\chi^2_1=4.90$, $p<.05$

¹⁸ $\chi^2_1=9.87$, $p<.01$

¹⁹ We “refreshed” the sample in the spring to maintain the total sample size of 200. This was done because we wanted to ensure that we had a sample of at least 200 to follow into the elementary school years.

Table 1: Sample Characteristics Fall 2009

Characteristic	Entire Sample, weighted ¹	By Provider Type, Unweighted		
		Community	DPS	Significance of Difference by Provider Type
Sex				$\chi^2_1=2.91$; $p<.10$
Female	49.1%	58.6%	46.5%	
Male	50.9%	41.4%	53.5%	
Ethnicity				$\chi^2_5=18.48$; $p<.01$
Hispanic	50.9%	30.3%	56.4%	
White (not of Hispanic origin)	24.9%	36.4%	21.8%	
African-American (not of Hispanic origin)	15.8%	19.2%	14.9%	
Asian/Pacific Islander	2.8%	6.1%	2.0%	
Multi-Racial	4.1%	8.1%	3.0%	
Other	1.6%	0.0%	2.0%	
Child's Primary Language				$\chi^2_1=9.80$; $p<.01$
English	62.3%	76.8%	58.4%	
Another Language	34.3%	18.2%	38.6%	
Not Reported	3.4%	5.1%	3.0%	
Home Language				$\chi^2_1=11.95$; $p<.001$
English	58.5%	73.7%	54.5%	
Another Language	36.7%	18.2%	41.6%	
Not Reported	4.8%	8.1%	4.0%	
DPP Income Tier ²				$\chi^2_6=7.35$; ns.
Tier 1	44.7%	41.4%	45.5%	
Tier 2	20.8%	17.2%	21.8%	
Tier 3	4.0%	4.0%	4.0%	
Tier 4	0.2%	1.0%	0.0%	
Tier 5	5.9%	2.0%	6.9%	
Tier 6	0.0%	0.0%	0.0%	
Tier 7—Income Reported	13.6%	20.2%	11.9%	
Tier 7—Income Not Reported	10.8%	14.1%	9.9%	
Star Level of Preschool				$\chi^2_4=10.42$; $p<.05$
Preschool Not Yet Rated	0.2%	1.0%	0.0%	
Provisional	0.2%	1.0%	0.0%	
Star 1	0.0%	0.0%	0.0%	
Star 2	7.0%	11.1%	5.9%	
Star 3	69.9%	53.5%	74.3%	
Star 4	22.7%	33.3%	19.8%	
Region of the City				$\chi^2_4=11.59$; $p<.05$
Central	14.1%	22.2%	11.9%	
Northeast	27.1%	32.3%	25.7%	
Northwest	20.1%	21.2%	19.8%	
Southeast	10.9%	11.1%	10.9%	
Southwest	27.8%	13.1%	31.7%	

¹The weighted sample results are representative of the population of children enrolled in DPP in Fall 2009.

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

FALL 2009

COMMUNITY SAMPLE

Children who were included in the community sample were compared to 712 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 child gender, indicating that the proportion of boys and girls in the sample was significantly different than the proportion of boys and girls in the population overall.²⁰ In the population of DPP children enrolled in community sites, 49% of children were girls and 51% boys. In the sample, 59% were girls and 41% boys. 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.²¹ That is, the community sample overrepresented girls but was otherwise representative of the population of enrolled children in August 2009.

DPS SAMPLE

Children who were included in the DPS sample were compared to 3128 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.²² That is, the DPS sample was representative of the population of enrolled children in September 2009.

SUMMER 2010

COMMUNITY SAMPLE

Children who were included in the community sample were compared to 2203 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 child gender, indicating that the proportion of boys and girls in the sample was significantly different than the proportion of

²⁰ Gender: $\chi^2_1=4.23$, $p<.05$

²¹ Ethnicity: $\chi^2_6=5.42$, n.s.; income tier: $\chi^2_6=6.70$, n.s.; Qualistar rating: $\chi^2_6=5.98$, n.s.; home language: $\chi^2_1=.01$, n.s.; child primary language: $\chi^2_1=.01$, n.s.; region of the city: $\chi^2_4=7.59$, n.s.

²² Gender: $\chi^2_1=.35$, n.s.; Ethnicity: $\chi^2_6=5.00$, n.s.; income tier: $\chi^2_6=7.95$, n.s.; Qualistar rating: $\chi^2_2=4.28$, n.s.; home language: $\chi^2_1=.01$, n.s.; child primary language: $\chi^2_1=.23$, n.s.; region of the city: $\chi^2_4=1.71$, n.s.

boys and girls in the population overall.²³ In the population of DPP children enrolled in community sites, 49% of children were girls and 51% boys. In the sample, 59% were girls and 41% boys.

The sample also significantly differed from the population of children enrolled in community sites at the end of the school year in terms of the region of the city in which they lived.²⁴ Follow-up tests revealed that the children in the sample were significantly more likely to reside in the northeast region than children in the population as a whole. A third of children in the community sample lived in the northeast region compared with just 22% in the population of children enrolled in community sites at the end of the school year as a whole.²⁵ In contrast, children in the sample were significantly less likely to reside in the southwest region than children in the population as a whole. Thirteen percent of children in the community sample lived in the southwest region compared to 26% of children in the population of children enrolled in community sites at the end of the school year as a whole.²⁶ These effects are largely due to a change in the distribution of children's residences across the city over the course of the school year. 1384 additional children enrolled in community sites between the time of sampling and the end of the school year. A large proportion of these children (30%) lived in the southwest region, changing the distribution across the city.

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, the community sample overrepresented girls, overrepresented children from the northeast region of the city, and underrepresented children from the southwest region of the city, but was otherwise representative of the population of enrolled children at the end of the 2009-10 school year.

DPS SAMPLE

Children who were included in the DPS sample were compared to 3715 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.

²³ $\chi^2_1=3.87, p<.05$

²⁴ $\chi^2_4=16.83, p<.01$

²⁵ $\chi^2_1=8.43, p<.01$

²⁶ $\chi^2_1=8.72, p<.01$

²⁷ Ethnicity: $\chi^2_6=7.12, n.s.$; income tier: $\chi^2_6=1.04, n.s.$; Qualistar rating: $\chi^2_6=3.20, n.s.$; home language: $\chi^2_1=.33, n.s.$; child primary language: $\chi^2_1=.34, n.s.$

²⁸ Gender: $\chi^2_1=.25, n.s.$; Ethnicity: $\chi^2_6=4.08, n.s.$; income tier: $\chi^2_6=6.40, n.s.$; Qualistar rating: $\chi^2_3=2.59, n.s.$; home language: $\chi^2_1=.00, n.s.$; child primary language: $\chi^2_1=.25, n.s.$; region of the city: $\chi^2_4=1.73, n.s.$

PROCEDURES

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

Parents were mailed a survey about their children’s socio-emotional development in January 2010. 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 2 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, and a substantial improvement over the response rate from last two years.

Table 2: Sample sizes by data collection type, Fall 2009 and Spring 2010

Data Collection Activity	Fall 2009	Spring 2010
Standardized Assessments—English	200	200
Standardized Assessments—Spanish	63	63
DECA—Teacher Report	194 (97%)	193 (97%)
DECA—Parent Report	N/A	185 (93%)

MEASURES

ARCHIVAL DATA

Information about demographic characteristics was obtained from Affiliated Computer Services (ACS), the contractor that handles enrollment and tuition payments for the Denver Preschool Program. Information about program quality was obtained from Qualistar Early Learning, which is responsible for conducting quality ratings of sites.

STANDARDIZED ASSESSMENTS OF CHILDREN

Children were assessed using a battery of standardized assessments (see Table 3). 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

Table 3: 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

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.

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.

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 3). The DECA is a 37-item measure with four subscales, 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 this case, scores were only computed if at least 75% of the items on the scale were completed.

RESULTS

PRELIMINARY ANALYSES

Table 4 presents preliminary results 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. Scores for the PPVT and TVIP tended to be lower than those for the WJ and WM. It is noteworthy that for all of these assessments, there is considerable variability in children's scores, with some children scoring quite low and some scoring rather high.

The DECA is scaled using T-scores, which have a mean of 50 and a standard deviation of 10. In both the fall and spring, teachers rated children, on average, very 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 slightly lower scores on Attachment. Once again there was substantial variability in all of the scores.

Since all children were assessed in English, regardless of their primary language, it is useful to consider whether children's scores on the English assessments differed if children spoke English as their primary language or not. 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 5. 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 over 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 6). For this round, once again, all three differences were statistically significant. Similar to the fall, the largest difference between the primary language groups was observed for the PPVT, about two standard deviations in magnitude. Differences between primary language groups for LWI and AP were slightly smaller than the fall, but still statistically significant. For

LWI, the difference between language groups was nearly one standard deviation in magnitude. For AP, the difference between the groups was about two-thirds of a standard deviation.

Table 4: Weighted Descriptive Statistics for Child Outcome Measures

Variable	Fall 2009				Spring 2010			
	N	Mean	Standard Deviation	Range	N	Mean	Standard Deviation	Range
All Children								
<i>Standardized Assessments</i>								
PPVT Standard Score	200	90.67	25.70	11-142	200	96.17	21.14	41-143
WJ LWI Standard Score	200	100.64	14.31	67-162	200	105.03	13.75	69-159
WJ AP Standard Score	200	107.05	14.34	70-138	200	109.54	11.65	58-136
<i>Teacher-Rated DECA</i>								
Initiative T-Score	193 ¹	52.74	6.92	31-61	193	54.73	7.09	30-61
Self-Control T-Score	194	56.93	9.17	30-72	193	58.59	9.04	33-72
Attachment T-Score	194	50.75	8.73	30-72	193	52.27	9.76	28-72
Total Protective Factors T-Score	194	53.39	8.75	28-72	193	56.02	9.42	31-72
Behavioral Concerns T-Score	185	48.69	9.30	31-72	184	47.93	9.32	31-72
<i>Parent-Rated DECA</i>								
Initiative T-Score	--	--	--	--	185	51.53	10.51	28-72
Self-Control T-Score	--	--	--	--	185	55.15	9.36	28-72
Attachment T-Score	--	--	--	--	185	48.48	11.34	28-72
Total Protective Factors T-Score	--	--	--	--	185	51.27	10.46	28-72
Behavioral Concerns T-Score	--	--	--	--	185	54.65	10.82	28-72
<i>Spanish-Speaking Children Only</i>								
<i>Standardized Assessments</i>								
TVIP Standard Score	63	85.80	16.67	55-114	63	89.27	22.04	0-125
WM LWI Standard Score	63	98.21	12.34	72-118	63	101.58	17.83	70-176
WM AP Standard Score	63	96.02	12.06	62-118	63	99.05	12.01	57-121

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

Table 5: Weighted English Assessment Scores by Child's Primary Language, Fall Round¹

Assessment	Primary Language						t
	English			Another Language			
	N	Mean	SD	N	Mean	SD	
PPVT Standard Score	135	100.80	17.16	57	62.47	15.18	13.21***
WJ LWI Standard Score	135	103.62	11.78	57	88.36	9.67	7.76***
WJ AP Standard Score	135	111.45	10.14	57	92.32	10.40	10.38***

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

¹Information about the child's primary language was missing for one child in the sample.

Table 6: Weighted English Assessment Scores by Child’s Primary Language, Spring Round¹

Assessment	Primary Language						t
	English			Another Language			
	N	Mean	SD	N	Mean	SD	
PPVT Standard Score	135	108.47	15.47	57	77.11	16.75	12.52**
WJ LWI Standard Score	135	108.87	13.96	57	95.61	13.24	6.10***
WJ AP Standard Score	135	113.22	10.54	57	102.39	12.97	6.06***

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

¹Information about the child’s primary language was missing for one child in the sample.

PRESCHOOL QUALITY

The 201 children in the sample were enrolled in 100 different preschools.³⁵ Ninety-eight of these preschools were Qualistar rated. Detailed information about the quality of these preschools was provided to Clayton Early Learning from Qualistar. One site became eligible for DPP because they had obtained Accreditation from the National Association for the Education of Young Children (NAEYC). For this site, the only quality data that is available is the number of stars.³⁶ Finally, one provider’s rating was still in process at the time this report was prepared. No quality data are available for this provider yet. Figure 1 presents the breakdown of programs by star level. Over two-thirds of the programs had 3 stars. Nearly another fifth of programs had 4 stars. There was one preschool with a provisional rating and none with one star, indicating that very few programs were of the lowest quality.

Table 7 presents descriptive statistics for the five component areas of the Qualistar rating for the 98 sites with a Qualistar rating.³⁷ 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. There was variability around this mean, however. One program earned no points for this area, four earned four points. The remainder earned between 8 and 10 points. For Adult-to-Child Ratios and Group Size, the average of the programs was relatively high, but there was still some variability around that mean. Two programs earned 4 or fewer points, 13 programs earned between 5 and 7 points. The remainder earned between 8 and 10 points. On average, programs earned about 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

³⁵ It is important to keep in mind that all of the preschool quality information provided is based on only a sample of 100 preschools where the children in the sample were enrolled. For information on the quality of all preschool programs participating in DPP during the 09-10 school year, readers are referred to the annual evaluation report prepared by Augenblick, Palaich and Associates.

³⁶ 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.

³⁷ More information about the five component areas of the Qualistar rating is available at:

<http://www.qualistar.org/professionals/components.php>

60% of the possible points for Learning Environment. There was a good deal of variability around this mean as well. 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=100 programs).

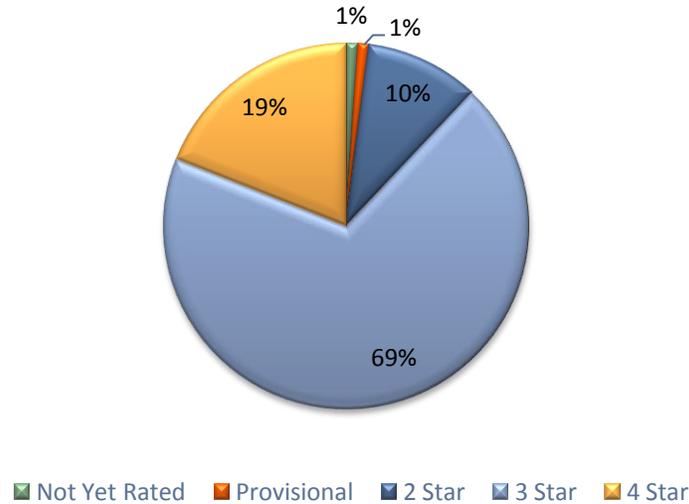


Table 7: Qualistar Rating Components for Programs Attended by Children in the Sample (n=98 programs).

Component	Possible Range	Mean	SD	Range
Learning Environment Points	0-10	6.18	1.63	0-10
Family Partnerships Points	0-10	9.14	1.67	0-10
Training and Education Points	0-10	5.59	1.91	0-9
Adult-to-Child Ratios and Group Size Points	0-10	8.81	1.48	3-10
Accreditation Points	0-2	0.04	0.28	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 differed significantly in total rating points earned,³⁸ with DPS programs earning significantly more points on average than Community programs (see Figure 2). There were also significant differences in the areas of Family Partnership Points³⁹ and Training and Education Points⁴⁰ (see Figure 3). DPS programs earned significantly more points in these areas than did Community programs. There was a trend toward a significant difference between program types on Ratio and Group Size,⁴¹

³⁸ $t(61.4)=2.19, p<.05$

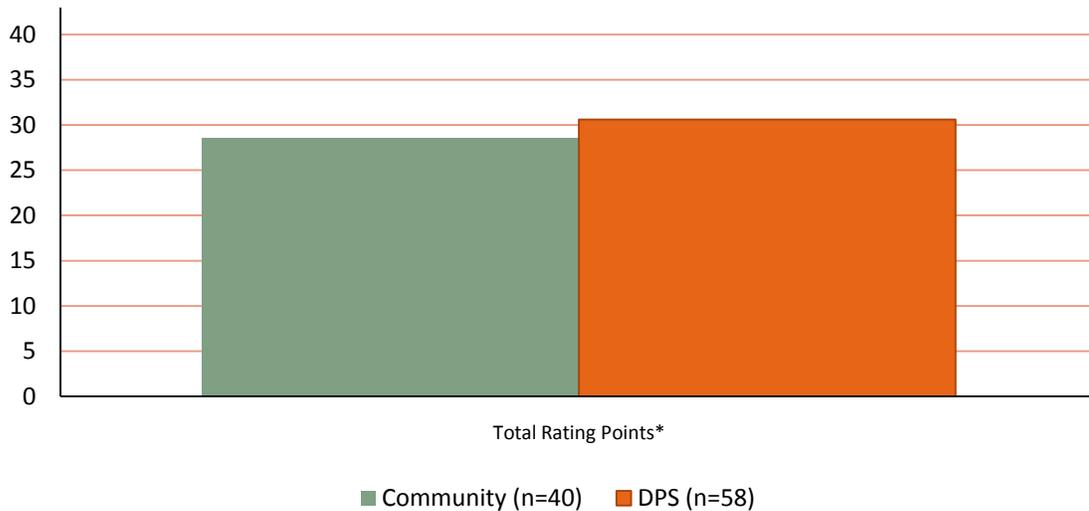
³⁹ $t(45.9)=2.87, p<.01$

⁴⁰ $t(66.1)=4.04, p<.001$

⁴¹ $t(95.2)=1.78, p<.10$

with Community programs scoring slightly higher in this area than DPS programs. There was not a significant difference between program types in the areas of Learning Environment⁴² or Accreditation.⁴³

Figure 2: Qualistar Total Rating Points for Programs Attended by Children in the Sample, by Provider Type (n=98 programs)¹



* $p < .05$

¹Standard deviations: Community=5.11, DPS=3.32

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 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 southeast and northeast regions of the city tended to be in the programs earning the highest number of rating points on average. Children in the southeast region of the city tended to be in programs that had earned significantly more rating points, on average, than children in the northwest and southwest regions of the city. Children in the northeast region of the city tended to be enrolled in programs that had earned significantly more rating points, on average, than children in the northwest region of the city. These differences were of small magnitude, however. The largest difference was between children in the

⁴² $t(64.4)=0.08$, n.s.

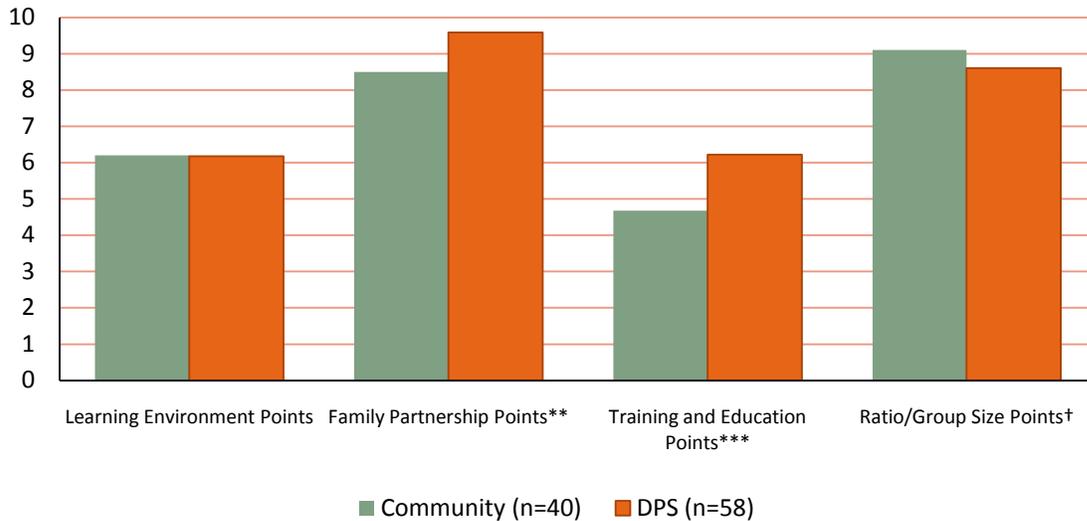
⁴³ Results for Accreditation Points are not depicted in Figure 7 because the possible range for Accreditation Points is rather small relative to the other areas (2 points). Further, as seen in Table 6, it was quite rare for programs to be accredited. The mean for accreditation points for Community sites was .1 ($sd=.44$). None of the DPS sites was accredited.

⁴⁴ Income tier: $F(3,187)=2.52$, 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 $F(4,187)=5.61$, $p < .001$

southeast region, who were enrolled in programs that earned about 32 points on average, and those in the northwest region, who were enrolled in programs that earned about 29 rating points on average.⁴⁶

Figure 3: Qualistar Rating Components for Programs Attended by Children in the Sample, by Provider Type (n=98 programs)¹



† p<.10; * p<.05; ** p<.01; *** p<.001

¹Standard deviations: Learning Environment Points: Community=1.96, DPS=1.37; Family Partnership Points: Community=2.30, DPS=0.82; Training and Education Points: Community=1.50, DPS=1.76; Ratio/Group Size Points: Community=1.06, DPS=1.70

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 4 presents the percent of children scoring 85 or above and 100 or above on each of the assessments at the spring time point.

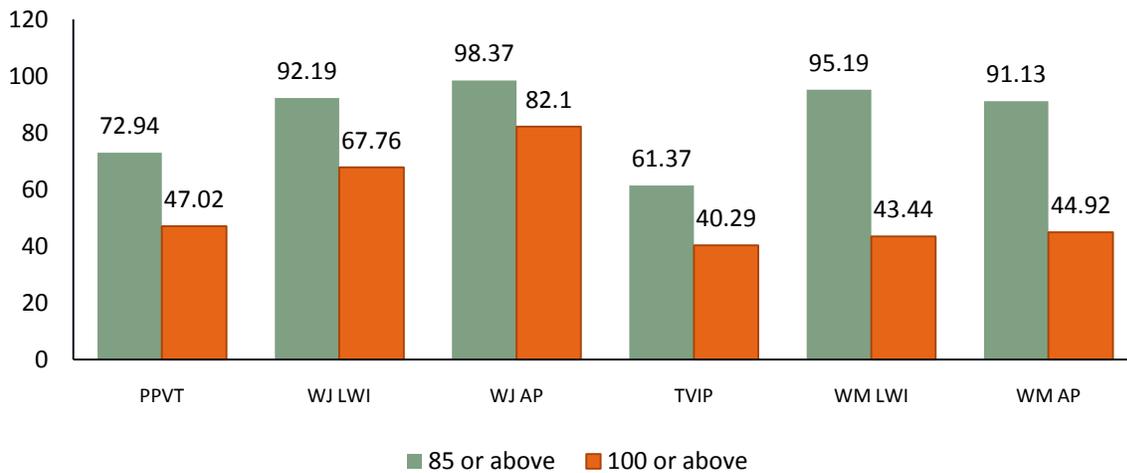
For the English assessments, the vast majority of children (over 90%) scored 85 or above on the WJ LWI and WJ Applied Problems assessments and over two-thirds of children scored 100 or above on these assessments.

⁴⁶ Rating points: Central Region-mean=30.58, sd=3.27; Northeast Region-mean=31.36, sd=3.37; Northwest Region-mean=28.60, sd=3.68; Southeast Region-mean=31.99, sd=3.80; Southwest Region-mean=29.59, sd=3.30

Nearly three-quarters 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 (97%) scored 85 or above on the PPVT as compared with a relatively small proportion of children whose primary language was not English (30%).⁴⁷ A similar, but less pronounced pattern was observed for WJ LWI and Applied Problems. Eighty-three percent of children whose primary language was something other than English scored 85 or above on WJ LWI as compared with 97% of children whose primary language was English.⁴⁸ For WJ Applied Problems, 96% of children whose primary language was not English scored 85 or above as compared with nearly all children whose primary language was English (99.75).⁴⁹

Figure 4: 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, 69% of children whose primary language was English earned a score of 100 or greater as compared with 10% of children with another primary language.⁵⁰ For WJ LWI, 81% of children whose primary language was English scored 100 or greater as compared with 43% of children with another primary language.⁵¹ Finally, for WJ Applied

⁴⁷ $\chi^2_1=99.43, p<.001$

⁴⁸ $\chi^2_1=11.31, p<.001$

⁴⁹ $\chi^2_1=3.84, p=.05$

⁵⁰ $\chi^2_1=62.47, p<.0001$

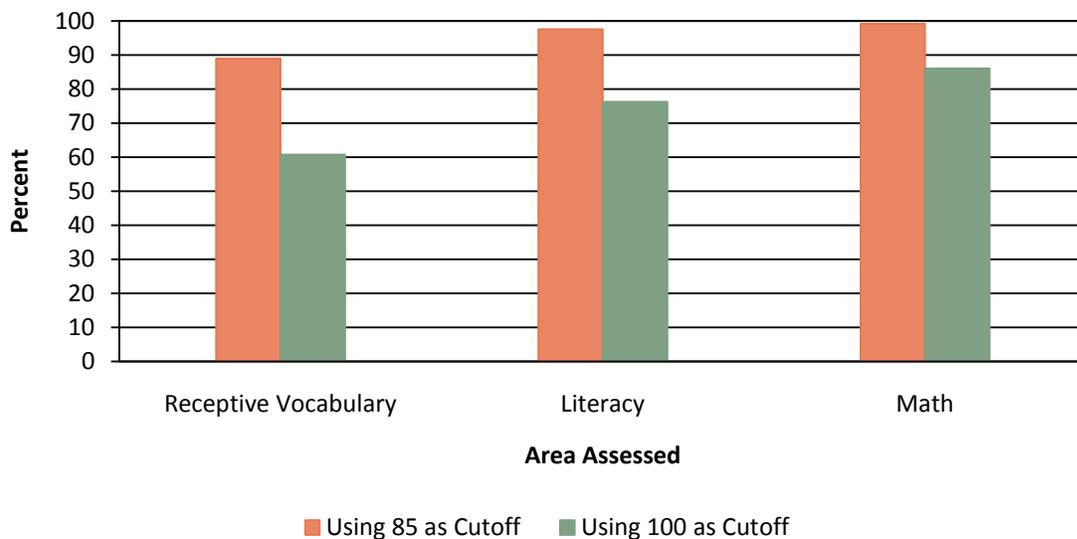
⁵¹ $\chi^2_1=28.40, p<.0001$

Problems, nearly all children whose primary language was English earned scores of 100 or above compared with about two-thirds of children whose primary language was something other than English.⁵²

For assessments administered in Spanish, scores were once again stronger for LWI and Applied Problems than for vocabulary (TVIP). Over 90% of children scored 85 or above on WM LWI and Applied Problems while about 60% of children scored 85 or above on the TVIP. About 40% scored 100 or above on the TVIP, slightly less than half scored 100 or above on the WM LWI and 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 5. Nearly 90% of children met or exceeded the cutoff of 85 in at least one language in the area of receptive vocabulary (i.e., PPVT or TVIP). Nearly all children met or exceeded the cutoff of 85 in at least one language on the literacy assessment (WJ-LWI or WM-LWI) and 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; over three-quarters met or exceeded this benchmark for both literacy and math.

Figure 5: Weighted Percent of Children Scoring in the Average Range or Above on Spring Standardized Assessments in Spanish or English

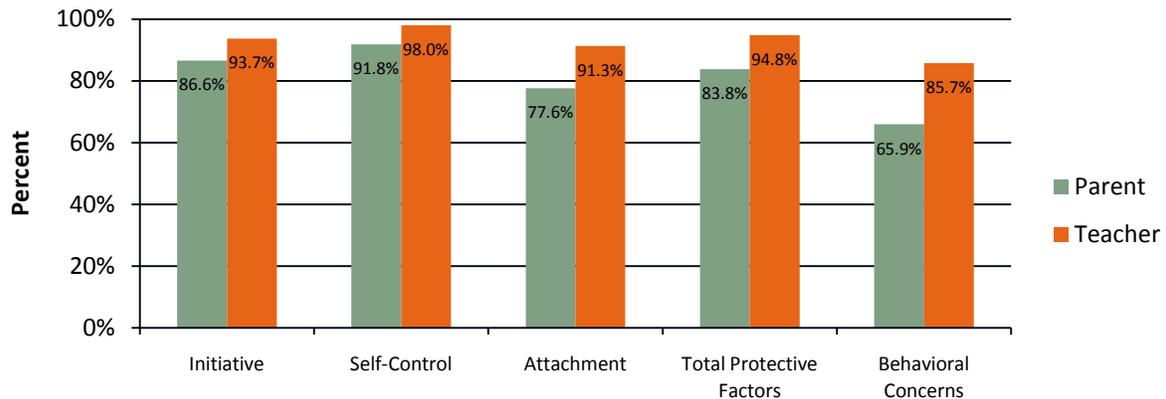


⁵² $\chi^2_1=17.52, p<.0001$

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 6, 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 three-fourths of children in the typical or strength range for Attachment and nearly two-thirds 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, Attachment and Protective Factors. They also rated nearly 90% of children in the typical range for Behavioral Concerns.

Figure 6: Weighted Percent of Children Scoring in the Average Range or Above on Spring Parent and Teacher DECA Surveys.



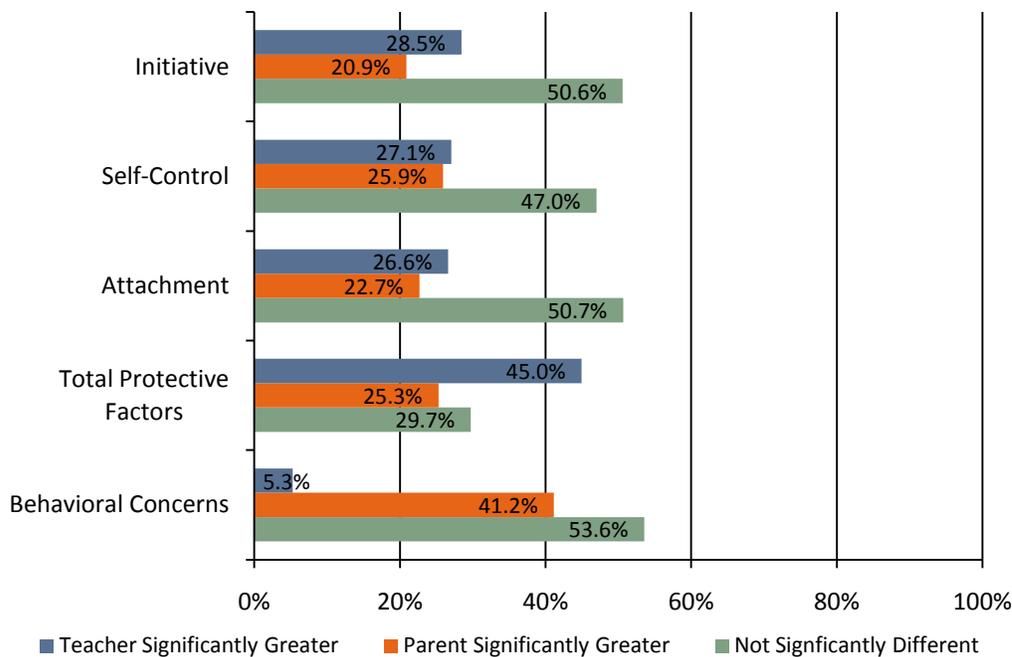
We examined the differences between teachers’ and parents’ ratings using guidelines from the authors of the DECA. The authors developed these guidelines to help users distinguish between differences in scores due to measurement error and differences that are likely due to a meaningful difference between 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, 3.6 ($sd=11.5$), did not exceed this threshold, indicating that, on average, teachers’ and parents’ ratings did not differ. As displayed in Figure 7, for about half of children, the difference between the teacher’s and parent’s report did not significantly differ. For slightly over a quarter of the sample, the teacher’s rating was significantly greater than the parent’s rating. For about a fifth of the sample, the parent’s rating was significantly greater than teacher’s rating.

For Self-Control, a difference of 10 is needed to conclude that there is a significant difference between the parent and teacher rating. The average difference between teachers’ and parents’ reports, 3.5 ($sd=12.6$), did not exceed this threshold, indicating that, on average, teachers’ and parents’ ratings did not differ. As displayed in Figure 7, for nearly half of children, the difference between the teacher’s and parent’s report did not significantly differ. The remainder of the sample was about equally split between the two remaining groups. That is, the parent’s score was significantly higher than the teacher’s about as often as the reverse was true.

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, 3.8 ($sd=14.4$), 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 about as common for the teacher’s score to be higher as 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 4.9 ($sd=13.2$), which was below that threshold. However, for 45% of children, the teacher’s rating was significantly greater than the parent’s rating (see Figure 7). For slightly less than a third of children, there was not a significant difference between raters.

Figure 7: Comparison of Parent and Teacher DECA Surveys, Weighted



For Behavioral Concerns, a difference of 14 points is needed to conclude that there is a significant difference between the parent’s and teacher’s ratings. The average difference in the sample was 7.2 ($sd=13.8$), 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 7). 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, there did not appear to be a tendency for one rater to rate children higher than the other. 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 8. There were significant increases in all three standardized assessments administered in English. For the PPVT and LWI, the increase was rather large, about a third of a standard deviation. For Applied Problems, the increase was about a sixth of a standard deviation. It is important to keep in mind that these scores are adjusted for age, so these increases are above and beyond what one would expect due to typical maturation. For the Spanish assessments, significant increases were observed for LWI and AP. The magnitude of both of the increase for both of these assessments was about a fifth of a standard deviation. Children's scores remained relatively constant on the TVIP.

Significant improvements were also observed in many of 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 (nearly one-quarter of a standard deviation), and the overarching Protective Factors scale (over a quarter of a standard deviation).

Table 8: 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	199	90.52 (25.71)	96.05 (21.16)	6.99***
WJ-LWI	199	100.46 (14.20)	105.05 (13.79)	7.08***
WJ-AP	199	107.04 (14.38)	109.50 (11.67)	3.52**
Standardized Assessments—Spanish				
TVIP	63	85.80 (16.68)	89.27 (22.05)	1.85
WM-LWI	63	98.22 (12.35)	101.58 (17.83)	2.01*
WM-AP	63	96.03 (12.07)	99.06 (12.01)	3.28**
Teacher Survey				
Initiative T-Score ¹	191	52.66 (6.91)	54.95 (6.89)	5.75***
Self-Control T-Score	192	56.83 (9.16)	58.64 (9.06)	3.78***
Attachment T-Score	192	50.71 (8.76)	52.35 (9.74)	3.09**
Total Protective Factors T-Score	192	53.31 (8.76)	56.14 (9.36)	5.76***
Behavioral Concerns T-Score	175	49.08 (8.91)	48.38 (9.12)	-0.24**

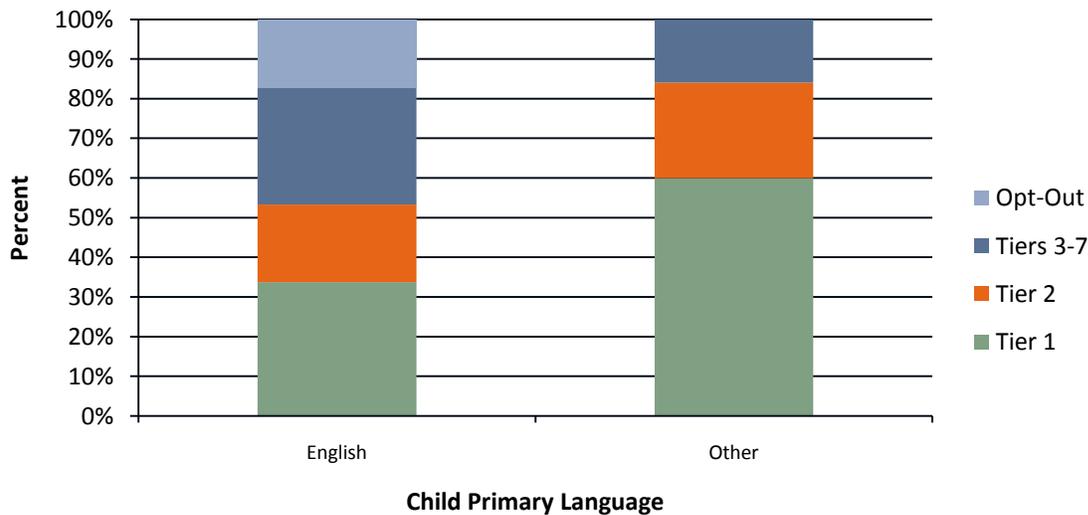
*p<.05, **p<.01, ***p<.001

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

CHANGE OVER TIME BY SUBGROUP

Further analyses were conducted to test whether the extent of the change over time varied by a variety of background characteristics: income tier, children’s primary language and Qualistar Rating of the preschool. Prior to conducting analyses by income tier, some data reduction was necessary since the number of participants for some of the income tiers was rather small (see Table 1). Income tier was collapsed into a new income tier group variable with 4 categories: tier 1, tier 2, tiers 3-7 and opt-out (i.e., parents who opted out of the requirement to report income and instead elected to automatically be assigned to tier 7).⁵³ It is important to note that two of these background characteristics, income tier and child’s primary language are strongly associated (see Figure 8).⁵⁴ Over 80% of children whose primary language is something other than English are from tiers 1 or 2 whereas only about half 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 8: Income Tier Groups, by Child Primary Language



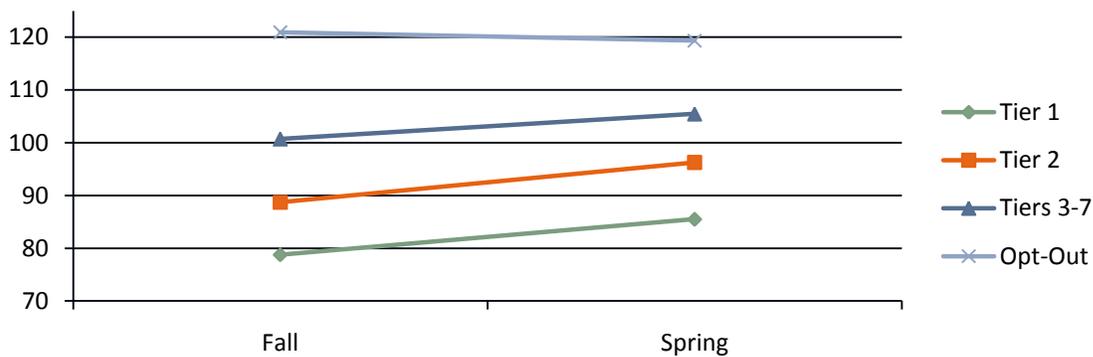
⁵³ For analyses of assessments administered in Spanish, a three-level income tier group variable was used because only one family with a child assessed in Spanish opted out of reporting income. These families were dropped from the analyses.

⁵⁴ $\chi^2_3=22.58, p<.0001$

INCOME TIER

A series of Repeated Measures ANOVAs⁵⁵ 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 was significant interaction between income tier group and time for PPVT,⁵⁶ WJ LWI,⁵⁷ and WJ AP.⁵⁸ There was a trend toward a significant interaction for DECA Attachment.⁵⁹ Results of these analyses are depicted in Figures 9-12. Figure 9 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 and opt-out category. Follow-up Tukey⁶⁰ tests revealed that children from Tiers 1 and 2 increased significantly more over time than children in the Opt-Out category. Pairwise comparisons of the other groups were not significant.

Figure 9: Weighted PPVT Standard Scores over Time, by Income Tier Group¹



¹Standard Deviations: Tier 1: Fall=20.53, Spring=17.35; Tier 2: Fall=23.54, Spring=19.60; Tiers 3-7: Fall=26.84, Spring=20.34; Opt-Out: Fall=8.69, Spring=10.53.

⁵⁵ 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.

⁵⁶ $F(3,195)=4.00, p<.01$

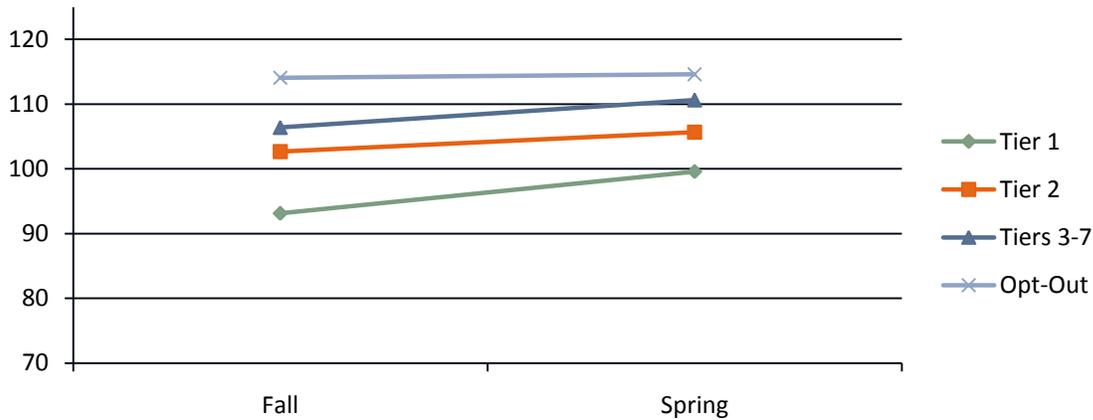
⁵⁷ $F(3,195)=3.18, p<.05$

⁵⁸ $F(3,195)=3.46, p<.05$

⁵⁹ $F(3,188)=2.23, p<.10$

⁶⁰ 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 10: Weighted WJ Letter-Word Identification Standard Scores over Time, by Income Tier Group¹

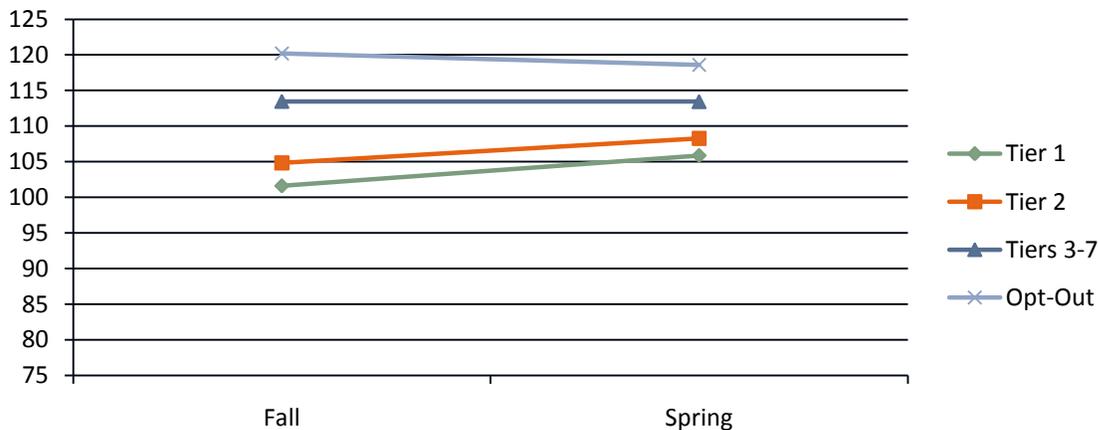


¹Standard Deviations: Tier 1: Fall=10.47, Spring=12.90; Tier 2: Fall=15.32, Spring=12.21; Tiers 3-7: Fall=13.79, Spring=13.08; Opt-Out: Fall=8.94, Spring=12.15.

Results of the analysis of WJ LWI are presented in Figure 10. Follow-up Tukey tests revealed a significant difference in change over time between Tier 1 and the Opt-Out group. Children in Tier 1 increased significantly more over the course of the school year than children in the Opt-Out group.

Results of the analysis of WJ Applied Problems are presented in Figure 11. Follow-up Tukey tests did not reveal any significant differences in the rate of change over time between groups, but there were two trends. Children in Tier 1 increased more over the course of the school year than children in Tiers 3-7 and the Opt-Out group, whose scores remained relatively stable over the course of the year.

Figure 11: Woodcock-Johnson Applied Problems Scores over Time, by Income Tier Group¹

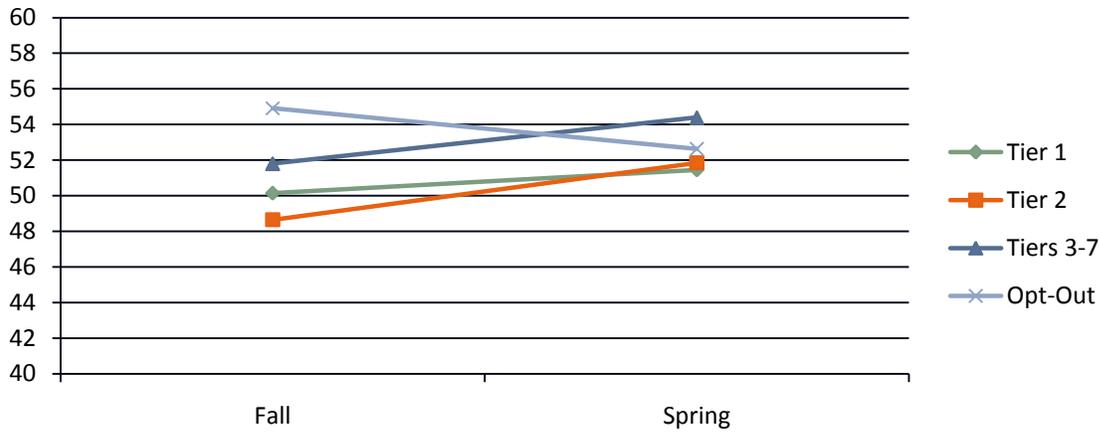


¹Standard Deviations: Tier 1: Fall=13.53, Spring=11.21; Tier 2: Fall=14.16, Spring=11.85; Tiers 3-7: Fall=13.06, Spring=11.23; Opt-Out: Fall=6.26, Spring=6.73.

Results for teachers' ratings on DECA Attachment over time are presented in Figure 12. Follow-up Tukey tests revealed a trend toward a significant difference between Tier 2 and the Opt-Out group. Children in Tier 2

increased more over time, on average, in teacher ratings of Attachment, than children in the Opt-Out group, who decreased over time.

Figure 12: Weighted Teachers' Ratings of Children's Attachment over Time, by Income Tier Group¹



¹Attachment was measured with the DECA. Standard Deviations: Tier 1: Fall=9.25, Spring=10.35; Tier 2: Fall=7.73, Spring=9.30; Tiers 3-7: Fall=8.30, Spring=9.40; Opt-Out: Fall=8.47, Spring=8.72.

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.⁶¹ There were significant interactions between primary language group and time for PPVT,⁶² WJ LWI,⁶³ and WJ AP.⁶⁴ There was a trend toward a significant interaction between primary language group and time for DECA Initiative.⁶⁵ Results of these analyses are presented in Figures 13-16.

⁶¹ 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.

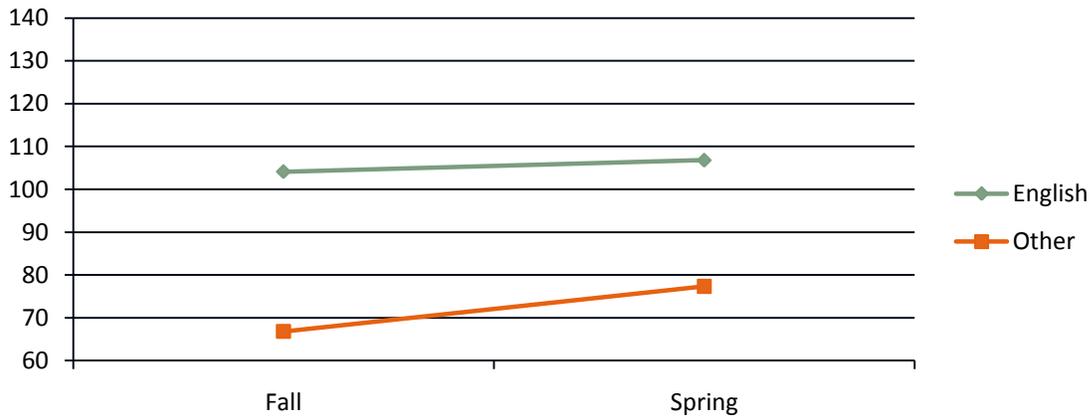
⁶² $F(1,189)=23.68, p<.0001$

⁶³ $F(1,189)=5.26, p<.05$

⁶⁴ $F(1,189)=42.70, p<.0001$

⁶⁵ $F(1,182)=2.98, p<.10$

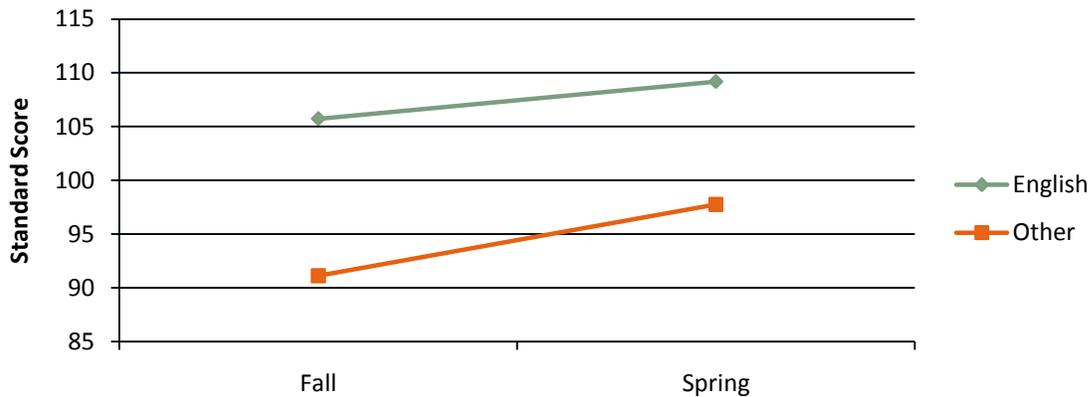
Figure 13: Weighted PPVT Scores over Time, by Child Primary Language¹



¹Standard Deviations: English: Fall=19.21, Spring=15.07; Other: Fall=17.98, Spring=18.19.

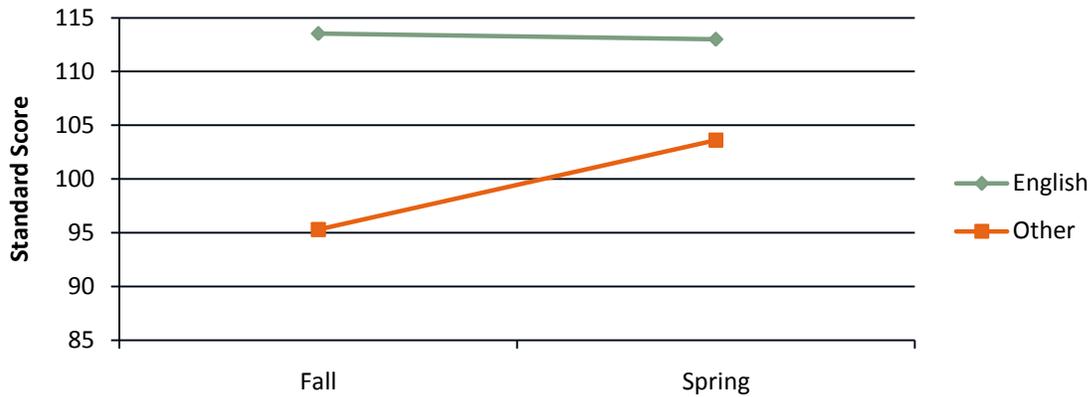
For all four assessments, children whose primary language was not English showed a larger increase in scores from fall to spring. For PPVT, children whose primary language was English increased only slightly, while their counterparts with another primary language increased an average of about two-thirds of a standard deviation. For WJ LWI, children whose primary language was English increased an average of nearly a third of a standard deviation, while their counterparts with another primary language increased an average of nearly half of a standard deviation. For WJ AP, scores children whose primary language was English remained relatively constant over the course of the school year, while scores for children with another primary language increased by over half of a standard deviation on average. Finally, for Initiative, scores for children whose primary language was English increased only slightly, while scores for children with another primary language increased by about a third of a standard deviation. For the remaining subscales of the teacher-rated DECA, 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 14: Weighted Woodcock-Johnson LWI Scores over Time, by Child Primary Language¹



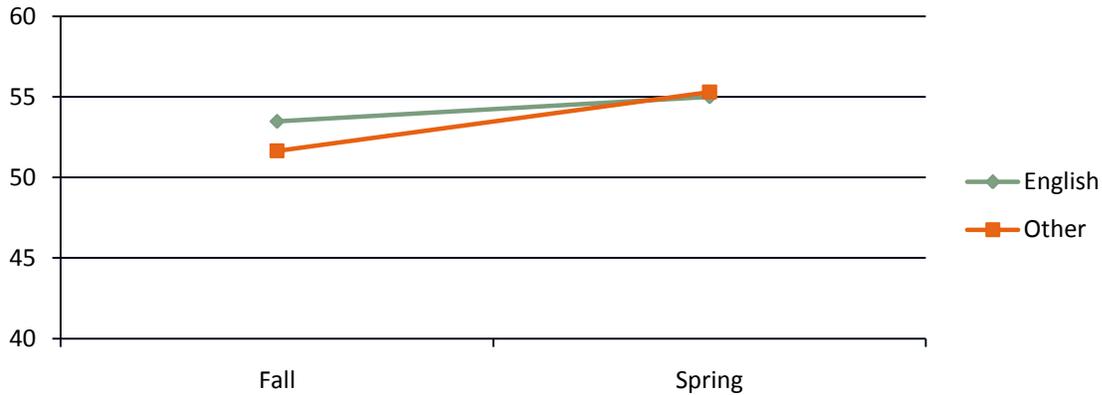
¹Standard Deviations: English: Fall=13.09, Spring=12.29; Other: Fall=11.46, Spring=14.02.

Figure 15: Weighted Woodcock-Johnson AP Scores over Time, by Child Primary Language¹



¹Standard Deviations: English: Fall=11.37, Spring=9.92; Other: Fall=12.32, Spring=12.69.

Figure 16: Weighted Teacher-Rated Initiative Scores over Time, by Child Primary Language¹



¹Standard Deviations: English: Fall=6.70, Spring=7.15; Other: Fall=6.51, Spring=5.22.

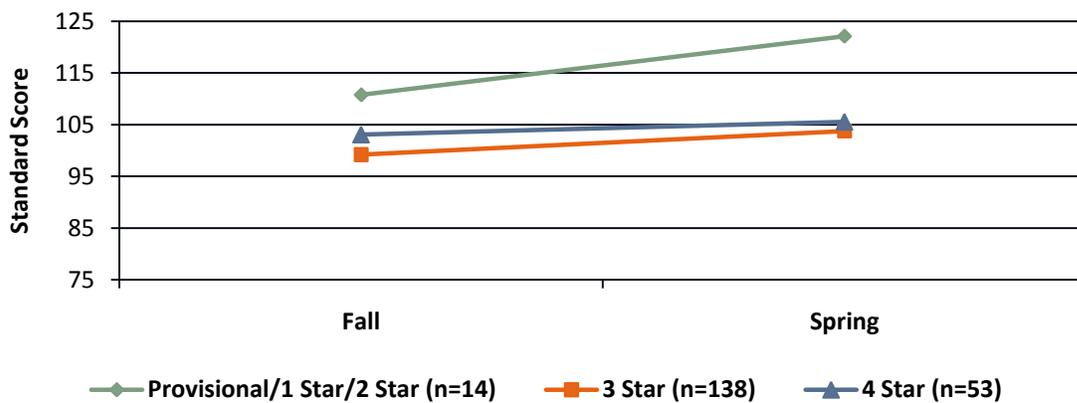
QUALISTAR RATING

A series of repeated measures ANOVAs were conducted to test for differences in change over time by the Qualistar rating of the preschool. Since previous analyses (described above under the heading “Preschool Quality”) indicated that region of the city was associated with the Qualistar rating of the preschool children attended, this variable was included in all the analyses as a control variable, in an effort to estimate the effect of the Qualistar rating net of the effect of region of the city. Prior to analysis, star level was collapsed into a new star group variable with three levels: Provisional/1 Star/2 Star, 3 Star, and 4 Star. This was necessary because relatively few children in the sample were enrolled in sites with a Provisional, 1 Star or 2 Star rating. Analyses were conducted for each of the standardized assessments administered in English, each of the assessments administered in Spanish, and teacher-rated DECA. The time by star rating group effect was significant for two assessments: WJ LWI

and WM LWI.⁶⁶ For all other assessments, the time by star rating group effect was non-significant, indicating that children in the three star rating groups, progressed on average, at a similar rate over time.

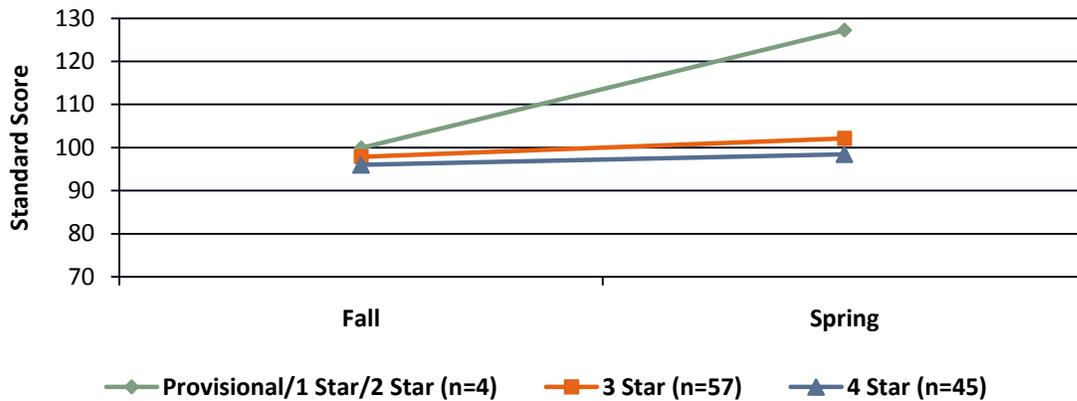
The effect for WJ LWI is depicted in Figure 17. Follow-up analyses⁶⁷ revealed that change over time for the Provisional/1 Star/2 Star group was significantly different than change for the other two groups. Contrary to what might be expected, children in the lowest quality programs started higher and increased more over time than their counterparts in the higher program quality groups. A similar effect was observed for the Spanish version of the same assessment, WM LWI (see Figure 18). Once again, follow-up tests revealed that change over time for the Provisional/1 Star/2 Star group was significantly greater than change for the other two groups. This group started higher and increased more over time than children in the higher quality groups.

Figure 17: Weighted WJ Letter-Word Identification Scores over Time, by Star Level of Program¹



¹Means presented are least-squared means after controlling for the effect of region of the city.

Figure 18: Weighted WM Letter-Word Identification Scores over Time, by Star Level of Program¹



¹Means presented are least-squared means after controlling for the effect of region of the city. N's provided are weighted.

⁶⁶ WJ LWI: $F(2,191)=4.67, p<.05$; WM LWI: $F(2,56)=5.94, p<.01$

⁶⁷ T-tests for differences in fall-to-spring change scores were conducted, using a Bonferroni adjustment for multiple comparisons.

MULTIVARIATE ANALYSES

One possible explanation for the unexpected findings for Qualistar rating as a predictor of children's outcomes over the course of their preschool year is that there was limited variability in the star rating. Collapsing the variable into 3 levels helped somewhat for analysis purposes, but the net result was still that only about 7% of children (weighted, see Table 1) were in programs with less than a 3 star rating. Examination of the demographic characteristics of this small subgroup revealed that they were more likely to be from the opt-out tier.⁶⁸ Twenty-two percent of children from the opt-out tier were enrolled in these lower quality programs compared with only 7% of children from other income tiers. The children enrolled in programs with less than a 3 star rating did not differ significantly from children in higher rated programs on any of the other demographic characteristics. Nonetheless, it is difficult to draw firm conclusions based on such a small number of children. A final set of multivariate analyses was conducted using more precise variables related to quality. Namely, we used number of Qualistar rating points in lieu of the number of stars, number of ratio/group size points, number of training and education points awarded, and mean ECERS-R⁶⁹ score for all DPP preschool classrooms at the site.⁷⁰ A series of regressions predicting children's scores on the outcome variables in the spring were conducted with child's primary language, income tier, ethnicity, and region of the city included as control variables. In addition, the children's score in the fall on the same measure was also included as a control. As a result these analyses, while not specifically focused on change over time (i.e., the actual difference between fall and spring scores), do examine "residualized gain," which can be understood as how children score in the spring after taking into account the differences between them in the fall.

The first set of regressions included number of rating points as the predictor of interest. Number of rating points emerged as a significant predictor of 2 of the 11 child outcome variables (3 standardized assessments in English, 3 standardized assessments in Spanish, and 5 teacher-rated DECA variables). For TVIP and WM LWI, there was a negative association between number of rating points and children's scores on these assessments in the spring, after controlling for demographic characteristics and the same assessment in the fall. For TVIP, an increase of one rating point was associated with about a 1.7 point decrease in TVIP scores.⁷¹ Similarly, an increase of one rating point was associated with about a 1.4 point decrease in WM LWI scores.⁷²

When number of ratio/groups size points earned was included as a predictor of interest, it was a significant predictor for 1 of the 11 dependent variables. Looking across children and adjusting for demographic characteristics and fall TVIP scores, there was a negative effect of ratio/group size points on TVIP. A one point increase in ratio/group size points was associated with nearly a one third of a standard deviation decrease in TVIP

⁶⁸ $\chi^2_1=5.15, p<.05$

⁶⁹ The Early Childhood Environment Rating Scale, Revised is an observational tool used by Qualistar as part of its rating. The ECERS-R includes items that are rated on a 7-point rating scale. (Harms, T., Clifford, R. M., & Cryer, D. (1998). *Early Childhood Environment Rating Scale, Revised Edition*. Teachers College Press, New York.).

⁷⁰ There was not sufficient variability in the other Qualistar rating component areas (ratio, family partnerships, and accreditation) to warrant including them in these analyses.

⁷¹ $b=-1.68, t=2.38, p<.05$

⁷² $b=-1.37, t=2.56, p<.05$

scores.⁷³ When number of training and education points earned was included as a predictor of interest, it did not emerge as a significant predictor in any of the models.

When ECERS-R scores were included as the predictor of interest, one significant effect was observed. Once again, it was a negative association with an assessment of Spanish speakers, WM LWI. Looking across children and adjusting for the effects of the control variables and WM LWI scores in the fall, a one point increase in average ECERS-R score was associated with a nearly half of a standard deviation decrease in WM LWI scores.

The number of significant associations is relatively small, 9% of the analyses conducted, but it exceeds what one would expect by chance (5%). In addition, an unexpected pattern of negative associations between measures of quality and assessments of Spanish-speaking children's language and literacy did emerge. Further, this finding echoes a similar unexpected finding in our DPP evaluation report for the 08-09 school year. Further investigation of this unexpected finding is warranted. It may be the case that some of the higher scoring preschools have explicit English-only language policies. These programs may be doing a good job of teaching children English, but have less of an emphasis on supporting children's Spanish language skills. Unfortunately, we do not have systematic information about the preschools' language policies, so are unable to test this hypothesis.

⁷³ $b=-4.91, t=2.46, p<.05$

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?
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, we were unable to address question 4 in this report because we were unable to obtain kindergarten assessment scores in time for inclusion in this report. We will prepare a separate report on the kindergarten assessment data for children enrolled in DPP during the 08-09 school year upon receipt of those data from DPS. In addition, we are only partially able to address question 5. This report provided results relevant to how preschool program quality ratings are associated with kindergarten readiness but not with long-term academic success.

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, literacy and math skills. Spanish-speaking children also made progress in their Spanish literacy skills and their math skills (assessed in Spanish) over the course of their preschool year. The gains observed were above and beyond what would be expected based on normal development. Progress was observed in socio-emotional development as well. Over the course of the preschool year, teachers reported that children demonstrated significantly more protective factors and fewer behavioral concerns.

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

Results of the evaluation suggest that the vast majority of children are ready for school, both academically and socio-emotionally. When considering both languages of assessment, we concluded that relatively few children had scores in the risk range (below 85) on assessments of their vocabulary, literacy and math skills. These standardized assessments are scaled such that 84% of the general population would be expected 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 approach that threshold. We also considered a more stringent criterion to

examine readiness, namely scores that met or exceeded the population average (a score of 100). The assessments are scaled such that half of children in the general population would be expected to meet or exceed this threshold. When both languages of assessment were considered, more children than would be expected (i.e., more than half) met this more stringent criterion: more than half for vocabulary and over 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 10% of children. Teachers' ratings of behavioral concerns were rather low on average. Teachers identified behavioral concerns as an area of concern for fewer than 15% of children. Parents identified protective factors as an area of concern for about 16% of children and behavioral concerns as an area of concern for about a third 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 twice the number of 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, over 80% of children whose primary language was not English were from the lowest two income tiers as compared with about half 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 the Opt-Out category, who are likely to be from higher income tiers, and/or Tiers 3-7. Finally, with respect to teachers' ratings of Attachment, children in Tier 2 increased significantly more than children in Opt-Out category, 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.

The results for primary language followed a similar pattern, in that there were significant effects for all three academic assessments in English, and one significant effect for the socio-emotional assessment. Children whose primary language was something other than English tended to start lower on the academic assessments. Again, with just one significant effect out of five socio-emotional scores examined, there is not a strong pattern of results for this area. This pattern of findings supports the conclusion that children who are learning English as a

second language make steeper progress over the course of the school year in academic areas than children who primarily speak English. In terms of socio-emotional development, children are making similar progress during their DPP year, regardless of their primary language.

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

There was some weak evidence for an unexpected association between star level and literacy scores both in English and Spanish. Children in the lowest quality programs started higher and increased more over time than children in higher quality programs. This association was not observed for the other 9 assessments examined, however. When interpreting these results, it is important to keep in mind that the vast majority of children in the evaluation sample were enrolled in 3 star or 4 star preschools. Only 7% of children were enrolled in lower quality preschools. This greatly limits our ability to address this question using the star rating, as it effectively reduces the question to one of whether outstanding preschools are associated with greater kindergarten readiness than very good preschools. However, the Qualistar rating is comprised of 5 component areas, three of which had greater variability than the star rating categories: ratio/group size points, training and education points and learning environment (as measured by mean ECERS-R scores). In these analyses, an unexpected, but fairly consistent pattern emerged. Being in a higher-quality program tended to be associated with lower scores on assessments administered in Spanish. This finding echoes an unexpected finding in our evaluation report for the 08-09 school year. In that report, we found that children in programs with higher scores for their learning environment tended to have lower Spanish vocabulary scores. Further examination of this unexpected finding is warranted. It may be the case that some of the higher scoring preschools have explicit English-only policies. These programs may do a good job of teaching children English, but have less of an emphasis on supporting children's Spanish language skills. Unfortunately, we do not have systematic information about the preschools' language policies, so we are unable to test this hypothesis.

SUMMARY AND FUTURE DIRECTIONS

This evaluation described children's progress during the course of their DPP preschool year. In general, children progressed in their academic skills as assessed in English at a rate which exceeded what would be expected simply because of maturation. Spanish speaking children made significant gains in their literacy and math skills (assessed in Spanish) but 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 and fewer behavior problems at the end of the school year than at the beginning. Growth in vocabulary, literacy and math 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 fairly consistent across subgroups.

Nearly all children in our sample were enrolled in a higher quality program. Over 90% of children attended a preschool that had earned either 3 or 4 stars. This greatly limited our ability to adequately test the association between preschool quality and child outcomes over time. However, a fairly consistent pattern of results emerged with children in higher quality programs scoring lower on assessments administered in Spanish in

the spring of their preschool year than their counterparts in lower-quality programs, after taking into account demographic characteristics and children's scores on the same assessment in the fall. One must be careful to interpret this finding with extreme caution as it was unexpected and the study included few variables that could help shed light on the meaning of this association. One alternative explanation is that the programs earning a higher learning environment rating were more likely to be English-only programs. Bilingual children who enter a preschool setting where teachers do not speak Spanish in the classroom would be expected to lose their Spanish skills over time *if their Spanish was not being adequately supported at home*. We have proposed to collect more information on program quality as part of our evaluation for the 10-11 school year. We will conduct classroom observations using the CLASS (Classroom Assessment Scoring System: Pianta, LaParo, and Hamre, 2008). While we are conducting these observations, we could also have our observers collect some information from teachers about how they use Spanish in the classroom as well as make notes on how Spanish is used during the observation. These additional sources of information could potentially shed light on these unexpected findings.

Consistent with findings in our 08-09 report, parents rated their children's socio-emotional functioning more negatively than did teachers. In particular, 40% of parents identified significantly more behavioral concerns than did their child's teacher. While it is certainly common for children's behavior to differ in the home and school settings, one would expect that children would vary with respect to which setting they exhibited fewer behavioral concerns. Data from Qualistar indicate that, by and large, these preschools have strong relationships with families (on average, programs earned over 9 out of the 10 possible points on this component of the Qualistar rating). Perhaps programs could consider using these strong relationships to support families facing challenges with their children's behavior, both in terms of supporting positive, protective factors and reducing behavior problems.

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. These early results are promising. Results from future years of this annual evaluation will provide the opportunity to replicate these findings as well as follow children from this evaluation cohort into elementary school.

APPENDIX

Table A1: Sample Characteristics—Spring 2010

Characteristic	Entire Sample, weighted ¹	By Provider Type, Unweighted		
		Community	DPS	Significance of Difference by Provider Type
Sex				$\chi^2_1=2.91$; $p<.10$
Female	49.1%	58.6%	46.5%	
Male	50.9%	41.4%	53.5%	
Ethnicity				$\chi^2_5=18.26$; $p<.01$
Hispanic	50.9%	30.3%	56.4%	
White (not of Hispanic origin)	25.6%	36.4%	22.8%	
African-American (not of Hispanic origin)	15.0%	19.2%	13.9%	
Asian/Pacific Islander	2.8%	6.1%	2.0%	
Multi-Racial	4.1%	8.1%	3.0%	
Other	1.6%	0.0%	2.0%	
Child's Primary Language				$\chi^2_1=9.80$; $p<.01$
English	62.3%	76.8%	58.4%	
Another Language	34.3%	18.2%	38.6%	
Not Reported	3.4%	5.1%	3.0%	
Home Language				$\chi^2_1=11.95$; $p<.001$
English	58.5%	73.7%	54.5%	
Another Language	36.7%	18.2%	41.6%	
Not Reported	4.8%	8.1%	4.0%	
DPP Income Tier ²				$\chi^2_6=7.35$; ns.
Tier 1	44.7%	41.4%	45.5%	
Tier 2	20.8%	17.2%	21.8%	
Tier 3	4.0%	4.0%	4.0%	
Tier 4	0.2%	1.0%	0.0%	
Tier 5	5.9%	2.0%	6.9%	
Tier 6	0.0%	0.0%	0.0%	
Tier 7—Income Reported	13.6%	20.2%	11.9%	
Tier 7—Income Not Reported	10.8%	14.1%	9.9%	
Star Level of Preschool				$\chi^2_4=9.59$; $p<.05$
Preschool Not Yet Rated	0.2%	1.0%	0.0%	
Provisional	0.2%	1.0%	0.0%	
Star 1	0.0%	0.0%	0.0%	
Star 2	7.0%	11.1%	5.9%	
Star 3	69.1%	53.5%	73.3%	
Star 4	23.4%	33.3%	20.8%	
Region of the City				$\chi^2_4=12.36$; $p<.05$
Central	13.3%	22.2%	10.9%	
Northeast	27.1%	32.3%	25.7%	
Northwest	20.1%	21.2%	19.8%	
Southeast	11.7%	11.1%	11.9%	
Southwest	27.8%	13.1%	31.7%	

¹The weighted sample results are representative of the population of children enrolled in DPP in Fall 2009.

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