PLAY TO LEARN EVALUATION

2012-2018 REPORT

Prepared by Sarah Faude
M.S.Ed, PhD Candidate

WWW.COUNTDOWNTOKINDERGARTEN.ORG/PLAY-TO-LEARN-GROUPS

Table of Contents

EXECUTIVE SUMMARY

The Play to Learn program ("Playgroup") is a Boston Public Schools (BPS) initiative organized and run through the Countdown to Kindergarten department. Playgroup is broadly designed to support children, families, and schools to promote school readiness. It serves children ages 1-3 in local communities, exposing them to developmentally appropriate activities that are comparable to those offered in high quality formal school settings such as those offered in the K0-K2 classrooms in BPS.

Purpose

This Play to Learn Evaluation asks: Is participation in the Play to Learn program a predictor of student academic success? This evaluation looks at the impact of Playgroup on the acquisition of early literacy and reading skills specifically and student performance more broadly, broken down by subgroup.

Scope

Since the 2012-2013 school year, over 20,000 K2 students have taken the beginning-of-year DIBELS test in Boston Public Schools, about 3% of whom had previously participated in the Play to Learn program. This report therefore compares the demographics and academic performance of the children enrolled in Boston Public Schools between those who participated in the Play to Learn program and those who did not. To do so, it provides statistical comparisons of demographics, K1 participation rates, beginning-of-year DIBELS results for K2 students, and MCAS/PARCC data across six years of data.

Overview

The 2012-2018 Play to Learn Evaluation is broken into five central parts:

(1) A brief overview of the methods used to collect and analyze the data in this report.

(2) An analysis of the demographics of Playgroup vs. Non-Playgroup students.

(3) An analysis of academic performance of Playgroup vs. Non-Playgroup students.

(4) An analysis of whether or not Playgroup participation predicts better academic outcomes.

(5) Evaluation recommendations.

Findings

1. Playgroup and Non-Playgroup students are different in several significant ways.

    a. Playgroup students are disproportionate in their share of:
        i. Latinx, White, and students identifying as Other
        ii. K1 participants
        iii. Arabic and Spanish speakers
        iv. LEP status students
        v. Non-Special Needs students
        vi. Non-Free or Reduced Priced Lunch students

2. How much Playgroup participation matters varies by standardized test.
    a. DIBELS
        i. Playgroup is the second best predictor of high performance, after speaking English. Playgroup students are 1.5 times more likely to be high performing as their Non-Playgroup peers.
    b. MCAS
        i. Playgroup participation is only statistically significant for the Math MCAS. For Math, Playgroup is the second best predictor of high performance, after identifying as Asian. Playgroup students are more than two times more likely to be high performing as their Non-Playgroup peers.
    c. PARCC
        i. Playgroup participation is not statistically significant on either PARCC test. This is likely due to small sample size since the exam was only administered for one year.

Evaluation Recommendations

1. Recruitment
    - Intentionally increase Playgroup participation for students who are members of historically underperforming subgroups on key standardized tests such as DIBELS, MCAS, and PARCC.
2. Student Achievement Metrics
    - Include other student performance indicators in future evaluations: student attendance data such as absences and latenesses; student behavioral data such as suspension, detention, or expulsion rates; and student academic performance data such as GPA or end of year grades in key subjects such as English and Math.
3. School System Readiness
    - Include other enrollment and assignment data beyond K1 enrollment, such as registration round or school quality.

Part 1: Methods

Data

This report includes data from six school years, beginning with the 2012-2013 school year and through the 2017-2018 school year. All students for whom there was complete demographic data and who took the beginning-of-year DIBELS test are included in this study. Due to inconsistencies with demographic data in the DIBELS record, student demographic data was pulled through the district's student information server, Aspen, and linked using student ID numbers.

This report builds off of an earlier report that examined 2012-2015 student data.[1] In an effort to increase sample sizes and illuminate longer-term trends, this report compares students in those three years (hereafter Cohort 1) to those students in the subsequent Cohort (hereafter Cohort 2) from 2015-2018. Nevertheless, given that only about three percent of all BPS students participate in Playgroups, sample size limitations persist. As a result, sections of this report necessarily speak only to trends in data rather than statistically significant differences.

Domains

Race and Ethnicity

The district data used for this report combines data on ethnicity and race. District data also categorizes students as Hispanic, and this is counted as a separate category from other racial categories. As a result, students who identify as Hispanic *and* Black are counted in this report as Hispanic only, as they would in other district data collection. Using more inclusive language, I refer to all "Hispanic" populations as Latinx in this report.

K1 Participation

While all students who attend the district's K2 are included in this evaluation, as of the 2017-2018 school year, students who attended K1 represented approximately half of all future K2 students in Boston Public Schools.[2] As the district and the city have committed to the

[1] Due to shifts in data management within the district, this evaluation report uses a newly compiled dataset of all years and variables. As a result, the descriptive statistics and analyses from 2012-2015 are similar, but not identical to the original report. While overall trends are consistent across both reports, particularly in demonstrating the value of the Play to Learn program on student achievement, the specific figures from that report are not referenced here. Instead, new analysis was conducted, and is available upon request.

[2] https://www.northeastern.edu/csshresearch/bostonarearesearchinitiative/wp-content/uploads/sites/2/2018/07/Final-Evaluation-of-Equity-in-BPS-HBAP.pdf

expansion of high quality K1 seats for students,[3] participation in K1 is increasingly an important indicator of school readiness. K1 participation is therefore a predictor of both academic preparedness in the future, as well as a measure of strong parent information networks and engagement at the time of enrollment.

Sex

Student sex is measured by the district as a binary: either male or female.

Language[4]

Student language is measured here through families' reports of their "first language," which is taken from the Home Language Survey administered at the time of registration by the district. Comparisons of student language are reported here in two ways. Language is first analyzed by categorizing the ten district languages into English speaking, Spanish speaking, and those students who speak any other language. Given that the majority of students across six years of data speak English (62.3%), with Spanish as the second largest language group (22.7%), all eight remaining languages combined amount to only 14.9% of all students. All ten district languages are also analyzed in order to show both changes over time and data on specific linguistic communities.

Limited English Proficiency (LEP)

In addition to students' home languages, the district also collects data on students' proficiency with the English language. This status is categorized in three different ways: (F) or (FLEP) for Formerly Limited English Proficient; (L) or (LEP) for Limited English Proficient; and (N) or (Never-LEP) for Never Limited English Proficient. Because this status is only for students who ever tested into this service, this is a subgroup of the overall sample of K2 students covered by this evaluation. However, this subgroup is important to examine as they show the intersection of linguistic diversity, academic proficiency, and targeted supports within the district.

Special Needs

Students with special needs are typically designated through a combination of teacher recommendation and an evaluation process conducted by the school and sometimes additional external testing. Following this, schools and families develop and agree on an Individualized

---

[3] https://www.bostonglobe.com/metro/2017/01/17/marty-walsh-tries-again-for-pre-kindergarten-for-all/qmqXRBHwSO2mHg0rV6FKEO/story.html

[4] There are several different ways that the district measures the diversity of languages spoken by BPS families. Historic reports on Play to Learn groups used a variable called "Home Language," which the district also collects when families first register with the district. While both measures are self-reported by parents, "Home Language" often undercounts the linguistic diversity of families. To better capture the range of languages spoken, and in particular to understand the role of language and language status with families as a demographic variable, this evaluation uses "First Language."

Education Plan (IEP) to determine the level of and type of student needs as well as the supports they are entitled to by law under the Individual with Disabilities Education Act (IDEA).

For this report, special needs is reported at the time of data collection (May 2018) and not at the time students took their K2 beginning-of-year DIBELS assessment due to missing data. Therefore the comparisons below are for students' *current* special needs status, and not their historical status.

Free/Reduced Priced Lunch

In Boston, an estimated 78 percent of students qualified for free and reduced priced lunch. Despite this, nearly twenty percentage points fewer students actually received these benefits. In the 2016-2017 school year, the district[5] shifted to providing free lunch to all students in partnership with the state of Massachusetts and the Community Eligibility Provision (CEP).[6]

This analysis does not distinguish between those students who receive free lunch and those who receive reduced priced lunch, but looks at both together compared to those students who do not. Future evaluations will need to shift away from looking at this domain and instead look at different measures of socioeconomic need.

DIBELS

The Dynamic Indicators of Basic Early Literacy Skills (DIBELS) test assesses the acquisition of early literacy skills and is designed to be a formative assessment tool to evaluate the effectiveness of interventions. Though DIBELS is not designed to be the sole measure of a child's academic performance, it is designed to measure a child's readiness for reading, one of several important predictors of later academic success. Based on their performance on this assessment, each student receives one of four designations: Above Benchmark; Benchmark; Below Benchmark; and Well Below Benchmark.

BPS administers the DIBELS district-wide in grades K2 to 2, up to three times per year. This evaluation is limited to K2 results obtained from the beginning-of-year assessment, when data was most complete. Student achievement is shown both for the four designations as well as for two composites, Above/At Benchmark and Below/Well Below Benchmark, which serve as proxies for high and low achievement.

MCAS

MCAS has been the standardized test of choice in Massachusetts and Boston for many years, with the exception of PARCC's administration during the 2014-2015 school year. This is not to

---

[5] https://www.bostonpublicschools.org/site/Default.aspx?PageType=3&DomainID=4&PageID=1&ViewID=047e6be3-6d87-4130-8424-d8e4e9ed6c2a&FlexDataID=3160

[6] http://www.doe.mass.edu/cnp/nprograms/cep/

say that the test has remained entirely the same however, as post-PARCC MCAS testing is now called "MCAS 2.0" and reflects many of the efforts to increase higher level thinking questions on the state assessment. In both MCAS and MCAS 2.0, hereafter referred to as just MCAS, each student receives one of four designations: Exceeding Expectations; Meeting Expectations; Partially Meeting Expectations; and Not Meeting Expectations.

This evaluation uses MCAS data from 2 years: 2015-2016; 2016-2017. While this is less than previous evaluations, this evaluation directly compares how individual students who were in playgroup performed on these assessments. This is the most accurate and robust way to assess whether or not Playgroup participation matters for high achievement on this assessment. Student achievement data is provided both for the four designations as well as for two composites, Exceeding/Meeting Expectations and Partially Meeting/Not Meeting Expectations, which serve as proxies for high and low achievement.

PARCC

The Partnership for Assessment of Readiness for College and Careers (PARCC) is a standardized assessment that is computer-based and used to assess Math and English Language Arts (ELA) for grades 3-11. It was administered in 13 states, aligned to the Common Core, and assessed higher level thinking. PARCC entails both a Performance Based Assessment and End of Year Assessment. Based on their performance on this assessment, each student received one of five designations: Exceeds Expectations; Met Expectations; Approached Expectations; Partially Met Expectations; and Did Not Meet Expectations.

PARCC was administered in Boston during the 2014-2015 school year, but only for grades 3-8 in the subjects of ELA and Math. This evaluation is therefore limited to one year of data for PARCC: the 3rd grade PARCC results for one year of students. Student achievement is shown for all of the five designations as well as for two composites, Exceeds/Met Expectations and Approached/Partially Met/Did Not Meet Expectations, which serve as proxies for high and low achievement.

Analysis

Descriptive Statistics

Descriptive statistics were conducted on all data. Each section that follows provides analyses that include both tables and charts, all of which are divided by Non-Playgroup and Playgroup students for ease of comparison. Within tables, descriptive statistics indicate the number of students represented by subgroup (N), the total number of students represented, as well as the percentage of students represented within either Non-Playgroup or Playgroup categories. For example, in the section "All Language," a percentage would represent the percentage of

Playgroup students who speak Arabic, rather than the percentage of district Arabic speakers in Playgroup.

Correlation Statistics

Significance testing was conducted on all data when sample sizes allowed,[7] using Pearson's Chi-Square. Asterisks (*) in the report correspond with the significance level, where $p < .05$=*, $p < .01$=**, and $p < .001$=***. This should be interpreted as follows: when a difference is statistically significant at the $p < .05$ level, it indicates that there is a *less than* 5% chance that the difference is due to randomness. The higher the significance, the more confident we can be that the findings are due to the differences between the two groups (here, participating in Playgroup) rather than random chance.

Inferential Statistics

For achievement data, a binary logistic regression was computed to determine the significance level (if any) of Playgroup participation on academic performance when controlling for demographic differences between Playgroup and Non-Playgroup students.

Comparing Different Standardized Assessments

While the three broader assessments in this evaluation, DIBELS, MCAS, and PARCC, do not use the same scoring thresholds or categories, both are normed and grouped into different qualitative categories that allow for broader comparison. Although they are different, DIBELS and MCAS are both grouped into four categories, while PARCC is grouped into five categories. To account for these variations and still offer comparison, each test's categories have been grouped to account for high and low achievement. This is reflected in the descriptive, correlational, and inferential statistics.

[7] Significance testing was not conducted in all comparisons due to small sample sizes of playgroup participants, particularly when divided into subgroups. Small sample sizes decrease the likelihood that a comparison can be found to be statistically significant.

Part 2: Student Demographics Findings

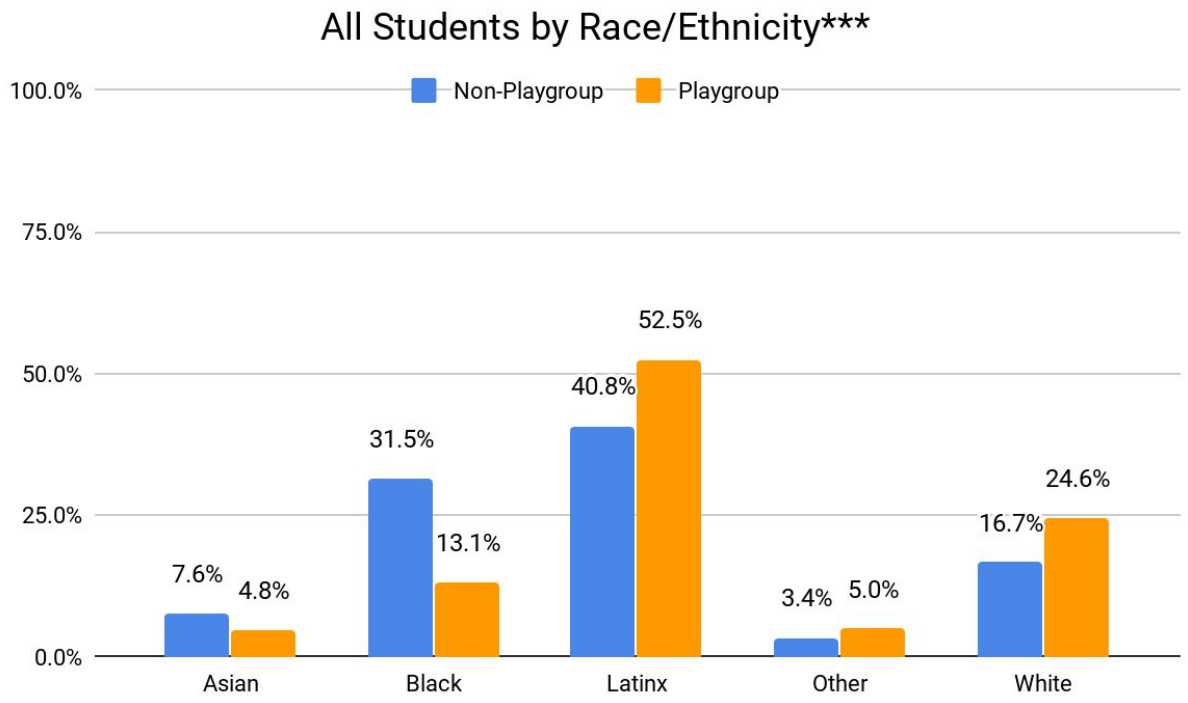
Part two of this evaluation compares the demographics and characteristics of Playgroup students to those of their Non-Playgroup peers across seven domains: Race and Ethnicity, K1 Participation, Sex, Language, Limited English Proficiency Status, Special Needs Status, and Free/Reduced Priced Lunch. In each section that follows, informed by Cohort level and aggregate analyses, are descriptions of trends over six years of data as well as notable shifts across Cohorts when relevant and/or significant.

This analysis reveals that over the past six years, the characteristics of Playgroup students have been statistically significant for several domains when compared to their Non-Playgroup peers. In particular, compared to Non-Playgroup students, Playgroup students are disproportionately Latinx, White, or of another race or ethnicity; enrolled in K1; Arabic and Spanish speakers; have a Limited English Proficiency (LEP) status; and less often identified as having Special Needs or receiving Free or Reduced Priced Lunch.

The rest of part two is broken down by domain. Findings are described alongside each table and chart for ease of comparison.

Race and Ethnicity

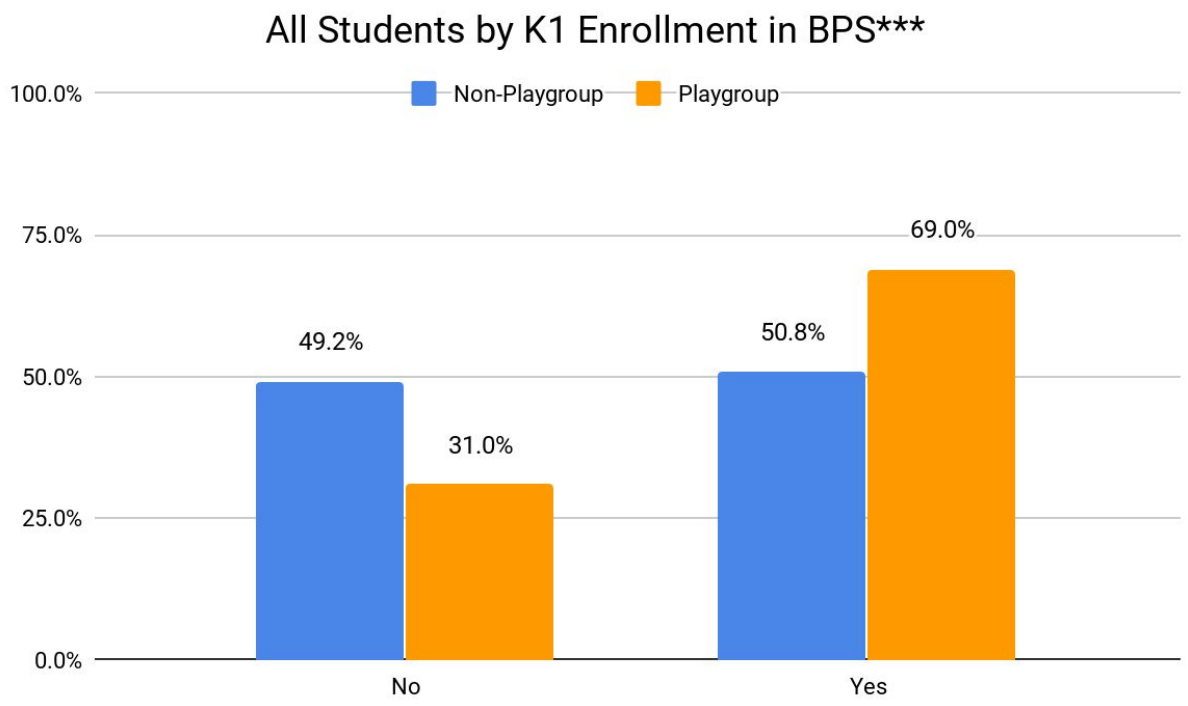
The race and ethnicity of Playgroup vs Non-Playgroup students is statistically significant at the p< .001 level for Cohort 1, Cohort 2, and across all six years. The table and chart for all six years of data is displayed below. Specifically, Playgroup has an overrepresentation of students who identify as Latinx (p< .001), White (p< .001), and Other (p< .05), and an underrepresentation of students who identify as Asian (p< .01) and Black (p< .001). While this has persisted across the two cohorts, the gaps in representation have sharply increased over time for Asian students, Black students, and most dramatically for White students.



	Non-Playgroup		Playgroup		Statistical Significance
Race	**N**	**%**	**N**	**%**	
Asian	1515	7.6%	30	4.8%	**
Black	6264	31.5%	82	13.1%	***
Latinx	8118	40.8%	328	52.5%	***
Other	668	3.4%	31	5.0%	*
White	3321	16.7%	154	24.6%	***
Total	**19886**	100%	**625**	100%	***

K1 Participation

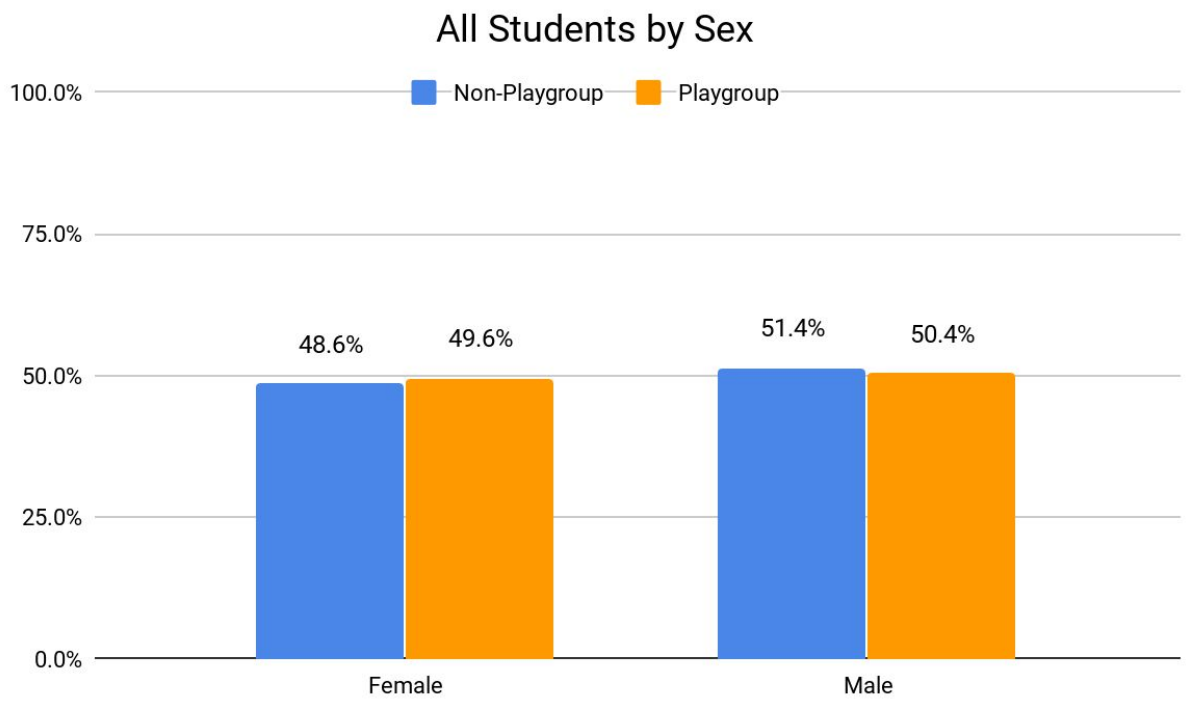
Overall the data show that students who participate in Playgroup are more likely than their Non-Playgroup peers to attend K1 at Boston Public Schools. This finding is statistically significant for Cohort 1 ($p < .01$), Cohort 2, ($p < .001$), and overall across the six years ($p < .001$).



	Non-Playgroup		Playgroup		Statistical Significance
Attend K1	**N**	**%**	**N**	**%**	
No	9785	49.2%	194	31.0%	***
Yes	10101	50.8%	431	69.0%	***
Total	**19886**	100%	**625**	100%	***

Sex

There is no statistically significant difference in student sex between Non-Playgroup and Playgroup participants in Cohort 1, Cohort 2, or across all six years.

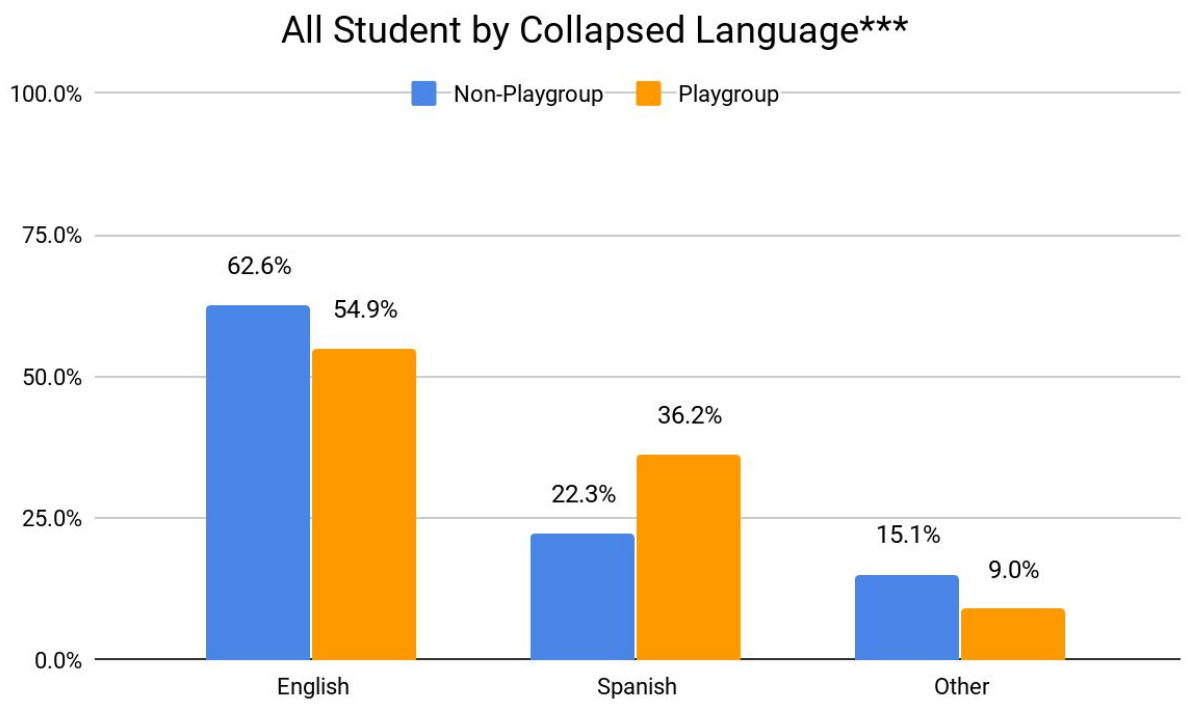


	Non-Playgroup		Playgroup		Statistical Significance
Sex	**N**	**%**	**N**	**%**	
F	9672	48.6%	310	49.6%	not significant
M	10214	51.4%	315	50.4%	not significant
Total	**19886**	100%	**625**	100%	not significant

Language

Language Collapsed

In a collapsed analysis of three groups (English, Spanish, Other), Spanish speaking participants are overrepresented in the Playgroup compared to their Non-Playgroup peers in both cohorts as well as overall across the six years. While statistically significant at the p< .001 level in both cohorts and overall, the differences have decreased between Cohort 1 and Cohort 2, which means that the Non-Playgroup and Playgroup populations have become more similar in recent years in terms of their first language.

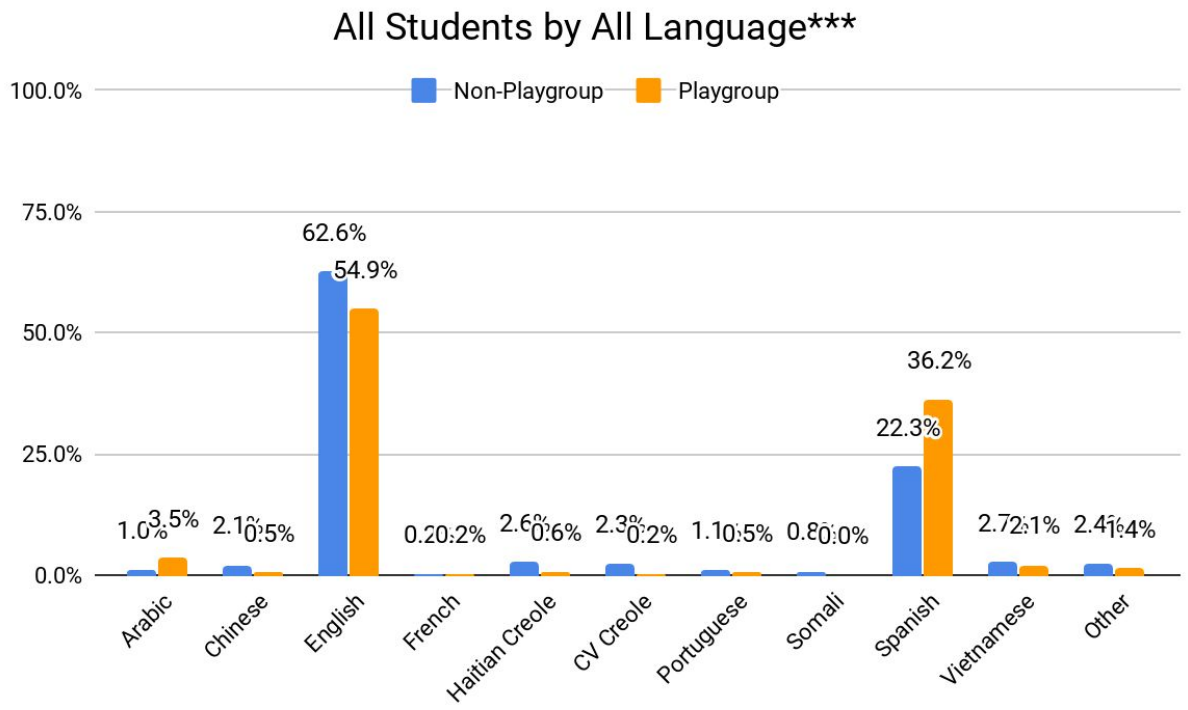


	Non-Playgroup		Playgroup		Statistical Significance
First Language	**N**	**%**	**N**	**%**	
English	12439	62.6%	343	54.9%	***
Spanish	4437	22.3%	226	36.2%	***
Other	3010	15.1%	56	9.0%	***
Total	**19886**	100%	**625**	100%	***

Language All

When looking at all ten linguistic groups over all six years, Playgroup and Non-Playgroup differences are statistically significant at the $p < .001$ level. Given small sample sizes for language groups other than English and Spanish, it is likely that the statistical significance is explained by the variation in these two majority languages, as seen in the previous table and chart.

In particular, Arabic and Spanish speakers are overrepresented in Playgroup compared to Non-Playgroup. All other language groups are underrepresented in Playgroup in aggregate across all six years. The only shift between cohort years with regards to this over/underrepresented divide is with French speakers, who are overrepresented in Cohort 1 (n=1 student) but underrepresented in Cohort 2 (n=0 students). However, given the very small number of Playgroup participants in many of these specific linguistic communities, the difference between overrepresentation and underrepresentation can be the difference in participation by only a handful of students, or even just one as we see with French speakers.



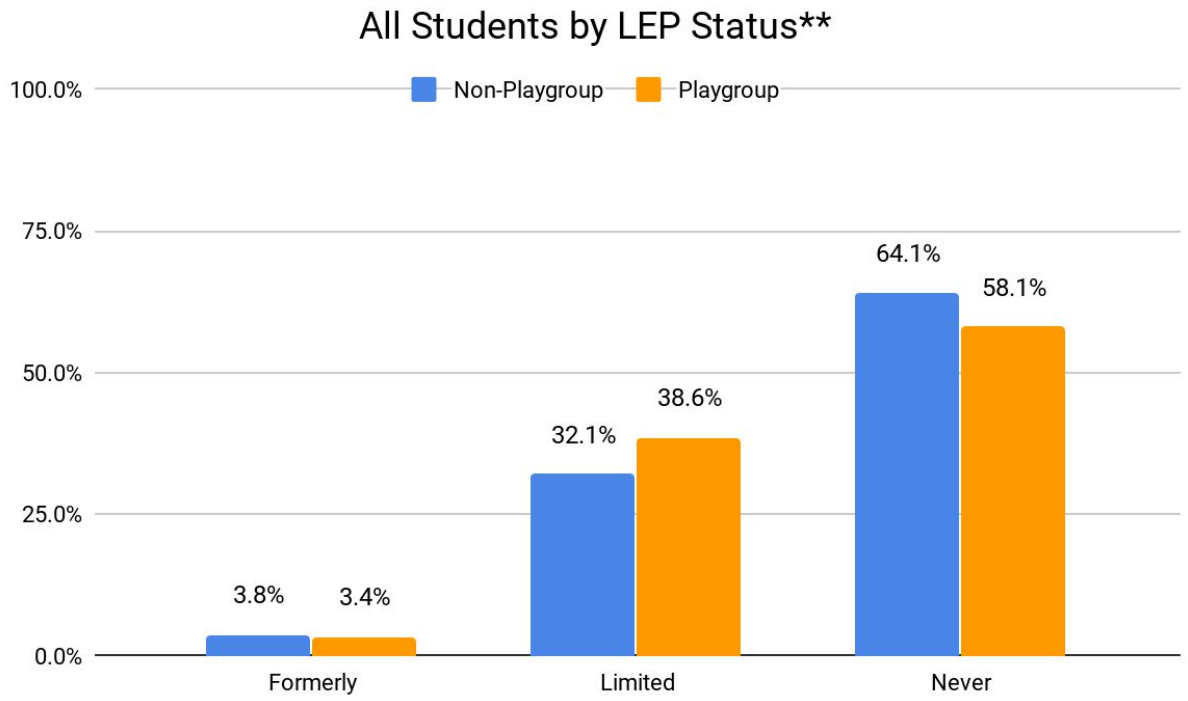
	Non-Playgroup		Playgroup		Statistical Significance
First Language	**N**	**%**	**N**	**%**	
Arabic	192	1.0%	22	3.5%	***
Chinese	421	2.1%	3	0.5%	**
English	12439	62.6%	343	54.9%	***
French	44	0.2%	1	0.2%	not significant
Haitian Creole	512	2.6%	4	0.6%	**
Cape Verdean Creole [8]	451	2.3%	1	0.2%	***
Portuguese	218	1.1%	3	0.5%	not significant
Somali	154	0.8%	0	0.0%	*
Spanish	4437	22.3%	226	36.2%	***
Vietnamese	533	2.7%	13	2.1%	not significant
Other	485	2.4%	9	1.4%	not significant
Total	**19886**	100%	**625**	100%	***

[8] Cape Verdean Creole is represented in charts as "CV Creole" for legibility reasons only.

Limited English Proficiency (LEP)

There is a statistically significant difference between Playgroup and Non-Playgroup students for LEP Status at Cohort 1 ($p < .01$), Cohort 2 ($p < .05$), and across all six years ($p < .01$). While overall a higher proportion of LEP students participated in Playgroup than did not, Playgroup students are also less likely overall to be designated as FLEP or Never-LEP. However within cohorts, Cohort 2 Playgroup students are actually very similar to their Non-Playgroup peers with regards to both LEP status and Never-LEP status, which explains why Cohort 1 and overall are more statistically significant than Cohort 2.

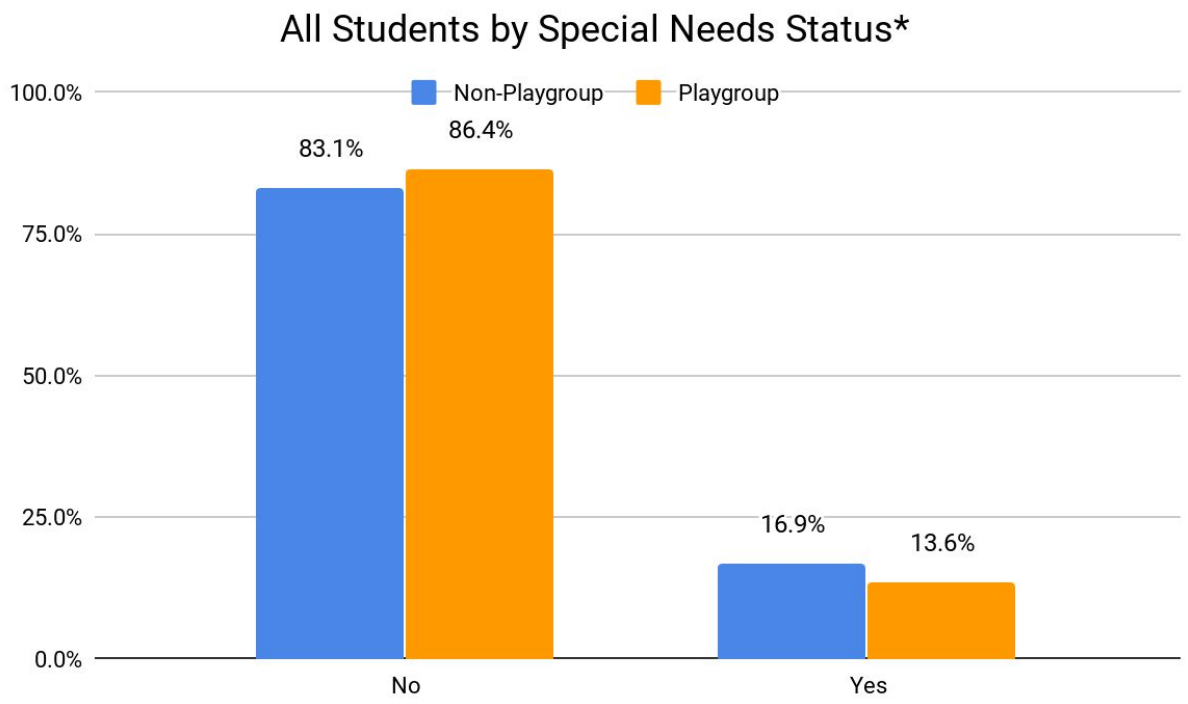
These trends are also likely linked to methodological decisions made for this evaluation. Due to incomplete demographic data collected at the time of the DIBELS test, LEP status was determined pulling current student records in Aspen. Therefore, we should expect that with more time in district programs, more students in Cohort 1 than in Cohort 2 will have achieve the designation of Formerly Limited English Proficient.



	Non-Playgroup		Playgroup		Statistical Significance
LEP Status	**N**	**%**	**N**	**%**	
F	764	3.8%	21	3.4%	not significant
L	6381	32.1%	241	38.6%	**
N	12741	64.1%	363	58.1%	*
Total	**19886**	100%	**625**	100%	**

Special Needs

There is a statistically significant difference between Playgroup and Non-Playgroup participants with regards to their special needs status when looking across all six years ($p < .05$) but not when looking within either individual cohort. The proportion of students with special needs is similar for both Playgroup and Non-Playgroup students, with a slightly higher proportion of special needs students not participating in Playgroup. This difference was most pronounced in Cohort 1, and the gap in representation was nearly halved by Cohort 2. This suggests that Playgroup and Non-Playgroup students have become more similar in this regard over time.

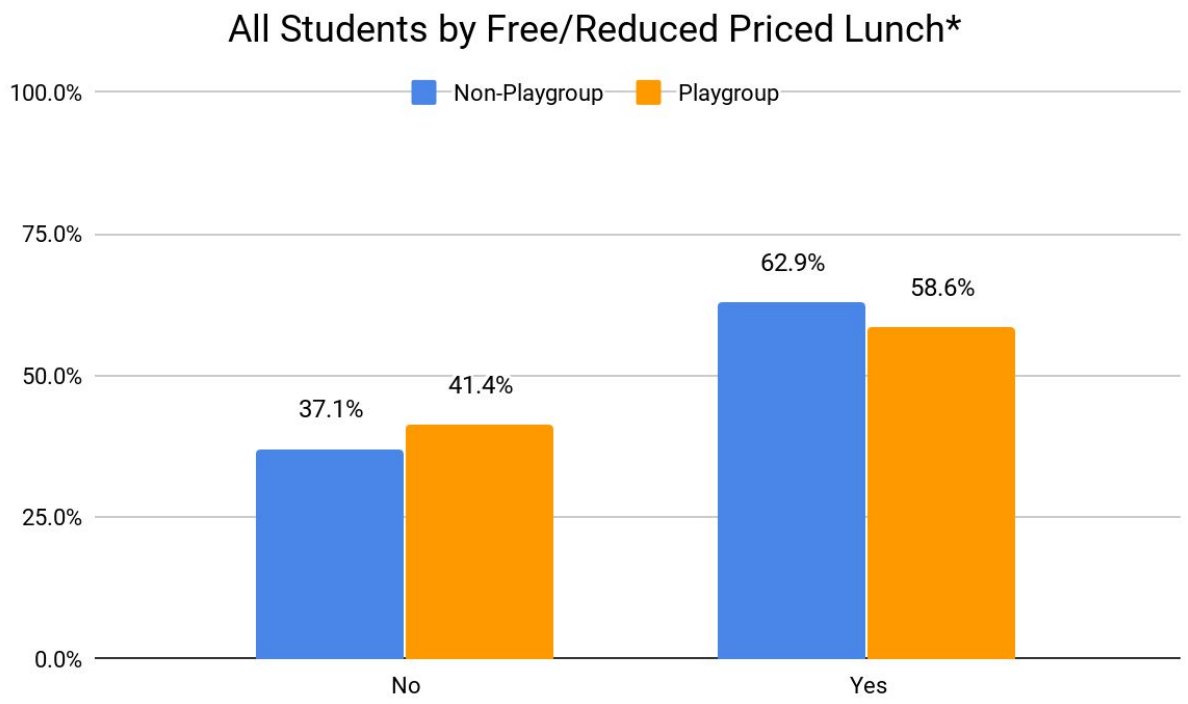


	Non-Playgroup		Playgroup		Statistical Significance
Special Needs	**N**	**%**	**N**	**%**	
No	16520	83.1%	540	86.4%	*
Yes	3366	16.9%	85	13.6%	*
Total	**19886**	100%	**625**	100%	*

Free/Reduced Price (F/RP) Lunch

Across all six years, we see that the difference between free or reduced priced lunch status of Non-Playgroup and Playgroup students is significant at the $p < .05$ level. Playgroup students are less likely than their Non-Playgroup peers to receive free or reduced priced lunch.

When the data is disaggregated by Cohort, each Cohort shows a different trend. In Cohort 1, Playgroup students were less likely to receive free or reduced priced lunch (55.8%) compared to Non-Playgroup Students (62.9%), although it is not statistically significant. In Cohort 2, however, Playgroup students are more frequently free and reduced priced lunch recipients (60.4%) than their Non-Playgroup peers (58.6%), and this difference is statistically significant at the $p < .01$ level. This suggests that while in aggregate, Playgroup students are more privileged in this regard, if recruitment trends mirror Cohort 2, this will likely shift in future years.



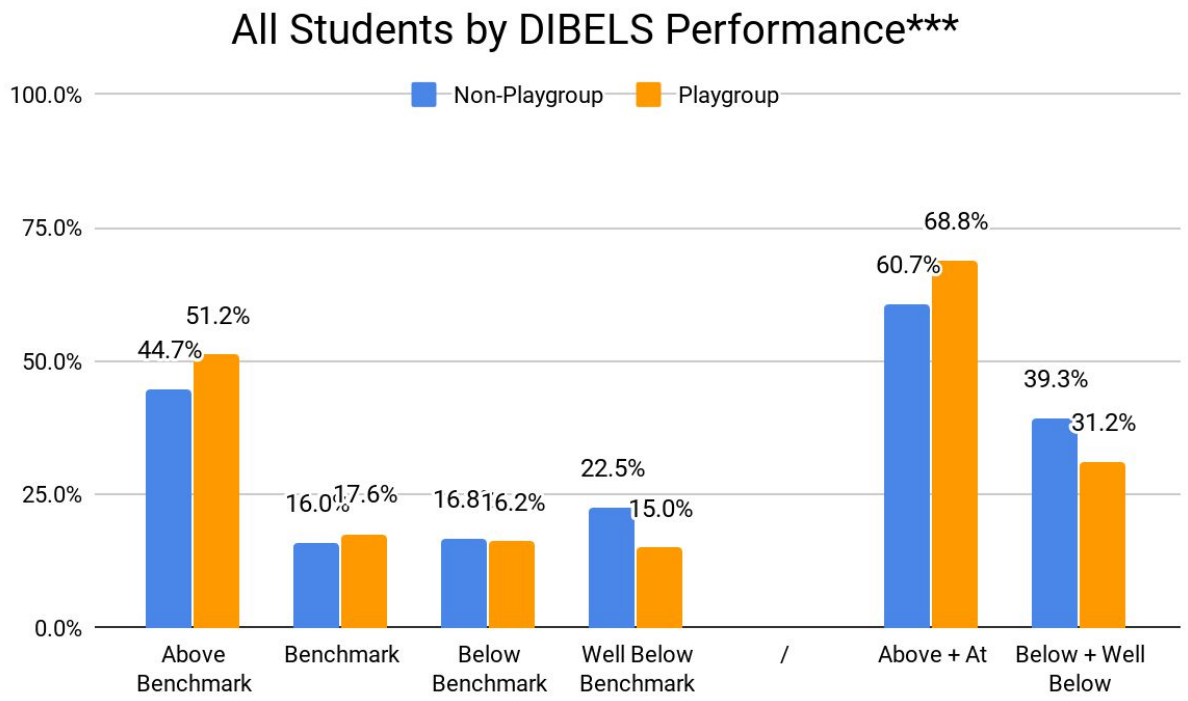
	Non-Playgroup		Playgroup		
Receive F/RP Lunch	**N**	**%**	**N**	**%**	**Statistical Significance**
No	7377	37.1%	259	41.4%	*
Yes	12509	62.9%	366	58.6%	*
Total	**19886**	100%	**625**	100%	*

Part 3: Student Academic Performance Findings

DIBELS

Overall, student DIBELS performance is statistically significant at the $p < .001$ level, with Playgroup students outperforming their Non-Playgroup peers. This persists when looking at combined categories of Above Benchmark and Benchmark as well as the combined categories of Below Benchmark and Well Below Benchmark. However, the data indicate that this significance is driven by the two most extreme categories of Above Benchmark and Well Below Benchmark, since the middle categories are not statistically significant.

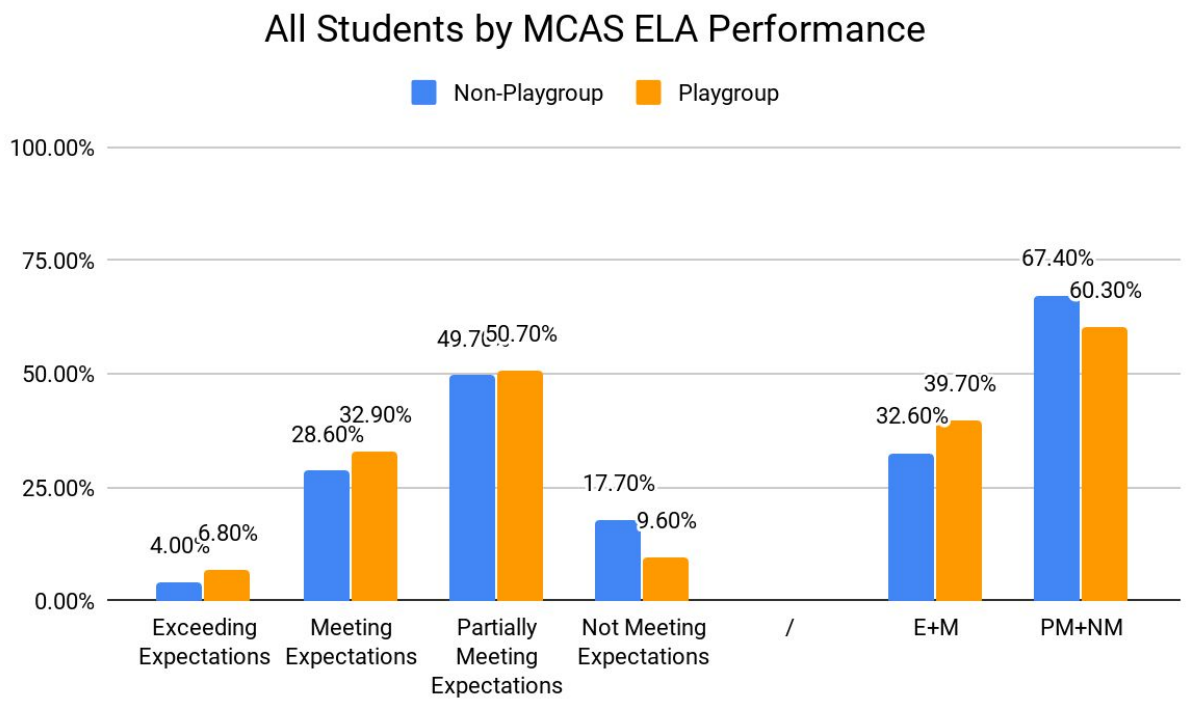
When looking within cohorts, Cohort 1 Playgroup participants had nearly identical rates of Above Benchmark performers as their Non-Playgroup peers, but were overrepresented in their proportion of Below Benchmark students. By Cohort 2, however, Playgroup students substantially outperform their peers; they are overrepresented in Above Benchmark and underrepresented in Well Below Benchmark categories. The pattern holds true at the Benchmark and Below Benchmark levels as well, indicating that Playgroup students by Cohort 2 are much more successful on this predictive assessment than their Non-Playgroup peers and even their Cohort 1 predecessors.



	Non-Playgroup		Playgroup		Statistical Significance
DIBELS Results	**N**	**%**	**N**	**%**	
Above Benchmark	8890	44.7%	320	51.2%	**
Benchmark	3176	16.0%	110	17.6%	not significant
Below Benchmark	3331	16.8%	101	16.2%	not significant
Well Below Benchmark	4475	22.5%	94	15.0%	***
A+A Total	12066	60.7%	430	68.8%	***
B+WB Total	7806	39.3%	195	31.2%	***
Total	**19872**	100%	**625**	100%	***

MCAS ELA

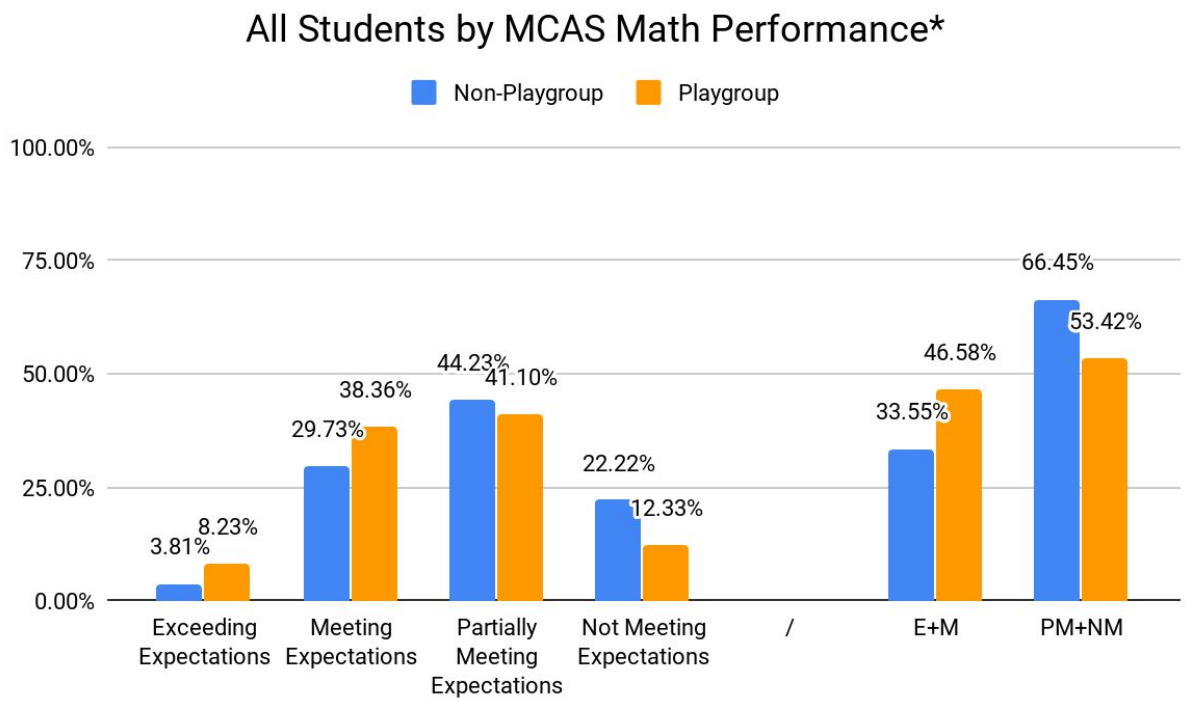
Overall, student ELA MCAS performance is not statistically significant, likely due to sample size of Playgroup students compared to their Non-Playgroup peers. When looking at both sets of combined categories, they are each significant at the p<.001 level, demonstrating that Playgroup students outperform Non-Playgroup students on the ELA MCAS exam. Data indicate that this significance is driven most by differences between lower performing Playgroup and Non-Playgroup students, as the difference between those who are Exceeding Expectations is not statistically significant.



	Non-Playgroup		Playgroup		Statistical Significance
MCAS ELA Results	**N**	**%**	**N**	**%**	
Exceeding Expectations	208	4.0%	5	6.8%	not significant
Meeting Expectations	1486	28.6%	24	32.9%	**
Partially Meeting Expectations	2582	49.7%	37	50.7%	***
Not Meeting Expectations	918	17.7%	7	9.6%	***
E+M Total	1694	32.6%	29	39.7%	***
PM+NM Total	3500	67.4%	44	60.3%	***
Total	**5194**	100.0%	**73**	100.0%	not significant

MCAS Math

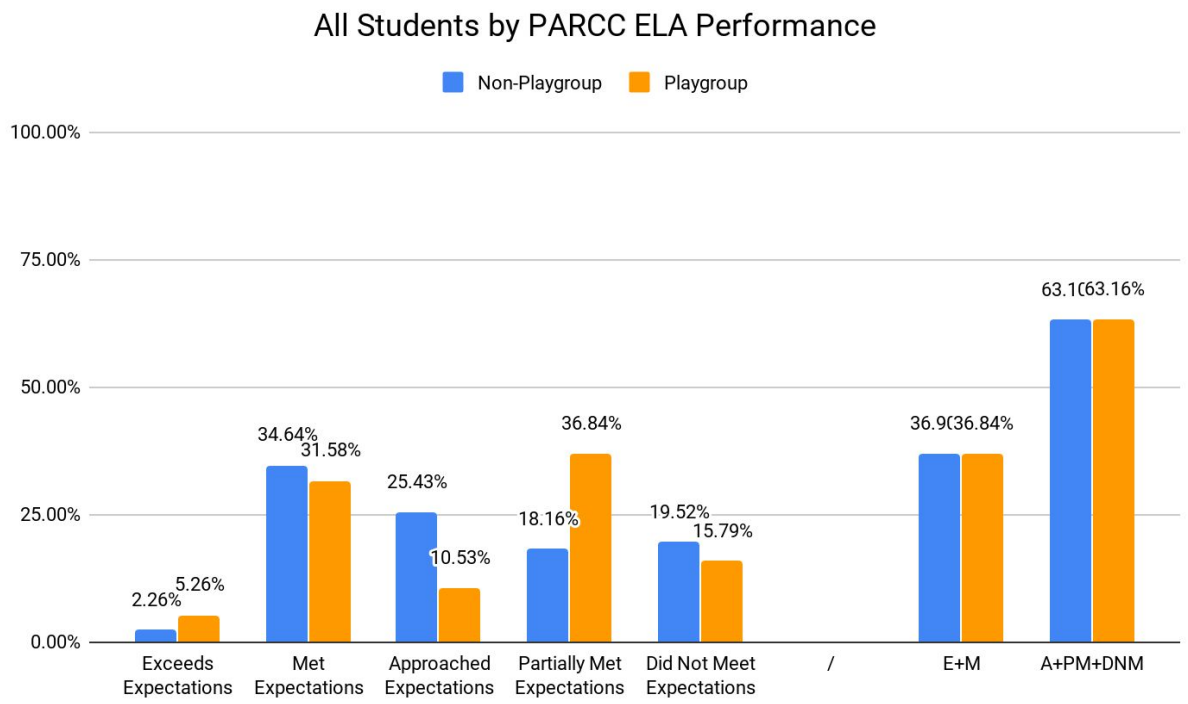
Overall, student Math MCAS performance is significant at the p< .05 level. When looking at high performers (those Exceeding and Meeting Expectations) versus low performers (those Partially Meeting or Not Meeting Expectations), the differences between Playgroup and Non-Playgroup are statistically significant for each at the p< .001 level. While those Exceeding Expectations is not significant, likely due to small sample size, it is clear from the data that Playgroup students outperform their Non-Playgroup peers on this assessment.



	Non-Playgroup		Playgroup		Statistical Significance
MCAS Math Results	**N**	**%**	**N**	**%**	
Exceeding Expectations	198	3.81%	6	8.23%	not significant
Meeting Expectations	1544	29.73%	28	38.36%	**
Partially Meeting Expectations	2297	44.23%	30	41.10%	***
Not Meeting Expectations	1154	22.22%	9	12.33%	***
E+M Total	1742	33.55%	34	46.58%	***
PM+NM Total	3451	66.45%	39	53.42%	***
Total	**5193**	100.00%	**73**	100.00%	*

PARCC ELA

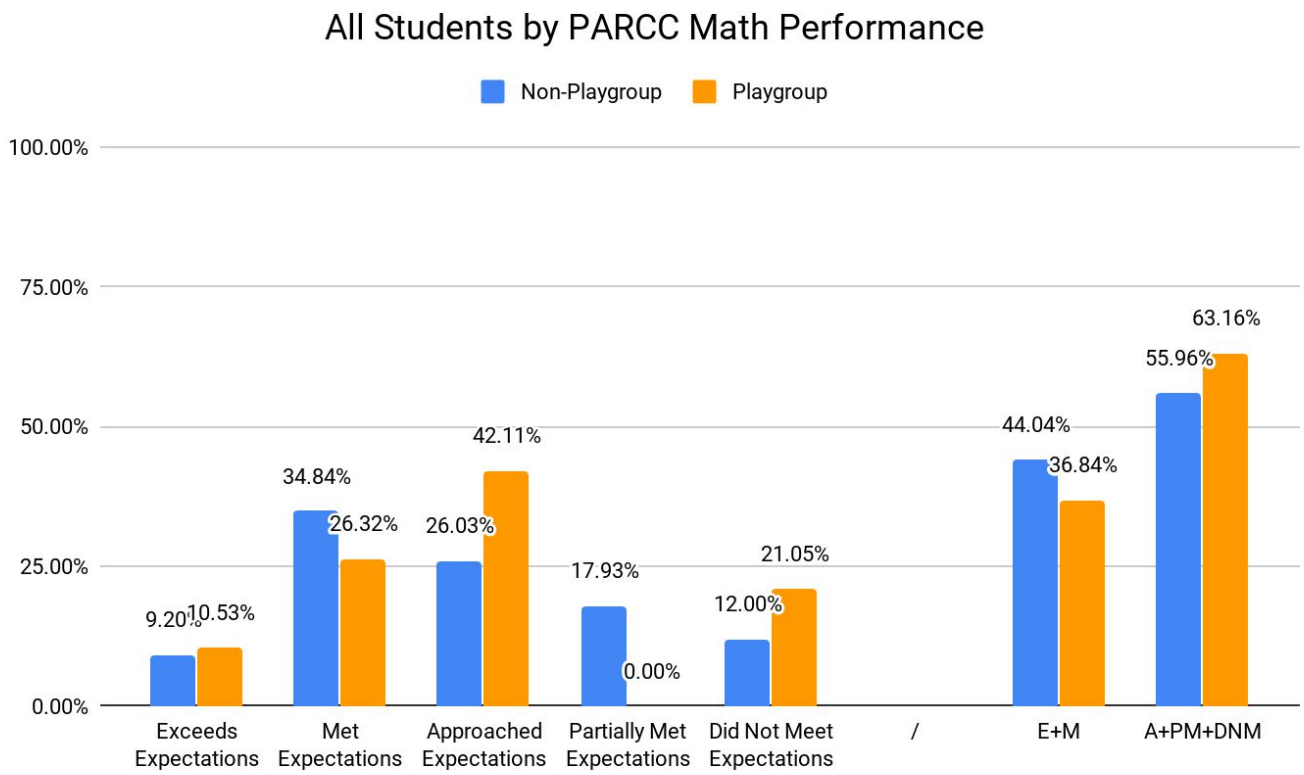
Overall, student ELA PARCC results are not significant and there is less consistent evidence of high Playgroup performance than on the ELA MCAS. However, both findings are likely due to the small sample size of Playgroup participants since PARCC was only administered for one year. In addition, the top category of Exceeds Expectations is not statistically significant. Although the collapsed groups of high performers (Exceeds and Met Expectations) and low performers (Approached, Partially Met, and Did Not Meet Expectations) appear nearly identical, there is a statistically significant difference between them at the p< .001 level, with Playgroup students slightly underperforming their Non-Playgroup peers.



	Non-Playgroup		Playgroup		Statistical Significance
PARCC ELA Results	**N**	**%**	**N**	**%**	
Exceeds Expectations	58	2.26%	1	5.26%	not significant
Met Expectations	891	34.64%	6	31.58%	***
Approached Expectations	654	25.43%	2	10.53%	***
Partially Met Expectations	467	18.16%	7	36.84%	*
Did Not Meet Expectations	502	19.52%	3	15.79%	**
E+M Total	949	36.90%	7	36.84%	***
A+PM+DNM Total	1623	63.10%	12	63.16%	***
Total	**2572**	100.0%	**19**	100.0%	not significant

PARCC Math

Overall, student Math PARCC results are not significant, and there is less consistent evidence of high Playgroup performance than on the Math MCAS. However, both findings are likely due to the small sample size of Playgroup participants since PARCC was only administered for one year. In addition, the collapsed group of high performers (Exceeds and Met Expectations), the collapsed group of low performers (Approached, Partially Met, and Did Not Meet Expectations), and the bottom category of Did Not Meet Expectations are all not statistically significant. However, the top four categories are, albeit with mixed findings. A higher proportion of Playgroup students are in the top category of Exceeds Expectations and Approached Expectations, but are underrepresented in the Met Expectations category.



	Non-Playgroup		Playgroup		Statistical Significance
PARCC Math Results	**N**	**%**	**N**	**%**	
Exceeds Expectations	236	9.20%	2	10.53%	*
Met Expectations	894	34.84%	5	26.32%	***
Approached Expectations	668	26.03%	8	42.11%	*
Partially Met Expectations	460	17.93%	0	0.00%	***
Did Not Meet Expectations	308	12.00%	4	21.05%	not significant
E+M Total	1130	44.04%	7	36.84%	not significant
A+PM+DNM Total	1436	55.96%	12	63.16%	not significant
Total	**2566**	100.0%	**19**	100.0%	not significant

Part 4: Does Playgroup Participation Predict Better Academic Outcomes?

Rationale and Approach

Given the statistically significant findings on DIBELS, MCAS, and PARCC performance *and* the statistically significant differences between Playgroup and Non-Playgroup students themselves, it is important to assess whether the differences in academic performance on the different assessments can be explained by the difference in group demographics. A binary logistic regression examines if participation in Playgroup is still a significant factor that explains student performance on each test.

For each assessment, performance categories were collapsed into high performing and low performing, in order to increase sample size and maximize the power of the analysis. A model was produced for each assessment, and included the demographic variables Sex, Race, Language Collapsed, Special Needs Status, and Free/Reduced Priced Lunch as well as Playgroup participation. The models that follow show the odds ratios[9] and significance levels[10] for each variable.

[9] Odds ratios explain the relative measure of effect, or how many more times likely something will happen for one group than for another.

[10] When a variable is statistically significant in the model, it means that it impacts academic performance *even after* controlling for the other variables in the model.

Findings

DIBELS

Playgroup participation predicts better DIBELS performance.
The model shows that participating in Playgroup is a strong predictor of high performance on DIBELS, second only to speaking English. Playgroup students have an odds ratio of 1.54 and it is statistically significant at the $p < .001$ level. The odds ratio indicates that for two students who are identical in terms of the demographics variables in the model, the student who participated in Playgroup is more than one and a half times more likely to perform well on DIBELS than their Non-Playgroup counterpart.

Speaking English is the strongest predictor of high performance on DIBELS as it is significant at the $p < .001$ level and has an odds ratio of 1.77. This suggests that if there are two students who are otherwise identical, the English speaking student will be 1.77 times more likely to perform Above or at Benchmark on the DIBELS test than the other student. All variables in this model are statistically significant, with the exception of speaking Spanish. While Spanish is not significant, it is of note that Latinx identity is, with these students (all other things equal) still perform only half as well than their non-Latinx counterparts. Therefore, while Playgroup participation makes a huge difference in DIBELS performance among similar children, it does not close opportunity gaps that exist between different demographic groups.

Characteristics Predicting High Performance on DIBELS

Variable	Odds Ratio
Sex (Female)	1.14***
F/R Priced Lunch (Yes)	.63***
Special Needs (Yes)	.37***
Race	
Asian	1.27*
Black	.79*
Latinx	.58***
White	1.24*
Language	
English	1.77***
Spanish	.90
Playgroup (Yes)	1.54***

MCAS

Playgroup participation predicts higher Math, but not ELA performance on MCAS.
The model below looks at ELA and Math MCAS performance separately. While the odds ratio for Playgroup students indicates a positive trend in performance on both tests, this difference is

only statistically significant for Math performance. For Math, Playgroup performance is the second highest predictor for high performance after identification as Asian, with an odds ratio of 2.02 and a significance level of p< .01. Therefore Playgroup students are two times more likely to be high performers on the Math MCAS than their otherwise identical Non-Playgroup peers.

Some variables are similarly predictive of ELA and Math. Free and reduced priced lunch status and special needs status are both statistically significant in this model and have similar odds ratios for both tests; they are less than half and less than one fifth as likely as their peers to be high performers on the MCAS. Speaking Spanish is also predictive of lower performance on both MCAS tests, with Spanish speakers about two thirds as likely as their English speaking peers to be high performers. Sex as a predictor flips between tests; female students are more likely to be high performers on the ELA test and male students are more likely to be high performers on the Math test.

For ELA performance, White and Asian identification are the most predictive variables of high ELA performance, with odds ratios of 2.41 and 2.40 respectively, both significant at the p< .001 level. This means that students with these racial groups, all other things equal, are nearly two and a half times more likely than their Black, Latinx, or Other peers to be high performers on the MCAS ELA assessment. Conversely, Black students are half as likely as their other identified peers to perform high on the MCAS ELA.

For Math performance, all variables except for speaking English are statistically significant. Identifying as Asian is the best predictor and these students, all variables held equal, are nearly four times more likely than their peers to be high performers on the Math MCAS. Black and Latinx students both are less likely to be high performers on the Math MCAS than they were on the ELA MCAS.

Characteristics Predicting High Performance on MCAS

	ELA MCAS	Math MCAS
Variable	Odds Ratio	Odds Ratio
Sex (Female)	1.24**	.82**
F/R Priced Lunch (Yes)	.48***	.47***
Special Needs (Yes)	.14***	.17***
Race		
Asian	2.40***	3.87***
Black	.55**	.40***
Latinx	.85	.60**
White	2.41***	1.66**
Language		
English	1.18	.98
Spanish	.67**	.67**
Playgroup (Yes)	1.53	2.02**

PARCC

Playgroup participation does not correlate with higher PARCC performance.

The model below presents models that look at ELA and Math PARCC performance separately. While the odds ratio for Playgroup students indicate positive performance on both tests, particularly Math, this difference is not statistically significant. This may be due to small sample sizes as PARCC was only administered for one year.

Across both tests, however, similar although intensified patterns hold for sex, White students, and Asian students as there were for the MCAS subject tests. Male students hold a similar advantage on Math, but female students hold a greater advantage on ELA. Asian students increase their likelihood of high performance on Math to more than five and half times their peers. White students increase their likelihood of high performance on Math to nearly three times that of their peers. Special needs students have similar outcomes on PARCC as they did for MCAS, and are more than eighty percent less likely to be high performing than their otherwise identical peers. Free and reduced priced lunch students fare slightly better on PARCC than MCAS, although still about half as likely to be high performers on both tests.

Characteristics Predicting High Performance on PARCC

	ELA PARCC	Math PARCC
Variable	Odds Ratio	Odds Ratio
Sex (Female)	1.35**	.84*
F/R Priced Lunch (Yes)	.56***	.57***
Special Needs (Yes)	.13***	.15***
Race		
Asian	2.12*	5.53***
Black	.51*	.67
Latinx	.69	.88
White	1.70	2.67**
Language		
English	1.26	.90
Spanish	.68*	.76
Playgroup (Yes)	1.46	1.04

Part 5: Recommendations

Recruitment

Playgroup participation is a strong predictor of high student performance in the BPS. Therefore, bridging this opportunity gap is a worthwhile investment. I therefore recommend that Play to Learn intentionally increase Playgroup participation for students who are members of historically underperforming subgroups on standardized tests.

Student Achievement Metrics

Given variation in standardized test administration, the small sample size of Playgroup makes it hard to determine the predictive power of this intervention on student performance. While it is beyond the purview of the Countdown to Kindergarten program to determine district-wide assessments, future evaluations may want to take other student performance indicators into consideration to understand the short and long term impacts of this initiative on student achievement.

Despite active academic and policy debates on what positively impacts of student learning and engagement, there is more consensus in what predicts school dropout. Johns Hopkins researchers identify that four factors can "identify 60% of the students who would ultimately fail to graduate from the school district."[11] These factors include poor attendance, poor behavioral grades, failing math in sixth grade, and failing English in sixth grade. Therefore, future evaluations of the Play to Learn Program should take some additional variables into consideration: attendance data on student absences and lateness; student behavioral data such as suspension, detention, or expulsion rates; and student academic performance data such as GPA or end of year grades in key subjects English and Math. These will be incredibly valuable measures, especially if standardized tests continue to change and create a challenge of small sample sizes.

School System Readiness

Given the dynamics of Boston's school assignment system and the uneven expansion of high quality, free, pre-kindergarten programs in the district, I recommend also looking at other enrollment and assignment data. In a binary logistic regression of K1 enrollment, using the same demographic variables as presented in Part 3 of this report, Playgroup participation is the *greatest predictor* of K1 enrollment of these variables. In fact, Playgroup participants, all other identifiers held constant, are more than two times as likely to enroll in K1 than their Non-Playgroup peers. Given that K1 enrollment is an important gateway to school enrollment, and on-time school registration is also an equity issue,[12] future evaluations should consider school enrollment patterns by registration round or school quality.

---

[11] https://new.every1graduates.org/wp-content/uploads/2012/03/preventing_student_disengagement.pdf

[12] https://chalkbeat.org/posts/us/2018/06/25/early-school-choice-deadlines-mean-affluent-parents-often-get-first-shot-at-coveted-schools-new-study-shows/