BROWN
The Policy Lab

Longitudinal Evaluation of Providence Public Schools' Multilingual Learners Program

Date: June 30, 2021

Authors[1]: The Policy Lab and the Providence Public School District

[1] Principal Authors: Sarah Thang, Edward Huh, and Kevin H. Wilson. From The Policy Lab: Attiyya Houston, Carycruz Bueno, and Yang Xu. From PPSD: Lindsay Lanteri and Jennifer Efflandt. With special thanks to Jake Bowers, Maya Dayan, and Cotter McCarthy. Additional thanks to Marco Andrade, Claire Hug, Maura Iv, David Yokum, and attendees of The Policy Lab's Demo Day. Communication may be directed to thepolicylab@brown.edu.

Executive Summary

In 2018, the U.S. Department of Justice (DOJ) identified 12 violations of the Equal Educational Opportunities Act of 1974 by the Providence Public School District (PPSD) specifically impacting Multilingual Learner (MLL) students[2]. To avoid litigation, PPSD and DOJ came to a settlement agreement, which outlined specific requirements for the district to meet by August 2021. Paragraph number 42 roughly outlines an evaluation of the effectiveness of PPSD's MLL programs over a three-year timespan for students who remained enrolled in the district and were either kindergarteners, third graders, sixth graders, or ninth graders during the 2016-17 school year. This report fulfills this requirement.

In concert with PPSD, The Policy Lab collated, cleaned, and analyzed administrative data from School Year (SY) 2016-17 through SY 2019-20 to answer these questions. This time period begins before PPSD and the DOJ entered into the settlement agreement. The following report provides exploratory, descriptive, and quasi-experimental analyses of the data collected.

During the years covered by this study, students in some of PPSD's English Language programs saw much more growth in English language acquisition and English Language Arts than their peers. In particular, elementary students in the Two Way Dual Language and Bilingual programs demonstrated, on average, the most growth. On the other hand, secondary students saw minimal gains in these and other subjects, regardless of program. Moreover, while no students in our study cohort repeated a grade during our study, exit rates from MLL programming were generally low, though students in the third through fifth grades had substantially higher exit rates than others.

While student absenteeism rates in PPSD were quite high relative to national averages, MLL students had lower absenteeism rates compared to PPSD students at large. On the other hand, MLL students in high school were less likely to enroll in AP courses and less likely to graduate in four years than other PPSD students.

We also find that while the students studied in our cohort were generally similar to the MLL population at large, there were some important differences. The MLL population was more male than PPSD as a whole, and among high school students, male MLL students were more likely than female MLL students to leave PPSD during the duration of the study. MLL students were also much more likely to be Hispanic or Latino than PPSD as a whole.

Finally, we find that while enrollment in certain programs was meant to be based on measured English language ability, students with a wide range of English language proficiency exam scores participated in these programs. This is notable because the different programs were meant for students of differing abilities. Many students in the cohort of interest and the broader MLL

[2] The DOJ settlement refers to Multilingual Learners (MLLs) as English Language Learners (ELLs or ELs). We use the term multilingual learners throughout this report to align with the terminology PPSD uses.

population were also missing standardized exam scores—including scores on some or all of the sections of the English language proficiency exam required to exit the MLL program. We note that this exam has several sections, and students seem to be missing scores from different sections at different rates. This indicates that some sections are likely either more difficult to schedule or more difficult for students to attend.

We conclude this report with several next steps that The Policy Lab and PPSD are planning to collaborate on to improve MLL instruction in the district. First among these is whether we can better understand the effect of teacher certifications on student outcomes. Second, the advent of community eligibility for free and reduced price lunch has reduced the utility of this traditional indicator of socioeconomic status. We hope to work together to build an alternative metric. Next, we hope to investigate ways to improve MLL student attendance, especially at critical times such as at English language proficiency exams. Finally, given the promising findings around the impact of elementary dual language and bilingual programs, we hope to rigorously evaluate the expansion of these programs' effect on MLL student outcomes.

Table of Contents

1. Background

1.1 Multilingual Learners

Multilingual learners (also known as English Language Learners (ELs or ELLs))[3] are one of the fastest growing student populations in the United States, with the National Education Association (NEA) estimating that one out of every four students across the United States will be classified as MLL by 2025 (National Education Association, 2020). The U.S. Department of Education (DOE) estimates that the number of kindergarten through 12th grade MLL students enrolled in public schools between School Year (SY) 2000-01 and SY 2016-17 rose by over one million students or by 28.1%, from 3.8 (8.1%) to 4.9 (9.6%) million students (Office of English Language Acquisition, 2020). Although the majority of MLL students nationwide speak Spanish as their home language, the languages represented by this student population are also incredibly varied (Soto, et al., 2015), as is their racial and ethnic diversity (de Brey, et al., 2019).

Unfortunately, MLL students are often underserved academically, with MLL students generally achieving worse academic outcomes than their peers (Fry, 2007; Kanno & Varghese, 2010; Zinth, 2013). Given the current and anticipated rise in the number of MLL students, there is a great deal of urgency to not only learn what works in MLL education, but how best to evaluate and refine education systems so that they continually strengthen their ability to meet the diverse needs of MLL students across all age groups and backgrounds.

The goal of this report is therefore twofold: first, to help Providence Public School District (PPSD) develop a baseline understanding of how their systems were functioning to support MLL students as well as to identify high-impact areas for improvement. Secondly, this report serves as a model for districts who wish to conduct a longitudinal evaluation of their MLL programs and how their systems could be better aligned to provide MLL students with the intended support. Finally, while many of the findings and recommendations in this report are district specific, these challenges and successes are unlikely to be unique to PPSD. As such, districts may wish to review the recommendations and next steps in the light of their own policies and practices.

1.2 Providence Public School District

In 2018, one in eight Rhode Island residents was an immigrant. Of these immigrant residents, 7,613 were children, the majority of whom were enrolled in PPSD. During the 2016-17 academic year, PPSD educated approximately 24,075 students, including 5,619 MLL students across 22 elementary schools, seven middle schools, 10 high schools[4] and two public district charter schools (Rhode Island KidsCount, 2018). Additionally, more than half of the district's students

[3] The DOJ settlement refers to Multilingual Learners (MLLs) as English Language Learners (ELLs or ELs). We use the term multilingual learners throughout this report to align with the terminology PPSD uses.
[4] As of SY 2020-21, PPSD has nine high schools, after closing Evolutions High School (Borg, 2020).

come from homes where English is not the primary language, representing about 55 languages in total.

In short, there is an extremely diverse cultural and linguistic landscape embedded in PPSD, requiring a correspondingly strategic and nimble education service provision. MLL services have been provided across eight programs[5] (see Appendix A for program definitions):

Table 1. Crosswalk of MLL program names as used by PPSD and listed in the DOJ settlement.

Program Name Used By PPSD[6]	**Program Name According to the DOJ Settlement**
ESL[7] Sheltered	English as a Second Language/Sheltered Instruction ("Sheltered ESL")
ESL Integrated	Integrated ESL
Newcomer ESL	ESL Newcomer
Consultation Model (SY 2016-17 Only)	The Consultation Model
Collaborative ESL	Collaborative ESL (or English Learner Collaborative (ELC)) ESL Push-In (Duplicate)
Transitional Bilingual	Transitional Bilingual
Developmental Bilingual	Developmental Bilingual
Two Way Dual Language	Dual Language

The Transitional Bilingual and Consultation Model programs are no longer offered (definitions for these programs can be found in Appendix B; see Appendix C for a timeline of which MLL programs were offered by school year).

[5] The ESL Push-in program was listed in the original DOJ settlement as a separate program, bringing the total number of programs to nine. However, we were informed by PPSD that ESL Push-In and the Collaborative ESL program provide the same MLL service delivery model.
[6] Throughout this report, we refer to MLL programs using the names in use by PPSD. We provide this table as a reference across program names.
[7] ESL stands for English as a Second Language.

Students are eligible for different programs based primarily on their English language proficiency level, as measured by the WIDA screener[8]. However, some programs were only available at certain schools and some programs were only available in certain grades. When registering for school, parents may enroll their students in any program with available seats for which they are eligible, so family preferences among the programs or proximity between home and school may play a role in enrollment patterns. In discussions with PPSD staff, we learned, for instance, that Bilingual/Dual Language programs tend to fill up faster than other ESL programs. Parents of eligible students may also elect not to have their child receive services, a decision that parents may change in later years.

At a high level, the different MLL programs mostly vary by the following attributes: (a) English language ability as measured by the state approved English Language Development assessment, (b) the grade levels in which they are offered, (c) percent of MLL students in the classroom, (d) percent of time taught in English and Spanish,[9] (e) teacher qualifications, and (f) the number of teachers available in the classroom.

In 2018, the U.S. Department of Justice (DOJ) identified 12 violations of the Equal Educational Opportunities Act of 1974, 20 U.S.C. §§ 1701 et seq. ("EEOA"), by the district specifically impacting MLL students. To avoid litigation, PPSD and the DOJ came to a settlement agreement, which outlined specific requirements for the district to meet by August 2021. Paragraph number 42 roughly outlines an evaluation of the effectiveness of PPSD's MLL programs over a three-year timespan for students who remained enrolled in the district and were either kindergarteners, third graders, sixth graders, or ninth graders during the 2016-17 school year. The language of the DOJ settlement outlining the requirements of the MLL cohort analysis is below:

> ***"The District will evaluate the effectiveness of each of its EL programs district-wide to determine whether they are overcoming language barriers within a reasonable period of time and enabling students to participate meaningfully and equally in its educational programs. To that end, the District agrees to conduct a three-year longitudinal cohort analysis of each of its programs at the elementary, middle, and high school levels by disaggregating and monitoring the following data by current, former, and never ELs, for each EL program utilized by the District: standardized test scores, exit rates, dropout rates, graduation rates, retention-in-grade rates, English Language Proficiency assessments, and enrollment in special education and enrichment programs (e.g., gifted, honors, and Advanced Placement classes). In conducting the analysis, the District will track a cohort of ELs who were enrolled in kindergarten, third grade, sixth grade, and ninth grade in SY 2016-17 and who remain enrolled in the District over the term of this Agreement. The***

[8] The WIDA screener is an initial English language proficiency assessment administered to identify potential English learners. In subsequent years, students' English language proficiency is monitored using the ACCESS assessment.

[9] While there are students with many language backgrounds, PPSD currently only offers Bilingual and Dual Language program models using English-language and Spanish-language instruction.

District will use the results of its longitudinal analysis to inform EL program decisions and ensure every EL program it uses is effective."

This document, a collaboration of The Policy Lab (TPL) and the Providence Public School District, answers, to the best of our abilities, the questions raised by the settlement agreement.

2. Analysis Plan

2.1 Overview

This analysis leverages administrative data from PPSD to study the uptake and effectiveness of the district's MLL programs. We create four cohorts of students—starting in kindergarten, third grade, sixth grade, and ninth grade—and follow them over a period of three years, from SY 2016-17 to SY 2018-19. Students are included in our cohort if they were enrolled in an MLL program for a minimum of one day in each of those three school years. We then examine outcomes for students at the end of SY 2018-19.

2.1.1 Outcomes

In alignment with DOJ-specified outcomes ("standardized test scores, exit rates, dropout rates, graduation rates, retention-in-grade rates, English language proficiency assessments, and enrollment in special education and enrichment programs [e.g., gifted, honors, and Advanced Placement classes]"), we use the following outcome variables when considering program efficacy:

- Standardized and progress monitoring test scores:
    - SY 2016-17 to SY 2018-19 STAR[10] Early Literacy, Reading, and Math
    - Rhode Island's state standardized tests, namely, SY 2016-17 Partnership for Assessment of Readiness for College and Careers (PARCC) English Language Arts (ELA) and Mathematics scores, as well as SY 2017-18 and SY 2018-19 Rhode Island Comprehensive Assessment System (RICAS) ELA and Mathematics scores
- English language proficiency assessments
    - SY 2016-17 to SY 2018-19 ACCESS assessment[11]
- SAT scores for relevant students

[10] Spanish STAR scores were excluded from the analysis because the dataset lacked correct identifiers. PPSD maintains a system of student identifiers called LASIDs and the Rhode Island Department of Education (RIDE) maintains a separate system of student identifiers called SASIDs. During our data cleaning process, we observed that some LASIDs matched multiple SASIDs and some SASIDs matched multiple LASIDs as well as other inconsistencies. Given RIDE's broader access to data from around the state, we decided to consider SASIDs the source of truth. However, in the Spanish STAR data, the listed SASIDs only had 6 or fewer digits, whereas actual SASIDs are all ten digits long. Moreover, 58% of the data did not have any unique identifiers. Therefore, we were unable to match this data for analysis.
[11] The ACCESS and ALT ACCESS assessments used different scoring systems. The ALT ACCESS had categorical scores while the ACCESS used numerical scores. We therefore omit the ALT ACCESS as we were unable to use these results to measure the difference between a student's score at the beginning and end of the three year study period. Please see Section 6. Limitations and Implications for more details on how this may affect the interpretation of this study.

- MLL program exit rates (as determined by the number of students that meet the exit criteria [Appendix D])
- Grade retention (i.e. students who repeated a grade over the course of the study)
- Graduation rate[12]
- Dropouts (i.e. high school students who leave school and do not return to school)[13]
- Enrollment in special education and enrichment programs (i.e. Honors and Advanced Placement classes[14]), as appropriate
- Student attendance (measured by days enrolled, present, and absent)

2.1.2 Student and Teacher Demographics

We take into account student demographics as these may affect their outcomes in the MLL program(s). We disaggregate descriptive analyses by student demographics, and, where appropriate, include them as covariates in our regression models to control for the variation in outcomes that can be attributed to these demographics. In addition, we consider how strongly these characteristics are correlated with student outcomes, and how they differ across our programs of interest.

Student demographic variables include a student's grade level, school, race, gender, Free and Reduced Price Lunch (FRPL) status, home language, special education (SPED) status, and Students with Interrupted Formal Education (SIFE) status.

We were unable to obtain sufficient historical data on teachers' years of experience and demographics[15] and are therefore unable to make any conclusions about the effect of an individual teacher's skill or experience on the efficacy of an MLL program.[16] However, we acknowledge the importance and impact of a teacher's skill and experience on students' learning and growth (e.g. Chetty et al., 2014), and encourage more robust data collection and tracking of a suite of relevant teacher variables so that future analyses of PPSD's MLL programs can account for these factors.

---

[12] We use data from SY 2019-20 to examine graduation rates, since most students would have been in the 11th grade in SY 2018-19 and therefore unlikely to be eligible for graduation.
[13] These rates are measured using the Rhode Island Department of Education's audited "exit code" data. For more information on that process, see the Rhode Island Department of Education's Education Data Hub at https://www.ride.ri.gov/InformationAccountability/RIEducationData/EnrollmentGraduationData.aspx.
[14] Although the DOJ Settlement mentions the Gifted Program as an enrichment program in its list of outcomes, we do not include this program as PPSD does not have a Gifted Program. PPSD does offer an Advanced Academics program; however, due to inconsistencies with the process for identifying qualifying students during our study period, we do not include Advanced Academics enrollment in our analysis.
[15] We note that PPSD does currently report on teacher certification as part of their ongoing reporting requirements to the DOJ, however, historical demographic data was incomplete or difficult to utilize.
[16] It is well-documented that teachers experience significant growth in their first three years of teaching (Kane, Rockoff, & Staiger, 2006; Staiger & Rockoff, 2010; Kini & Podolsky, 2016).

2.1.3 Program Types

There were a total of eight MLL programs in PPSD during the period of the study (Table 1; also Appendix A). Additionally, students who opt-out of receiving MLL services are considered "eligible not enrolled" (ENE) students and are MLL students who are not participating in an MLL program. For the purposes of this analysis, we treat ENE status as a "program," bringing the total number of comparison groups to nine.

Throughout our analysis, we observe and compare outcomes for students in different programs while accounting for MLL demographic variables and other covariates. However, we acknowledge that there are a variety of reasons behind a family's choice to opt out of MLL services and the potential underlying self-selection factors unique to ENE students that are not quantifiable in this analysis.[17] Since we are unable to tease out these factors, we urge caution in drawing strong conclusions about ENE students relative to other MLL students.

Students were also able to move between programs throughout the duration of our study. There were several reasons a student might switch MLL programs, including:

- a student moving to a different school with different MLL program offerings
- a student requesting a transfer to another program
- a place opening up in a program of a student's choice
- a program closing
- a student making sufficient progress on English language acquisition that they qualify for a different program
- a student opting to receive (or not to receive) MLL services
- a student qualifying to exit MLL status
- a student who recently exited MLL status (often called "Monitored Year 1" and "Monitored Year 2") and who re-entered the MLL program

Figure 1 and Table 2 illustrate the movement of students across programs between SY 2016-17 and SY 2018-19. Notably, ENE students as well as students in the Collaborative ESL and Two Way Dual Language programs tend to remain in the same programs; most students in the Developmental Bilingual program remained in the same program from SY 2016-17 to SY 2017-18, but many switched programs going into SY 2018-19. Likewise, most students in

[17] PPSD asks the person opting a student out of MLL services their reasons for doing so, though no reason is required. Typically this person is a student's parent or guardian, though in later grades it may be the student themself. Among those who gave a reason, some reasons given in SY 2019-20 were personal confidence in the student's English language ability, including a former General Education placement either in PPSD or another jurisdiction; a desire to be enrolled at a particular school or in a particular program, including to keep siblings together; a need to take enough credits to graduate on time; a preference to be enrolled in remote instruction during the pandemic; a lack of confidence in PPSD's MLL programming to help the student; and other reasons.

Sheltered ESL remained in the same program during the first two years of the study, while roughly half moved to the Collaborative ESL program in SY 2018-19, with a small number exiting MLL services into Regular Ed.

We address the role of students' movement across programs in our analyses by focusing on the program students were enrolled in during SY 2016-17, a so-called Intent to Treat (ITT) approach (see Section 2.2.4 Quasi-experimental Analyses for details).

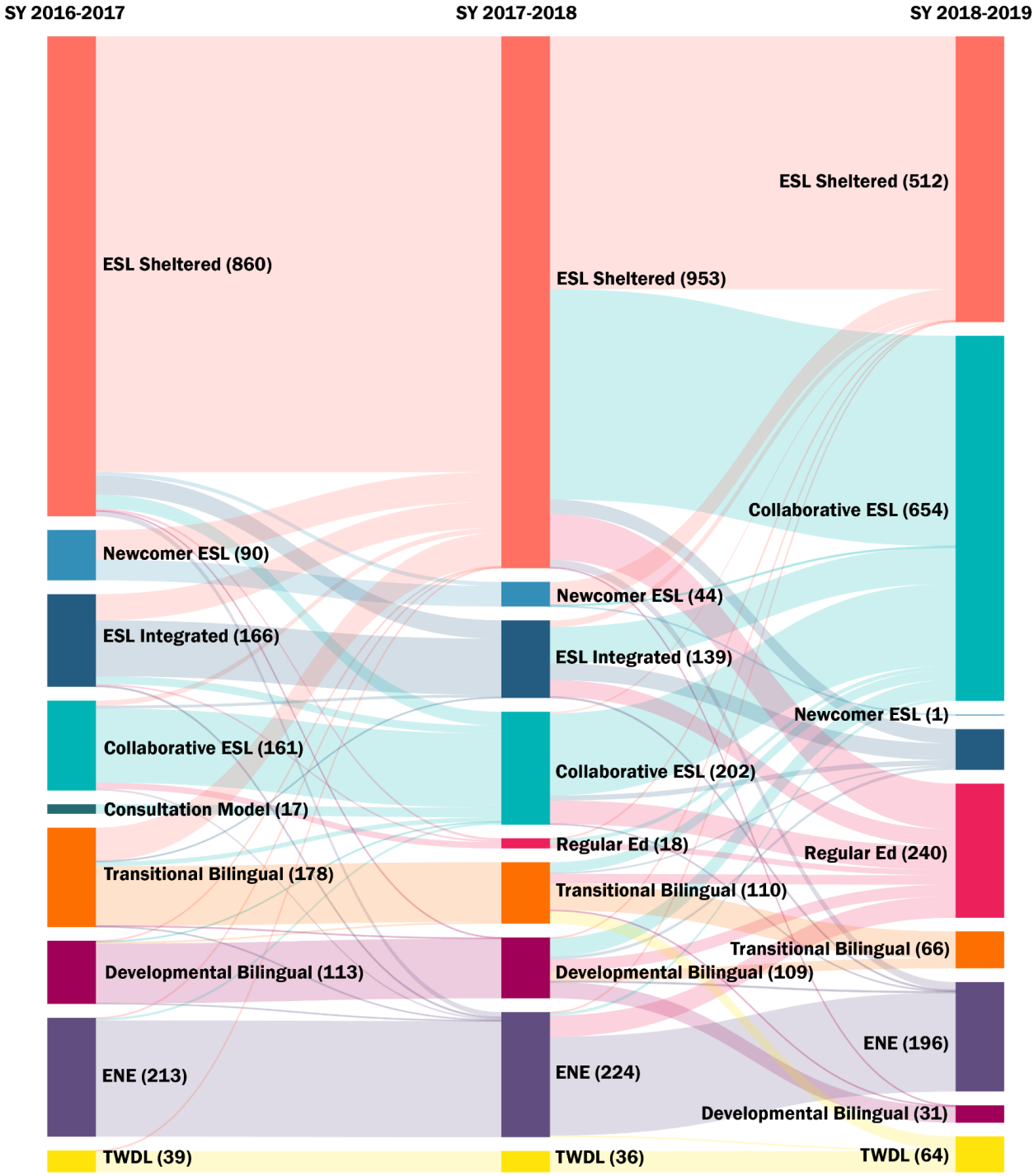
Finally, to determine a student's program type, we use the academic year (see Appendix C for years in which each program was offered) and the student's education type (i.e ESL, Bilingual, etc.). We also consulted with staff to clarify any administrative inaccuracies documenting students' program assignments, and referenced a list of students enrolled in the Consultation program provided by PPSD, as these students could not be easily identified in the administrative data. While we made significant effort to correctly identify a student's program type, we are unable to guarantee that our classifications are completely error-free given inconsistencies with the administrative data, the fact that we analyze data going back up to five years, and due to MLL program staff turnover (see Section 6. Limitations and Implications for a discussion of data challenges with identifying Student Type, and Appendix E for details about how we attempted to clarify program assignment business rules).

Table 2. Student movement across MLL Programs from SY 2016-17 to SY 2018-19[18]

		SY 2016-17	SY 2017-18	SY 2018-19
Kindergarten	ESL Integrated	66 (16.2%)	49 (12.0%)	17 (4.2%)
	ESL Sheltered	116 (28.5%)	132 (32.4%)	104 (25.6%)
	Consultation Model	*	0 (0.0%)	0 (0.0%)
	Collaborative ESL	66 (16.2%)	66 (16.2%)	129 (31.7%)
	Transitional Bilingual	74 (18.2%)	73 (17.9%)	66 (16.2%)
	Developmental Bilingual	53 (13.0%)	51 (12.5%)	31 (7.6%)
	Two Way Dual Language	20 (4.9%)	20 (4.9%)	43 (10.6%)
	ENE	*	*	*
	Regular Ed	0 (0.0%)	*	*
	Cohort	407 (100.0%)	407 (100.0%)	407 (100.0%)
3rd Grade	ESL Integrated	100 (14.7%)	90 (13.2%)	56 (8.2%)
	ESL Sheltered	187 (27.5%)	232 (34.1%)	40 (5.9%)
	Consultation Model	*	0 (0.0%)	0 (0.0%)
	Collaborative ESL	66 (9.7%)	99 (14.5%)	275 (40.4%)
	Transitional Bilingual	104 (15.3%)	37 (5.4%)	0 (0.0%)
	Developmental Bilingual	60 (8.8%)	58 (8.5%)	0 (0.0%)
	Two Way Dual Language	*	*	*
	ENE	138 (20.3%)	141 (20.7%)	106 (15.6%)
	Regular Ed	0 (0.0%)	*	*
	Cohort	681 (100.0%)	681 (100.0%)	681 (100.0%)
6th Grade	ESL Sheltered	251 (82.6%)	237 (78.0%)	115 (37.8%)
	Newcomer ESL	*	*	*
	Consultation Model	*	0 (0.0%)	0 (0.0%)
	Collaborative ESL	19 (6.2%)	25 (8.2%)	123 (40.5%)
	ENE	28 (9.2%)	34 (11.2%)	44 (14.5%)
	Regular Ed	0 (0.0%)	*	*
	Cohort	304 (100.0%)	304 (100.0%)	304 (100.0%)
9th Grade	ESL Sheltered	306 (68.8%)	352 (79.1%)	253 (56.9%)
	Newcomer ESL	87 (19.6%)	38 (8.5%)	0 (0.0%)
	Consultation Model	*	0 (0.0%)	0 (0.0%)
	Collaborative ESL	*	*	127 (28.5%)
	ENE	40 (9.0%)	41 (9.2%)	*
	Regular Ed	0 (0.0%)	*	*
	Cohort	445 (100.0%)	445 (100.0%)	445 (100.0%)

[18] Throughout this report, we suppressed table cells which contain small numbers. Per PPSD's standards, all number less than 10 have been replaced with a "*". To reduce the potential for "backsolving" the suppressed numbers using other data available in the table, we further applied a "complementary suppression" algorithm based on the standards of the Connecticut State Department of Education (Connecticut State Department of Education, 2015). See https://github.com/thepolicylab/zentables for the implementing code.

Figure 1: Flow of students across MLL programs in PPSD from SY 2016-17 to SY 2018-19



The above Sankey diagram traces students throughout the study period. Each color represents a program a student could be enrolled in, and each column represents a school year. The bands between the columns represent the flow of students between programs between school years. For instance, the coral colored vertical bar at the top of the middle column represents the 953 students in our cohort who were enrolled in ESL Sheltered in SY 2017-18. The light blue band to the right of it represents those students who moved from ESL Sheltered to Collaborative ESL in SY 2018-19.

2.2 Methods

2.2.1 Creating the Cohorts

We create four cohorts in accordance with the DOJ specification that we "track a cohort of ELs who were enrolled in kindergarten, third grade, sixth grade, and ninth grade in SY 2016-17 and who remain enrolled in the District over the term of this Agreement" (i.e. from SY 2016-17 to SY 2018-19).

A student must meet the following requirements to be included in a cohort:

1. They must be a student who was enrolled in PPSD and must have attended school at PPSD for a minimum of one day in each of the three years of the study (i.e. SY 2016-17 to SY 2018-19).
2. They must have been enrolled in kindergarten, third grade, sixth grade, or ninth grade during SY 2016-2017.[19] Grade levels in subsequent years do not matter, meaning that a student who is held back a year, for example, will remain in the cohort.
3. They must have been actively enrolled in an MLL program at the start of SY 2016-17. Students who exited the MLL program *prior* to SY 2016-17 and were considered Monitored Year 1 or Monitored Year 2 during SY 2016-17 were excluded.[20]

As mentioned above, ENE students are included in our cohorts. Thus, our final sample (N=1837) includes four cohorts: K Cohort (n=407), third grade Cohort (n=681), sixth grade Cohort (n=304), and ninth grade Cohort (n=445). The number of students within each cohort remains consistent across all three years given the first criteria in the cohort definitions listed above.

2.2.2 Descriptive Analyses

We begin our analyses by providing an overview of the demographic characteristics of PPSD's MLL population as well as of the students within our cohorts. We will observe how these demographic characteristics change over time within the nine MLL programs (Appendix A), including ENE students, and may compare these demographics across programs.

We then create descriptive statistics of changes in the outcome variables outlined above (see Section 2.1.1 Outcomes), both in aggregate and by program where possible and/or meaningful.

[19] We do not distinguish between students who repeated a grade in SY 2016-17 (e.g. a student who was in sixth grade in SY 2015-16 and again in SY 2016-17) and students who entered the grade for the first time.
[20] Students who meet the criteria to be in the cohort and who exited MLL services between SY 2016-17 and SY 2018-19 were included in the analysis.

We also compare students across program types (including ENE students) to determine if some programs seem to contribute to better outcomes for MLL students.

Finally, as preparation for the quasi-experimental analysis (see 2.2.4 Quasi-experimental Analyses), we consider whether these demographic and behavioral factors predict whether a student remains enrolled for all three years of our study, whether they enroll in particular programs, and whether they complete all required examinations that we consider in our analyses.

2.2.3 Complete Cases

Following the descriptive analyses, we use a quasi-experimental matched-comparison group analysis designed to attempt to measure the actual effects of MLL program enrollment on academic outcomes. To do so, we need to further limit our sample. In particular, even though we are primarily focused on outcomes as measured by standardized tests, we do not have scores for all students on all assessments. This could be for many reasons, including parents opting students out of statewide testing (McGowan, 2015) or difficulties in scheduling certain examinations.[21] The different types of tests students may be required to take compounds the issue of missing test scores. Depending on a student's grade, they may take some combination of language acquisition tests (ACCESS, annually), progress monitoring tests (STAR, three or more times a year), statewide standardized tests (PARCC or RICAS, annually), and college readiness tests (SAT, up to multiple times a year). The more tests a student is required to take, the more opportunities they have to miss at least one of their required test scores. As such, when performing our quasi-experimental analysis, we only examine students for whom we have complete test information.

[21] In discussions with PPSD staff, we learned that the ACCESS exam is usually administered over several days. While it is possible for kindergarteners to complete the assessment in one day, for students in higher grades that are more likely to take a computerized exam, PPSD recommends that each domain assessment (there are four total) be scheduled on separate days owing to the potentially lengthy time required (there is no time limit to complete any single domain). This means that there are logistical challenges to ensuring that all students have a complete ACCESS assessment. Examples of these challenges include needing schools to shift schedules flexibly to accommodate students who require more time to complete the assessment than the scheduled block allocated, as well as navigating student absences. Additionally, certain parts of the assessment, such as the speaking proficiency component, cannot be administered in a large group and is therefore both more time-intensive to administer and challenging to schedule. This is reflected in many students having complete ACCESS exam scores *except for* the speaking proficiency section of the exam.

We define information from a test as "complete" only if all sections were completed[22] and the test was taken in the first (SY 2016-17) and last (SY 2018-19) year of our study period.[23] Specifically, we require a student to have the test scores as described in Table 3.

Importantly, this requirement further limits our ability to make strong causal claims about the efficacy of PPSD's different MLL programs. For instance, a student who believes that they are likely to pass their ACCESS exam and thus exit the MLL program may be more motivated to attend the exam or put greater effort into achieving a good score. The opposite might also be true: a student who is worried that exiting an MLL program will reduce the amount of time they spend with their friends may avoid taking the ACCESS exam or put less effort into completing the assessment. Without full test coverage, these scenarios are indistinguishable.

Finally, we note that the completeness issue is especially important for understanding MLL exit rates. Having a complete ACCESS score is required to successfully exit the MLL program, and so not taking it is tantamount to remaining in the program for an extra year. We urge the reader to keep this limitation in mind while reading this report.

Table 3: Assessments a student in our cohort must have taken to qualify as a complete case.

Cohort	Complete Case Definition
Kindergarten Cohort	SY 2016-17 and SY 2018-19 ACCESS scores
Third Grade Cohort	SY 2016-17 and SY 2018-19 ACCESS scores, SY 2016-17 PARCC scores, and SY 2018-19 RICAS scores
Sixth Grade Cohort	SY 2016-17 and SY 2018-19 ACCESS scores, SY 2016-17 PARCC scores, and SY 2018-19 RICAS scores
Ninth Grade Cohort	SY 2016-17 and SY 2018-19 ACCESS scores, SY 2016-17 PARCC scores, and SY 2018-19 SAT scores

[22] For example, a student must have completed the listening, reading, writing, and speaking portions of the ACCESS test. Missing any one of these components would result in a student being omitted from the "Complete Case" group.

[23] In the case of statewide summative test scores, since the required test changed between the first and last year of our study, we treat this pair of tests as "one" test for the purpose of determining complete cases.

2.2.4 Quasi-experimental Analyses

In general, the form of our research questions is as follows: in a subset of each cohort of students in PPSD, typically defined by eligibility to participate in some subset of MLL programs (Appendix A), we will measure an outcome variable which is the result of a standardized test, progress monitoring test, or English language proficiency assessment at the end of a cohort period,[24] which we think should relate to:

- a behavioral outcome, such as student attendance
- participation in a particular MLL program, or
- student demographics, SY 2016-17 test scores, and certain school indicators.

These variables might explain why certain eligible students end up participating in an MLL program and/or dropping out of it.

We want to evaluate whether differences in MLL programs led to differences in outcomes. Since we cannot observe the same students experiencing multiple programs, we compare students who experienced one (or more) MLL programs to other similar students enrolled in other MLL programs.[25]

Imagine a kindergarten student enrolled in the Two Way Dual Language Program in SY 2016-17 with outcomes measured in SY 2018-19 at the end of second grade. Would this student's English language proficiency and test scores have been worse had they not participated in this program but in some other MLL program? How might we model the counterfactual trajectory of this child?

To answer these questions, we employ a *matching, stratification-based* strategy[26] whereby we find another child who represents this counterfactual trajectory. We included the following covariates in this comparison: grade, race, gender, FRPL status, SPED status, SIFE status, home language type, and baseline test scores. For example, we might create a matched pair by finding and grouping our student of interest with another child who was eligible for the Two Way Dual Language program but instead opted to not enroll in any MLL program. By comparing test scores within the matched pair, we will know that any differences in scores are not due to differences in the variables that we used to create the set (or only reflect those differences a small amount, not enough to change our statistical conclusions). If there are no reasons other than the

[24] We note that another possibility is that a student exits the MLL program or graduates from PPSD. Exiting the MLL program is, in part, a downstream result of standardized and language proficiency tests (see Appendix D, which enumerates exit criteria), so the primary outcomes that we will measure here will remain the test scores themselves.
[25] In our pre-analysis plan, we described that we would attempt to compare students only within schools. However, in practice, we found such matches to be incredibly impractical.
[26] For more information on matching techniques, see Rosenbaum (2010), Part II.

observed reasons for why a student would end up in a particular MLL program, then we can ascribe any observed effects to the programs themselves.[27] We emphasize that this is a very large assumption. For instance, perhaps a family chooses a program based on its availability at the school closest to their home. We cannot and do not observe such preferences, but they could influence program enrollment. As such, the reader should interpret all of our quasi-experimental analyses in light of this assumption.

However, PPSD does not have just a single MLL program, but eight MLL programs and the option to not enroll (ENE). For outcomes where we are interested in how all programs comparatively perform relative to each other, such as those in 5.1 Research Question 1, all 9 x 8 / 2 = 36 comparisons are not possible across all potential outcomes. This is due to lack of comparable populations in each program and due to sample size constraints.

As such, we will present program-by-program average effectiveness estimates and confidence intervals according to the above procedures. This will allow us to say how well each program serves the average student in its enrolled population. However, it will not allow us to perform direct comparisons between programs. For instance, suppose that a student qualifies for Target Program if they have a baseline screener score between 0 and 100. But to qualify for Alternate Program A, they must receive a score between 0 and 50, and to qualify for Alternate Program B, they must receive a score between 50 and 100. This analysis will allow us to ask about the *overall* effectiveness of Target Program, but we will not be able to say if Target Program performs better than Program A or better than Program B as different students are qualified for the latter programs.

Technical Specifications

We utilize the optimal full matching procedure of Hansen and Klopfer's `optmatch` R package (version 0.9-13) (Hansen & Klopfer, 2006). Specifically, following Rosenbaum (2010) Chapter 10, we create a distance measure based on the ranked Mahalanobis distance between the vector of covariates consisting of a student's

- gender,
- race,
- FRPL status,
- student with interrupted formal education (SIFE) status,
- whether they are enrolled in special education, and
- whether their home language was listed as Spanish, some other language, or unlisted (see Table 7 and Appendix F for details).

[27] More formally, if we have blocked all confounders, then this causal interpretation is possible (Morgan & Winship, 2014).

As English language ability is critical to performance on these standardized exams, we also require that matched students' baseline ACCESS scores be within ¼ of a standard deviation from each other.

Our primary outcomes of interest are improvement on standardized test scores. As such, we also require that, in any matched pair, baseline standardized test scores be within ¼ of a standard deviation from each other. We note that as Rhode Island's state tests changed over the course of this study, the baseline test *may not be the same* as the outcome test. In particular, we utilize the baselines for each test and each cohort as outlined in Table 4.

Table 4. Outcome variables we examine in our matching analyses and the baseline variables we match on.

	Outcome Test (SY 2018-19)	**Baseline Test (SY 2016-17)**
Kindergarten	Overall ACCESS Score	Overall ACCESS Score
	STAR Reading[28]	STAR Early Literacy
Third Grade	Overall ACCESS Score	Overall ACCESS Score
	RICAS ELA	PARCC ELA
	RICAS Math	PARCC Math
Sixth Grade	Overall ACCESS Score	Overall ACCESS Score
	RICAS ELA	PARCC ELA
	RICAS Math	PARCC Math
Ninth Grade	Overall ACCESS Score	Overall ACCESS Score
	PARCC ELA	SAT ELA
	PARCC Algebra	SAT Math

[28] Although we do not use STAR Reading or Early Literacy to determine "complete cases", we examine outcomes on these assessments for the kindergarten cohort as these students do not take the state assessment. Doing so provides a sense of how these students are doing on ELA, which we were not able to obtain by solely examining ACCESS test scores, as the ACCESS assessment measures English language proficiency. Here, we conduct analyses on STAR Reading and Early Literacy test scores only for students who otherwise satisfy the criteria for "complete cases" and also have the two listed STAR exams.

With this distance function, we can then employ the `pairmatch` procedure of Hansen and Klopfer as implemented in the `optmatch` package, which minimizes the total distance between all pairs.

We then perform a fixed effects regression of the form:

$$\text{outcome}_i \sim 1 + t_i + p_j$$

where t_i indicates whether student *i* was enrolled in the program of interest, p_j is a fixed effect for the pair that student *i* belongs to, and outcome_i is the outcome of interest. We compute this regression with the `lm_robust` function of the `estimatr` R package (version 0.30.2) (Blair et al., n.d.). We report confidence intervals and *p*-values using design-justified HC2 standard errors.

All code for this analysis is available upon request of the authors.

Intent to Treat

We focus estimation on the intent to treat (ITT) effect of a child enrolling in one of the nine MLL programs because children change schools and/or switch between programs. That is, in this analysis, we will not attempt to account for the dynamic effects of switching schools or programs on student's outcomes, nor will we focus on the effects of each specific trajectory of schools and programs chosen by each child. Instead, we focus on the effects of the program a student was enrolled in SY 2016-17. As an example, an MLL student who participates in the ESL Sheltered program in SY 2016-17 will be flagged as "ESL Sheltered" for the duration of the study. We do not distinguish, for instance, between an MLL student who participates in the ESL Sheltered program for a year before transfering to a different program, versus an MLL student who participates in the ESL Sheltered program all three years.

There are several reasons for making this choice: First, we hope that the effects of a particular program would continue to be felt throughout a student's educational career. An ITT analysis helps answer the question of how much a particular decision to enroll in a particular program at a particular point in time affects later achievement.

Second, as seen in Figure 1, the number of possible paths students can take through various different MLL programs is quite large. Indeed, *a priori*, a student might take any[29] of 9 x 8 x 8 = 576 different paths through each of the different programs over the course of three years. Such a large number of potential paths relative to the number of students in our cohorts makes

[29] The Consultation Model was only available during SY 2016-17, and so the latter two numbers in our math are eight.

measuring the dynamic effects of programs challenging. As such, an ITT analysis is perhaps the cleanest analysis that we can offer with the data available.

Three Important Caveats[30]

There are three important caveats in our procedure. First, as discussed above, we do not match on the school in which someone is enrolled. Given the high potential for school-level effects, this is a big limitation of this study.

Second, we note that our baseline scores are not fully pre-treatment scores. In the case of state-mandated standardized tests such as the RICAS, PARCC, and SAT, this is primarily due to data limitations in which we either did not have access to SY 2015-16 scores, or students, such as kindergartners, were too young to have taken such examinations. In our pre-analysis plan, we had planned to try to replace certain baseline scores with the STAR progress monitoring exams. However, we found that coverage of STAR exams was not robust, especially in older cohorts (see Section 4.1.2), and so we had to abandon these plans. Thus, our measure is effectively of improvement over the latter *two* years of the cohort study.

The one exception to this caveat is for our kindergarten cohort. Here, we utilize the beginning of year (BOY) STAR Early Literacy exam as our matching variable and use as our outcome variable the end of year (EOY) STAR Reading exam. While the dates students take these exams are not fixed, it is true that the BOY exam tends to be taken in September and the EOY exam tends to be taken sometime between April and June.

Third, recall that we have limited our analysis to complete cases. We make one exception to this analysis: for kindergarten students, STAR exam participation was not universal among our cohort. However, 301 of 366 (or 82%) of students in our cohort took both exams. We believe that this is sufficient coverage that we will separately examine students who had taken both the BOY STAR Early Literacy exam in SY 2016-17 and the EOY STAR Reading exam in SY 2018-19.

Adjusting p-values

In our quasi-experimental analyses, we are interested in keeping the family-wise error rate below a significance level of 0.05. As such, we employ a Bonferroni correction to the *p*-values obtained from our HC2 standard errors above. To do so, we declare that each cohort and each outcome variable defines a family of hypotheses. Each family f, will have some number of m_f of matching procedures that actually succeed. If p_{fi} is the p value for the ith test in family f, its Bonferroni corrected value is $m_f p_{fi}$. If this number is less than 0.05, we will consider the result significant.

[30] See 6. Limitations and Implications for discussion of other considerations.

For the reader's convenience, we report uncorrected *p*-values, though in our quasi-experimental analysis write ups, we only discuss those cases where the Bonferroni-corrected *p*-value is less than 0.05. They will be starred and bolded in the regression tables of Appendix H.

Robustness and Other Checks

Due to the non-experimental nature of our identification strategy, we perform several checks to understand whether or not any particular result might be considered spurious.

First, from the result of each matching, we perform a balance test to understand whether or not our matching strategy actually results in a balanced sample across our demographic groups of interest. Specifically, we employ the omnibus `xBalance` test of Hansen and Bowers's `RItools` package (version 0.1-17) (Hansen & Bowers, 2008). If this omnibus test reports a significant finding (defined for our purposes as a *p*-value of less than 0.05), then downstream regressions should be viewed skeptically.

Second, we check whether an effect persists if we measure a difference in rank as opposed to a difference in mean. To perform this check, we utilize a permutation test on the pair assignments. Specifically, we utilize the approximative Wilcoxon-Mann-Whitney test as implemented in the R `coin` package (version 1.4-1) (Hothorn et al., 2008).

3. Student Demographics

In this section, we discuss the winnowing of students from the overall PPSD population, to those enrolled in MLL programs, to those who are ultimately in our cohort, i.e. those who are enrolled in PPSD for all three years of the study. At each stage, the total population that is studied is reduced (see Appendix G for details on how each filter affects the total population being studied). In this section we focus on topline demographics. For statistical comparisons between our cohort and the full MLL population in PPSD, see Appendix I. We note that while these populations differ, this study is focused on students who (a) were enrolled in PPSD for three years starting in SY 2016-17 and (b) completed all the exams listed in Table 4.

3.1 MLL Students in PPSD

MLL students made up roughly 30% of the PPSD student population, or 6,749 out of 24,751 students in SY 2016-17. Between SY 2016-17 to SY 2018-19, the proportion of MLL students in PPSD increased from 27.3% (n=6,749) to 32.4% (n=7,965), an increase of over 1,000 MLL students, even though the total number of students in PPSD remained roughly consistent (Table 5).

Across PPSD, males make up a slightly larger proportion (55.4%) of MLL students. Most MLL students are Hispanic/Latino (83.6%) and qualify for FRPL (87.5%) (Table 6).

We observe some variation in the number and proportion of MLL students by grade, with middle schools typically having the smallest number and proportion of MLL students, regardless of study year. The number of students then rises in ninth grade, before gradually decreasing over the next three years. For instance, the percentage of MLL students in sixth grade during SY 2016-17 was just 19.1% (n=364), compared to 28.1% (n=513) in kindergarten and 34.3% (n=789) in ninth grade (Table 5 and Figure 2).

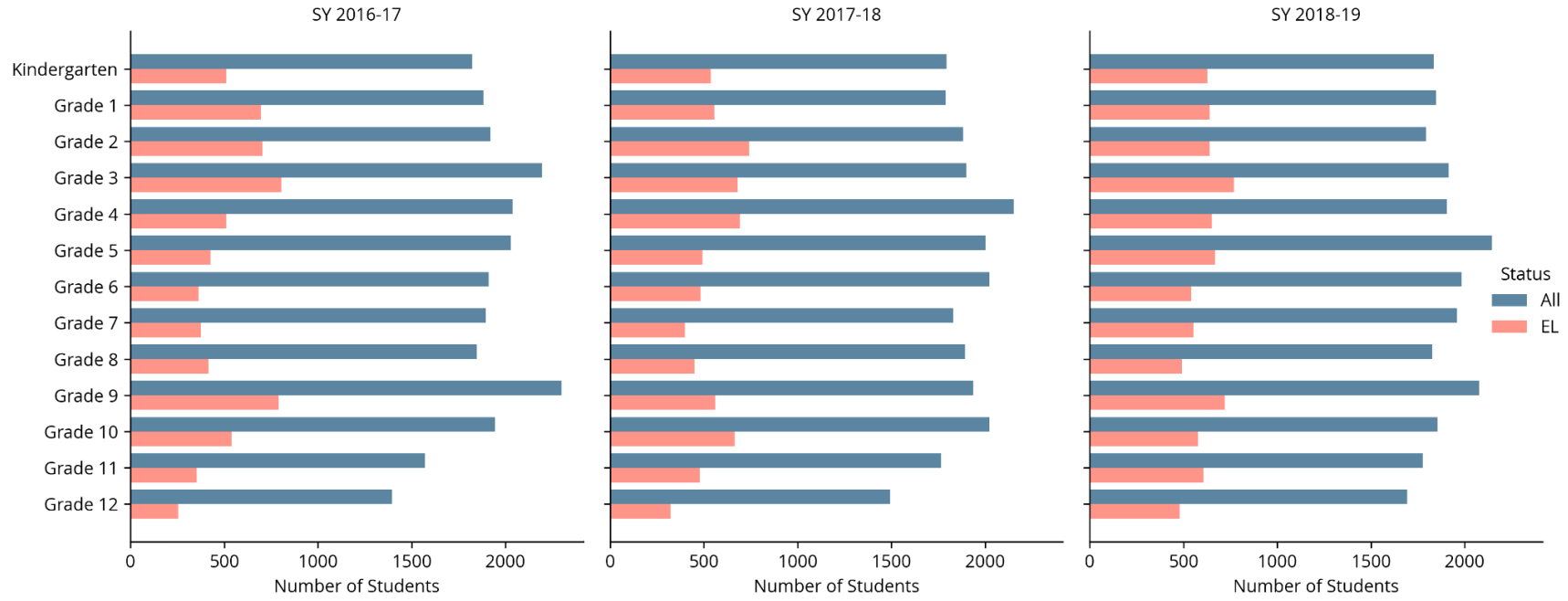
Table 5. Numbers and percentages of MLL students in PPSD by grade.

	SY 2016-17			SY 2017-18			SY 2018-19		
	MLL Students	All Students	%	MLL Students	All Students	%	MLL Students	All Students	%
Kindergarten	513	1824	28.1%	537	1795	29.9%	628	1834	34.2%
Grade 1	696	1883	37.0%	555	1791	31.0%	639	1847	34.6%
Grade 2	705	1919	36.7%	742	1883	39.4%	640	1794	35.7%
Grade 3	806	2195	36.7%	680	1901	35.8%	769	1914	40.2%
Grade 4	512	2038	25.1%	693	2154	32.2%	650	1906	34.1%
Grade 5	426	2030	21.0%	493	2004	24.6%	669	2146	31.2%
Grade 6	364	1910	19.1%	482	2023	23.8%	541	1985	27.3%
Grade 7	376	1895	19.8%	400	1830	21.9%	555	1959	28.3%
Grade 8	416	1847	22.5%	450	1893	23.8%	491	1828	26.9%
Grade 9	789	2300	34.3%	559	1936	28.9%	721	2078	34.7%
Grade 10	538	1944	27.7%	665	2023	32.9%	576	1855	31.1%
Grade 11	353	1570	22.5%	479	1765	27.1%	606	1777	34.1%
Grade 12	255	1396	18.3%	324	1494	21.7%	480	1694	28.3%
Total	6749	24751	27.3%	7059	24492	28.8%	7965	24617	32.4%

Table 6. Demographic information of MLL students in PPSD by grade.

	K	Grade 1	Grade 2	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8	Grade 9	Grade 10	Grade 11	Grade 12	Total
Female	236 (46.0%)	332 (47.7%)	355 (50.4%)	385 (47.8%)	234 (45.7%)	150 (35.2%)	158 (43.4%)	165 (43.9%)	190 (45.7%)	301 (38.1%)	210 (39.0%)	164 (46.5%)	133 (52.2%)	3013 (44.6%)
Male	277 (54.0%)	364 (52.3%)	350 (49.6%)	421 (52.2%)	278 (54.3%)	276 (64.8%)	206 (56.6%)	211 (56.1%)	226 (54.3%)	488 (61.9%)	328 (61.0%)	189 (53.5%)	122 (47.8%)	3736 (55.4%)
Asian	15 (2.9%)	34 (4.9%)	43 (6.1%)	38 (4.7%)	17 (3.3%)	13 (3.1%)	13 (3.6%)	13 (3.5%)	*	24 (3.0%)	*	*	*	237 (3.5%)
Black	33 (6.4%)	45 (6.5%)	39 (5.5%)	36 (4.5%)	39 (7.6%)	31 (7.3%)	28 (7.7%)	25 (6.6%)	31 (7.5%)	77 (9.8%)	37 (6.9%)	27 (7.6%)	27 (10.6%)	475 (7.0%)
Hispanic/Latino	426 (83.0%)	558 (80.2%)	579 (82.1%)	706 (87.6%)	433 (84.6%)	357 (83.8%)	302 (83.0%)	321 (85.4%)	355 (85.3%)	640 (81.1%)	455 (84.6%)	297 (84.1%)	211 (82.7%)	5640 (83.6%)
Native American	0 (0.0%)	*	*	*	*	*	*	*	0 (0.0%)	*	*	*	*	40 (0.6%)
Pacific Islander	*	*	0 (0.0%)	*	0 (0.0%)	*	0 (0.0%)	*	0 (0.0%)	*	*	*	0 (0.0%)	15 (0.2%)
Two or More Races	*	14 (2.0%)	11 (1.6%)	*	*	*	*	*	*	10 (1.3%)	*	*	0 (0.0%)	87 (1.3%)
White	31 (6.0%)	40 (5.7%)	26 (3.7%)	17 (2.1%)	17 (3.3%)	14 (3.3%)	12 (3.3%)	14 (3.7%)	15 (3.6%)	28 (3.5%)	19 (3.5%)	11 (3.1%)	11 (4.3%)	255 (3.8%)
Free/Reduced Price Lunch	453 (88.3%)	632 (90.8%)	644 (91.3%)	731 (90.7%)	456 (89.1%)	370 (86.9%)	313 (86.0%)	321 (85.4%)	343 (82.5%)	679 (86.1%)	448 (83.3%)	296 (83.9%)	219 (85.9%)	5905 (87.5%)
Not Free/Reduced Price Lunch	60 (11.7%)	64 (9.2%)	61 (8.7%)	75 (9.3%)	56 (10.9%)	56 (13.1%)	51 (14.0%)	55 (14.6%)	73 (17.5%)	110 (13.9%)	90 (16.7%)	57 (16.1%)	36 (14.1%)	844 (12.5%)
SPED	61 (11.9%)	80 (11.5%)	77 (10.9%)	93 (11.5%)	92 (18.0%)	78 (18.3%)	71 (19.5%)	52 (13.8%)	54 (13.0%)	45 (5.7%)	48 (8.9%)	*	*	780 (11.6%)
Non-SPED	452 (88.1%)	616 (88.5%)	628 (89.1%)	713 (88.5%)	420 (82.0%)	348 (81.7%)	293 (80.5%)	324 (86.2%)	362 (87.0%)	744 (94.3%)	490 (91.1%)	*	*	5969 (88.4%)
Overall	513 (100.0%)	696 (100.0%)	705 (100.0%)	806 (100.0%)	512 (100.0%)	426 (100.0%)	364 (100.0%)	376 (100.0%)	416 (100.0%)	789 (100.0%)	538 (100.0%)	353 (100.0%)	255 (100.0%)	6749 (100.0%)

Figure 2: Numbers of MLL students vs. all students in PPSD by grade.



3.2 MLL Students in Our Cohort

Our cohort consists of a total of 1837 MLL students, i.e. 27.2% of all MLL students or 7.4% of all students in PPSD during SY 2016-17 (see Appendix G for details on how we derived the cohort population). Overall, the MLL students in our cohort generally mirror the MLL population in PPSD at large, though there are some significant differences among the ninth grade cohort (see Appendix I.1). In this section, we break down the demographic characteristics of the MLL students in our cohort (Table 7).

Grade. We observe that the number of students in each cohort are not evenly distributed, with the third grade cohort (n=681) about twice the size of the sixth grade cohort (n=304) and roughly one and a half times that of the kindergarten (n=407) and ninth grade (n=445) cohorts . As such, the third grade cohort makes up approximately 37% of our overall cohort, while the sixth grade cohort makes up about 17% of our overall cohort. While this variation in cohort size by grade appears large, we are less concerned about this uneven distribution across grades than we might otherwise be as it mirrors that of the greater MLL population at PPSD (Table 6).

We also observe that the distribution of students by program[31] sometimes varied greatly by cohort in SY 2016-17. For instance, a higher proportion of students in the kindergarten cohort (16.2%) participated in the Collaborative ESL program compared to the other cohorts (less than 10%), and although almost 20% (n=87) of ninth grade cohort students participated in the Newcomer program, just three students in the sixth grade cohort did. Additionally, ESL Sheltered had the greatest proportion of student participation regardless of grade, though this varied from almost 30% in the elementary level cohorts to over 80% of students in the sixth grade cohort. Finally, about 20% of the students in the third grade cohort were ENE, compared to less than 10% each in the sixth and ninth grade cohorts, while virtually no kindergarten cohort students were ENE.

Gender. Overall, we see a higher proportion of male students in our MLL Cohort makeup (54.3% male vs. 45.7% female); the greater MLL population likewise consists of a greater proportion of male students. However, among those female students in the greater MLL population, they are much more likely to stay enrolled in PPSD for all three years of study (see Appendix I.1).

Across programs, male students also tend to make up a slightly larger proportion of students than female students, with a few exceptions. For instance, the Transitional Bilingual program for the kindergarten (n=74) and third grade (n=104) cohorts had more female than male participants (Table 8, Table 9); the difference in both cases was over ten percentage points. The same is true for students in the third grade cohort who were enrolled in the Collaborative ESL program (n=66) (Table 9). Other times, the ratio of male to female students differed greatly from that of the overall cohort. Across the third, sixth, and ninth grade cohorts, there was at least a

[31] For the purposes of this analysis, we consider ENE a “program”.

20 percentage point difference between the percentage of male versus female students who were ENE, with male students outweighing female students in each instance. Notably, the Newcomer ESL program for the ninth grade cohort (n=87) had more than twice the number of male (n=60) compared to female (n=27) students.

Race and Ethnicity. Throughout this report, self-identified race will always be one of the following categories:

- Asian
- Black
- Hispanic/Latino[32]
- Two or more races (multiracial)
- Pacific Islander
- White

As with the broader MLL population, most MLL students in our cohort (84.8%) are Hispanic/Latino. Likewise, Hispanic/Latino students make up the majority of participants across all program types, regardless of grade. However, it is worth noting that certain programs (such as the Transitional Bilingual, Developmental Bilingual, and Two Way Dual Language programs in the kindergarten and third grade cohorts) had few to no students from other racial/ethnic backgrounds. In contrast, certain programs (such as the ESL Integrated program in the kindergarten and third grade cohorts, the ESL Sheltered program across all cohorts, and the Newcomer ESL program in the ninth grade cohort) saw a greater representation of participants from other racial/ethnic backgrounds (Tables 8-11).

Finally, within our cohort, we observe that the ninth grade cohort had a slightly higher proportion of Black students (10.6%, compared to no more than 6% in the other cohorts) and a slightly lower proportion of Hispanic/Latino students (almost 79%, compared to over 84% in the other three cohorts).

Socio-economic Status (SES). Throughout this report, we use a student's FRPL status as a proxy for their socio-economic status (SES).[33] We find that across our cohorts, the proportion of students who qualify for FRPL is generally consistent and remains high (over 84%)[34] regardless of cohort grade or MLL program enrollment.

---

[32] Note that Hispanic and Latino are conflated in this enumeration, a feature of PPSD's administrative data.

[33] Importantly, in SY 2017-18, PPSD expanded its free lunch program to *all* elementary students, regardless of whether or not they had applied. They expanded this program again in SY 2019-20 to *all* students, elementary and secondary. As such, while SY 2016-17 FRPL status remains a reasonable proxy for socioeconomic status for the purposes of this study, it just barely does so. Follow up studies will need to find other measures to account for socioeconomic status differences (List, 2019).

[34] Programs with very low student enrollment may have skewed percentages; we do not take these programs into account here.

Special Education (SPED) Status. Compared to the overall MLL population, students in our cohort had similar rates of being enrolled in SPED programs after accounting for their grade.

Newcomer and Students with Interrupted Formal Education (SIFE) Status. Overall there were 143 students throughout our cohorts who were either enrolled in the Newcomer ESL program or were labeled as being SIFE. Of these 143 students, 84 (58.7%) were both Newcomer and SIFE students. However, we note that 117 (81.8%) of all such students were in the ninth grade. Of these, 30 students were labeled as SIFE only, and they were all enrolled in ESL Sheltered.

Home Language. The majority of students in our cohort (76.0%, n=1397) listed Spanish as their primary language on the home language survey. Only 10.5% (n=192) of students listed a language other than Spanish as their primary home language. Many of these were students from the ninth grade cohort, which had almost a fifth of its students (n=81) list a language other than Spanish as their primary home language. An additional 13.5% (n=248) of students had no primary home language data available. In particular, 45.2% (n=184) of students from the kindergarten cohort had missing primary home language information, compared to less than 6% in the other cohorts.

Within each program, Spanish generally remained the dominant primary home language. However, the Newcomer ESL program for the ninth grade cohort stands out, with 41.4% (n=36) of students listing a language other than Spanish as their primary home language (see Appendix F for details). The ESL Sheltered programs in the upper elementary and secondary cohorts also had relatively high proportions of students (>10%) with a language other than Spanish listed as their primary home language, as did third grade cohort students in the ESL Integrated program.

Table 7. Breakdown of MLL Cohort Demographics

	Kindergarten	3rd Grade	6th Grade	9th Grade	Total
Female	191 (46.9%)	324 (47.6%)	127 (41.8%)	197 (44.3%)	839 (45.7%)
Male	216 (53.1%)	357 (52.4%)	177 (58.2%)	248 (55.7%)	998 (54.3%)
Asian	*	28 (4.1%)	*	19 (4.3%)	67 (3.6%)
Black	24 (5.9%)	28 (4.1%)	18 (5.9%)	47 (10.6%)	117 (6.4%)
Hispanic/Latino	345 (84.8%)	606 (89.0%)	256 (84.2%)	351 (78.9%)	1558 (84.8%)
Native American	0 (0.0%)	*	*	*	*
Pacific Islander	*	*	0 (0.0%)	*	*
Two or More Races	*	*	*	*	18 (1.0%)
White	26 (6.4%)	12 (1.8%)	10 (3.3%)	19 (4.3%)	67 (3.6%)
Free/Reduced Price Lunch	369 (90.7%)	617 (90.6%)	268 (88.2%)	401 (90.1%)	1655 (90.1%)
Not Free/Reduced Price Lunch	38 (9.3%)	64 (9.4%)	36 (11.8%)	44 (9.9%)	182 (9.9%)
SPED	47 (11.5%)	66 (9.7%)	66 (21.7%)	35 (7.9%)	214 (11.6%)
Non-SPED	360 (88.5%)	615 (90.3%)	238 (78.3%)	410 (92.1%)	1623 (88.4%)
Newcomer/SIFE	*	14 (2.1%)	*	117 (26.3%)	143 (7.8%)
Not Newcomer/SIFE	*	667 (97.9%)	*	328 (73.7%)	1694 (92.2%)
Spanish	200 (49.1%)	597 (87.7%)	255 (83.9%)	345 (77.5%)	1397 (76.0%)
Other	23 (5.7%)	57 (8.4%)	31 (10.2%)	81 (18.2%)	192 (10.5%)
Missing	184 (45.2%)	27 (4.0%)	18 (5.9%)	19 (4.3%)	248 (13.5%)
Total	407 (100.0%)	681 (100.0%)	304 (100.0%)	445 (100.0%)	1837 (100.0%)

Table 8. Demographic information by program for the kindergarten cohort.

	Collaborative	Consultation Model	Developmental Bilingual	ENE	ESL Integrated	ESL Sheltered	Transitional Bilingual	Two Way Dual Language	Total
Female	31 (7.6%)	*	27 (6.6%)	*	29 (7.1%)	47 (11.5%)	41 (10.1%)	10 (2.5%)	191 (46.9%)
Male	35 (8.6%)	*	26 (6.4%)	*	37 (9.1%)	69 (17.0%)	33 (8.1%)	10 (2.5%)	216 (53.1%)
Asian	*	0 (0.0%)	0 (0.0%)	0 (0.0%)	*	*	0 (0.0%)	0 (0.0%)	*
Black	*	*	*	*	*	14 (3.4%)	*	0 (0.0%)	24 (5.9%)
Hispanic/Latino	56 (13.8%)	*	51 (12.5%)	*	52 (12.8%)	91 (22.4%)	68 (16.7%)	18 (4.4%)	345 (84.8%)
Native American	0 (0.0%)	0 (0.0%)	*	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	*	0 (0.0%)
Pacific Islander	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	*	0 (0.0%)	0 (0.0%)	*
Two or More Races	*	0 (0.0%)	0 (0.0%)	*	0 (0.0%)	0 (0.0%)	*	0 (0.0%)	*
White	*	0 (0.0%)	0 (0.0%)	0 (0.0%)	*	*	*	*	26 (6.4%)
Free/Reduced Price Lunch	*	*	*	*	56 (13.8%)	101 (24.8%)	*	20 (4.9%)	369 (90.7%)
Not Free/Reduced Price Lunch	*	*	*	0 (0.0%)	10 (2.5%)	15 (3.7%)	*	0 (0.0%)	38 (9.3%)
SPED	18 (4.4%)	*	*	0 (0.0%)	*	*	*	*	47 (11.5%)
Non-SPED	48 (11.8%)	*	*	*	*	*	*	*	360 (88.5%)
Newcomer/SIFE	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	*	0 (0.0%)	0 (0.0%)	*
Not Newcomer/SIFE	66 (16.2%)	*	53 (13.0%)	*	66 (16.2%)	*	74 (18.2%)	20 (4.9%)	*
Spanish	*	*	36 (8.8%)	*	*	53 (13.0%)	37 (9.1%)	*	200 (49.1%)
Other	*	0 (0.0%)	*	0 (0.0%)	*	11 (2.7%)	*	*	23 (5.7%)
Missing	36 (8.8%)	*	*	*	33 (8.1%)	52 (12.8%)	*	10 (2.5%)	184 (45.2%)
Total	*	*	*	*	*	*	*	*	*

Table 9. Demographic information by program for the third grade cohort.

	Collaborative	Consultation Model	Developmental Bilingual	ENE	ESL Integrated	ESL Sheltered	Transitional Bilingual	Two Way Dual Language	Total
Female	38 (5.6%)	*	29 (4.3%)	53 (7.8%)	47 (6.9%)	83 (12.2%)	60 (8.8%)	*	324 (47.6%)
Male	28 (4.1%)	*	31 (4.6%)	85 (12.5%)	53 (7.8%)	104 (15.3%)	44 (6.5%)	*	357 (52.4%)
Asian	*	*	0 (0.0%)	*	*	10 (1.5%)	0 (0.0%)	0 (0.0%)	28 (4.1%)
Black	*	*	*	*	*	*	0 (0.0%)	0 (0.0%)	28 (4.1%)
Hispanic/Latino	59 (8.7%)	*	58 (8.5%)	124 (18.2%)	81 (11.9%)	158 (23.2%)	102 (15.0%)	*	606 (89.0%)
Native American	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	*	0 (0.0%)	0 (0.0%)	*
Pacific Islander	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	*	0 (0.0%)	0 (0.0%)	0 (0.0%)	*
Two or More Races	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	*	0 (0.0%)	*	0 (0.0%)	*
White	*	0 (0.0%)	*	0 (0.0%)	0 (0.0%)	*	*	0 (0.0%)	12 (1.8%)
Free/Reduced Price Lunch	*	*	*	127 (18.6%)	87 (12.8%)	164 (24.1%)	*	*	617 (90.6%)
Not Free/Reduced Price Lunch	*	0 (0.0%)	*	11 (1.6%)	13 (1.9%)	23 (3.4%)	*	0 (0.0%)	64 (9.4%)
SPED	13 (1.9%)	*	*	24 (3.5%)	10 (1.5%)	*	*	*	66 (9.7%)
Non-SPED	53 (7.8%)	*	*	114 (16.7%)	90 (13.2%)	*	*	*	615 (90.3%)
Newcomer/SIFE	*	*	0 (0.0%)	0 (0.0%)	*	11 (1.6%)	0 (0.0%)	0 (0.0%)	14 (2.1%)
Not Newcomer/SIFE	*	*	60 (8.8%)	138 (20.3%)	*	176 (25.8%)	104 (15.3%)	*	667 (97.9%)
Spanish	55 (8.1%)	*	58 (8.5%)	123 (18.1%)	81 (11.9%)	154 (22.6%)	101 (14.8%)	*	597 (87.7%)
Other	*	*	*	15 (2.2%)	*	23 (3.4%)	*	0 (0.0%)	57 (8.4%)
Missing	*	0 (0.0%)	*	0 (0.0%)	*	10 (1.5%)	*	0 (0.0%)	27 (4.0%)
Total	*	*	*	*	*	*	*	*	*

Table 10. Demographic information by program for the sixth grade cohort.

	Collaborative	Consultation Model	ENE	ESL Sheltered	Newcomer ESL	Total
Female	*	*	*	112 (36.8%)	*	127 (41.8%)
Male	*	*	*	139 (45.7%)	*	177 (58.2%)
Asian	0 (0.0%)	0 (0.0%)	0 (0.0%)	12 (3.9%)	0 (0.0%)	12 (3.9%)
Black	*	0 (0.0%)	*	17 (5.6%)	0 (0.0%)	18 (5.9%)
Hispanic/Latino	18 (5.9%)	*	28 (9.2%)	204 (67.1%)	*	256 (84.2%)
Native American	0 (0.0%)	0 (0.0%)	0 (0.0%)	*	0 (0.0%)	*
Pacific Islander	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
Two or More Races	*	0 (0.0%)	0 (0.0%)	*	0 (0.0%)	*
White	0 (0.0%)	0 (0.0%)	*	10 (3.3%)	*	10 (3.3%)
Free/Reduced Price Lunch	*	*	*	217 (71.4%)	*	268 (88.2%)
Not Free/Reduced Price Lunch	*	0 (0.0%)	*	34 (11.2%)	0 (0.0%)	36 (11.8%)
SPED	19 (6.2%)	*	16 (5.3%)	28 (9.2%)	*	66 (21.7%)
Non-SPED	*	0 (0.0%)	12 (3.9%)	223 (73.4%)	*	238 (78.3%)
Newcomer/SIFE	0 (0.0%)	0 (0.0%)	0 (0.0%)	*	*	11 (3.6%)
Not Newcomer/SIFE	19 (6.2%)	*	28 (9.2%)	*	*	293 (96.4%)
Spanish	17 (5.6%)	*	28 (9.2%)	206 (67.8%)	*	255 (83.9%)
Other	*	0 (0.0%)	*	30 (9.9%)	0 (0.0%)	31 (10.2%)
Missing	*	*	*	15 (4.9%)	0 (0.0%)	18 (5.9%)
Total	*	*	*	*	*	*

Table 11. Demographic information by program for the ninth grade cohort.

	Collaborative	Consultation Model	ENE	ESL Sheltered	Newcomer ESL	Total
Female	*	*	16 (3.6%)	149 (33.5%)	27 (6.1%)	197 (44.3%)
Male	*	*	24 (5.4%)	157 (35.3%)	60 (13.5%)	248 (55.7%)
Asian	*	0 (0.0%)	*	12 (2.7%)	*	19 (4.3%)
Black	*	*	0 (0.0%)	31 (7.0%)	14 (3.1%)	47 (10.6%)
Hispanic/Latino	*	*	35 (7.9%)	242 (54.4%)	65 (14.6%)	351 (78.9%)
Native American	0 (0.0%)	0 (0.0%)	0 (0.0%)	*	*	*
Pacific Islander	0 (0.0%)	0 (0.0%)	*	0 (0.0%)	0 (0.0%)	*
Two or More Races	0 (0.0%)	0 (0.0%)	0 (0.0%)	*	*	*
White	0 (0.0%)	0 (0.0%)	*	15 (3.4%)	*	19 (4.3%)
Free/Reduced Price Lunch	*	*	*	277 (62.2%)	77 (17.3%)	401 (90.1%)
Not Free/Reduced Price Lunch	*	*	*	29 (6.5%)	10 (2.2%)	44 (9.9%)
SPED	*	*	*	15 (3.4%)	*	35 (7.9%)
Non-SPED	*	*	*	291 (65.4%)	*	410 (92.1%)
Newcomer/SIFE	*	0 (0.0%)	0 (0.0%)	30 (6.7%)	*	117 (26.3%)
Not Newcomer/SIFE	10 (2.2%)	*	40 (9.0%)	276 (62.0%)	*	328 (73.7%)
Spanish	*	*	37 (8.3%)	250 (56.2%)	51 (11.5%)	345 (77.5%)
Other	*	0 (0.0%)	*	41 (9.2%)	36 (8.1%)	81 (18.2%)
Missing	*	*	*	15 (3.4%)	0 (0.0%)	19 (4.3%)
Total	*	*	*	*	*	*

4. Descriptive Analyses

4.1 Test Scores

Throughout this section, we provide descriptive analyses on test scores (ACCESS, STAR, PARCC/RICAS, SAT) only for students who met the complete case requirements laid out in Section 2.2.3 Complete Cases.

4.1.1 ACCESS Scores

The WIDA ACCESS is a suite of summative assessments designed for monitoring K-12 students' English language proficiency annually. ACCESS is a standards-referenced test, meaning that students can achieve any score, and that their performance is compared against a set of WIDA-defined English language development standards, rather than against each other or the expected English language proficiency of students whose first language is English (WIDA Consortium, 2021a).

All MLL students in PPSD take the ACCESS test annually (typically between January and February) to monitor their progress in English language acquisition. Students who exit MLL services during our study period no longer complete the ACCESS test, but still remain in our cohort, and hence may be missing this test score.

The ACCESS exam provides two interpretations of their results: a scale score and a performance level. The performance level provides a quick way for educators to understand what English language level a child has achieved, whereas the scale score is more useful in regression analyses. The performance level is simply a transformation of the scale score. For descriptive purposes, we use the performance level, but for our quasi-experimental analyses in the next section we use the scale score. Table 12 compares SY 2016-17 and SY 2018-19 mean ACCESS performance levels by cohort.

On average, ACCESS performance levels improved between the SY 2016-17 and SY 2018-19 across all cohorts and across most programs. We observe that students in the kindergarten cohort started, as expected, with the lowest mean ACCESS performance level (1.77); by SY 2018-19, these students saw a 75.1 percentage change in average ACCESS performance levels, the largest across all cohorts. Students in the third grade cohort had the highest starting mean ACCESS score (3.14) in SY 2016-17, and had a 27.7 percentage change in mean ACCESS performance level by SY 2018-19. Because of the higher beginning mean ACCESS performance level, we may expect the percentage change to be smaller relative to what students in the kindergarten cohort experienced, i.e. we may expect that students who begin at a lower

proficiency level will make larger gains on the ACCESS initially (e.g. within a school year, a student may move from level 1 to level 2 more quickly than from level 3 to level 4).

In contrast to the language proficiency growth observed in students in the kindergarten and third grade cohorts, students in the sixth and ninth grade cohorts demonstrated limited change in mean ACCESS performance levels over the three year period. The average test scores in both cohorts were similar; both had lower starting mean ACCESS performance levels (2.90 in the sixth grade cohort and 2.91 in the ninth grade cohort) compared to the third grade cohort, and saw a percentage change in mean ACCESS performance levels of 0.1 and 0.2, respectively.

Change in mean ACCESS performance levels from SY 2016-17 to SY 2018-19 was uneven across programs.[35] Notably, kindergarten cohort students in the Developmental and Transitional Bilingual programs more than doubled their average ACCESS performance levels in the final year of our study, from 1.42 to 3.12 and from 1.51 to 3.14, respectively. Students in the Two Way Dual Language program also saw tremendous growth, increasing their average ACCESS performance levels by 81.3% from 1.82 to 3.30. In the third grade cohort, students in the Developmental Bilingual (n=42) and ESL Sheltered (n=106) programs saw the greatest increase in average ACCESS scores, at 34.0% and 32.6%, respectively.

On the other hand, some programs saw a slight decrease in mean ACCESS performance levels over the same time period, as such with sixth grade cohort students in the Collaborative ESL program (n=16) and those ENE (n=22), though we note that these sample sizes are small.

Relatedly, the availability of ACCESS performance levels varied greatly by cohort. Of the 407 MLL students in our kindergarten cohort, 89.9% (n=366) of them had available overall ACCESS performance levels[36] in both time periods; this proportion falls to 64.2% (n=437) and 69.7% (n=212) for the 3rd and sixth grade cohorts' 681 and 304 students, respectively, and is lowest for the ninth grade cohort, in which just 39.1% (n=174) of their 445 students had a complete pair of ACCESS performance levels. The large degree to which these scores are missing, particularly in the upper grades, may limit the certainty with which we can draw even observational conclusions about student progress on the ACCESS test.

Finally, although each MLL program is in theory designed to meet the unique needs of students with different levels of English language proficiency (as evidenced by each program's entry criteria [see Appendix A]), in practice, most programs during the duration of this study have students with a range of English language proficiency enrolled, as seen in the wide distribution of SY 2016-17 and SY 2018-19 ACCESS performance levels in many MLL programs (Figure 3).

[35] SY 2018-19 results were unavailable by program for the Consultation Model that had been discontinued during the duration of our study (see Appendix B and Appendix C). In certain cases, SY 2018-19 ACCESS scores were unavailable if no student in that grade cohort had participated in the specific MLL program that year.

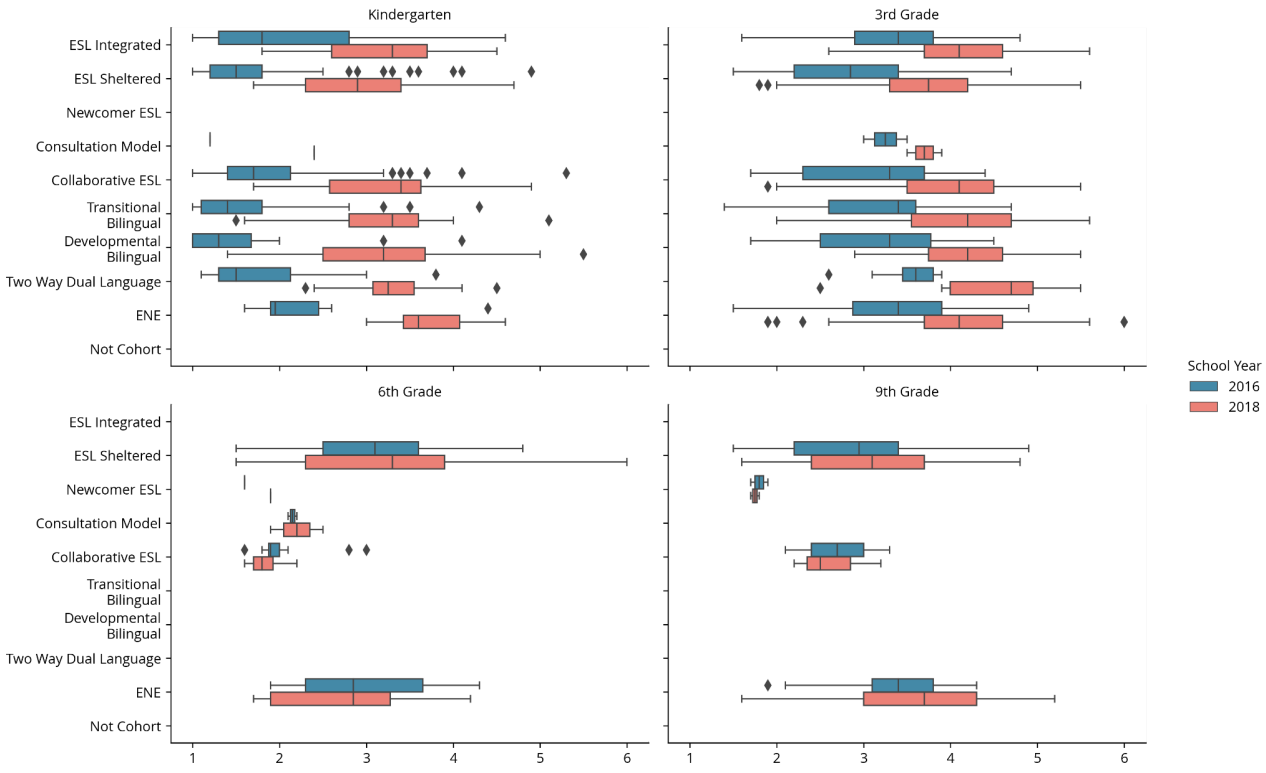
[36] For overall ACCESS scores to be available, a student must have completed all four ACCESS test components - listening, speaking, reading, and writing.

Table 12. Descriptive statistics of ACCESS scores by cohort and program, SY 2016-17 vs SY 2018-19

		2016		2018	
		n	Mean (SD)	n	Mean (SD)
Kindergarten	Collaborative	60	2.0 (0.8)	114	3.4 (0.7)
	Developmental Bilingual	46	1.4 (0.6)	26	3.3 (1.0)
	ENE	*	*	*	*
	ESL Integrated	61	2.2 (1.0)	*	*
	ESL Sheltered	105	1.7 (0.7)	93	2.7 (0.6)
	Transitional Bilingual	67	1.5 (0.6)	59	3.1 (0.7)
	Two Way Dual Language	20	1.8 (0.8)	42	3.2 (0.6)
	Cohort	366	1.8 (0.8)	366	3.1 (0.8)
3rd Grade	Collaborative	37	3.1 (0.8)	231	3.9 (0.6)
	Developmental Bilingual	42	3.1 (0.8)	0	N/A
	ENE	96	3.3 (0.7)	97	4.1 (0.8)
	ESL Integrated	68	3.3 (0.7)	34	3.3 (0.7)
	ESL Sheltered	106	2.8 (0.8)	15	3.3 (1.0)
	Transitional Bilingual	75	3.2 (0.7)	0	N/A
	Two Way Dual Language	11	3.5 (0.4)	10	4.7 (0.6)
	Cohort	437	3.1 (0.8)	437	4.0 (0.8)
6th Grade	Collaborative	16	2.0 (0.4)	110	3.1 (0.9)
	Developmental Bilingual	0	N/A	0	N/A
	ENE	22	3.0 (0.7)	38	3.2 (1.0)
	ESL Integrated	0	N/A	0	N/A
	ESL Sheltered	171	3.0 (0.8)	55	2.5 (0.8)
	Transitional Bilingual	0	N/A	0	N/A
	Two Way Dual Language	0	N/A	0	N/A
	Cohort	212	2.9 (0.8)	212	3.0 (1.0)

	Collaborative	*	*	91	3.4 (0.7)
	Developmental Bilingual	0	N/A	0	N/A
	ENE	29	3.3 (0.6)	31	3.5 (0.9)
	ESL Integrated	0	N/A	0	N/A
9th Grade	**ESL Sheltered**	140	2.8 (0.8)	50	2.4 (0.6)
	Transitional Bilingual	0	N/A	0	N/A
	Two Way Dual Language	0	N/A	0	N/A
	Cohort	174	2.9 (0.8)	174	3.1 (0.9)

Figure 3. Comparison of ACCESS levels by cohort and program, SY 2016-17 vs. SY 2018-19



ALT ACCESS Scores

Certain students enrolled in the SPED program were eligible to take the ALT ACCESS assessment, a paper-based assessment that is administered individually to SPED students with significant cognitive disabilities (WIDA Consortium, 2021b). Within our cohort, a total of 29 students took the ALT ACCESS assessment in SY 2016-17, 69.0% of whom were enrolled in the Collaborative ESL program (Table 13).

Almost all of these 29 students met the criteria for "complete cases". Overall, this was 12.6% of SPED students in our cohort, or 1.5% of students in our cohort overall. These students were excluded from our analyses as the ACCESS and ALT ACCESS assessments used different scoring systems (see Footnote 8 and Section 6. Limitations and Implications for more details).

Table 13: Breakdown of number of students who took ALT ACCESS by MLL program and SPED status.

		Collaborative ESL	Consultation Model	ENE	Newcomer ESL	ESL Sheltered	Total
SPED	**Kindergarten**	*	*	0 (0.0%)	0 (0.0%)	0 (0.0%)	*
	3rd Grade	*	*	*	0 (0.0%)	0 (0.0%)	15 (100.0%)
	6th Grade	*	0 (0.0%)	*	0 (0.0%)	0 (0.0%)	*
	9th Grade	*	0 (0.0%)	*	*	*	*
Not SPED	**3rd Grade**	0 (0.0%)	*	0 (0.0%)	0 (0.0%)	0 (0.0%)	*
	9th Grade	*	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	*
Total		20 (69.0%)	*	*	*	*	29 (100.0%)

4.1.2 Renaissance STAR Progress Monitoring Test Scores

The Renaissance STAR Early Literacy and STAR Reading assessments are computer-adaptive, benchmark, and progress monitoring assessments of students' reading skills, while the Renaissance STAR Math is a computer-adaptive, benchmark, and progress monitoring assessment of students' math skills. Students may take these assessments multiple times over the course of a school year.

Students in kindergarten or first grade typically complete the STAR Early Literacy assessment, a criterion-referenced assessment which measures foundational skills (i.e. "skills in ten content areas essential to reading readiness") (Renaissance Learning, Inc., 2018, p.1), in lieu of the

STAR Reading assessment, a norm-referenced assessment that measures reading comprehension (Renaissance Learning, Inc., 2018). Students who meet the criteria to exit STAR Early Literacy are able to complete the STAR Reading assessment instead; conversely, students in the second grade who have not met the criteria to exit to STAR Reading may complete the STAR Early Literacy assessment in addition to the STAR Reading.

We use student's SY 2018-19 EOY STAR Early Literacy, Reading, and Math test scores to conduct the following descriptive analyses; however, as mentioned above, we omit the STAR test scores from the ninth grade cohort as test result data was only available for a small subset of students.

Kindergarten Cohort

310 kindergarten students took the STAR Early Literacy assessment at the beginning of SY 2016-17. Students scored an average of 403.5 on the assessment. For MLL programs with a minimum of 20 students enrolled (MLL program enrollment varied widely across programs, and both the ENE and Consultation Model programs had <10 students enrolled), average test scores ranged from 386.1 (Transitional Bilingual) to 421.1 (ESL Integrated) (Table 14).

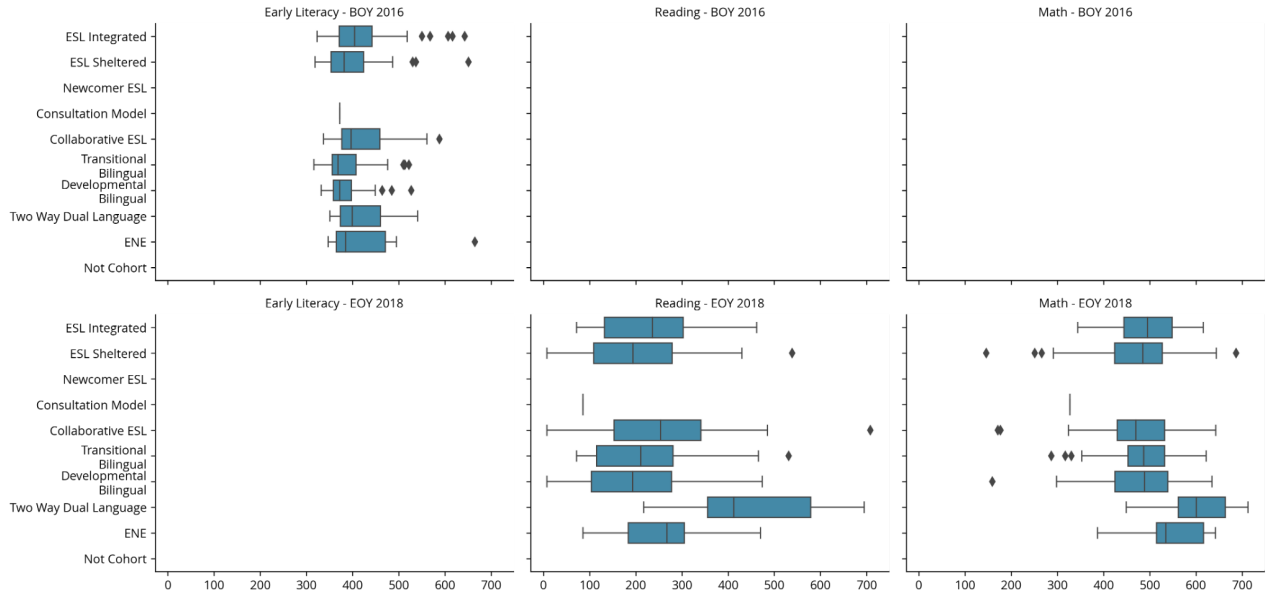
At the end of SY 2018-19, 339 students in our kindergarten cohort completed the STAR Reading assessment. The average test score across all students was 232.0. Students in the Two Way Dual Language program (n=20) had a mean score of 451.9, the highest test average across all programs. Notably, students in this program had average STAR Early Literacy test scores in SY 2016-17 that were higher than the kindergarten cohort average. In contrast, students in the ESL Sheltered and Developmental Bilingual programs had the lowest STAR Reading scores, at 200.9 and 201.9, respectively (although the average in the Developmental Bilingual program appears to be driven down by eight students [Figure 4]). Both programs along with the Transitional Bilingual program also had some of the lowest program averages on the STAR Early Literacy; however, while Transitional Bilingual students had comparable average STAR Early Literacy scores, their average on the STAR Reading was about 15 points higher than students in the ESL Sheltered and Developmental Bilingual programs.

We observe a much wider distribution of mean test scores across programs on the STAR Math assessment. The cohort average was 486.77 in SY 2018-19, with MLL program test averages ranging from 470.6 (Collaborative) to 608.3 (Two Way Dual Language). As with the STAR Reading assessment, students in the Two Way Dual Language MLL program (n=20) had the highest program test average.

Table 14. Descriptive statistics of STAR Early Literacy, Reading, and Mathematics scores for the kindergarten cohort, by program, Fall 2016-17 vs. Spring 2018-19.

		BOY 2016		EOY 2018			
		Early Literacy		Math		Reading	
		n	Mean (SD)	n	Mean (SD)	n	Mean (SD)
Kindergarten	**Cohort**	310	403.51 (60.85)	331	486.77 (90.68)	354	232.19 (125.57)
	ESL Integrated	52	421.08 (73.42)	58	493.03 (70.11)	59	224.49 (104.62)
	ESL Sheltered	90	394.50 (53.21)	93	472.52 (89.74)	100	200.93 (100.05)
	Consultation Model	*	*	*	*	*	*
	Collaborative ESL	54	420.02 (63.52)	55	470.60 (96.26)	55	262.31 (133.23)
	Transitional Bilingual	56	386.05 (47.27)	52	482.12 (72.62)	67	215.72 (107.97)
	Developmental Bilingual	31	386.58 (44.38)	46	475.76 (96.19)	46	201.91 (111.96)
	Two Way Dual Language	20	419.55 (60.10)	20	608.30 (65.20)	20	451.90 (137.14)
	ENE	*	*	*	*	*	*

Figure 4. Comparison of STAR scores for the kindergarten cohort by program and assessment year.



Third and Sixth Grade Cohorts

We compare average SY 2016-17 BOY STAR Reading and Math test scores with those from EOY SY 2018-19 to make observations about student growth. Across both cohorts and all MLL programs, average STAR Reading and Math test scores increased (Table 15, Figure 5).

In the third grade cohort, we observe a 232.2 point increase in average STAR Reading test scores, and a 193.8 point increase in average STAR Math test scores. Students' average STAR Reading scores were much lower than their average STAR Math scores at the beginning of our study, at 177.2 and 449.7, respectively. Consequently, though students' average test scores saw a greater increase in Reading, at the conclusion of our study, average Math test scores were higher (643.5) than for Reading (409.4).

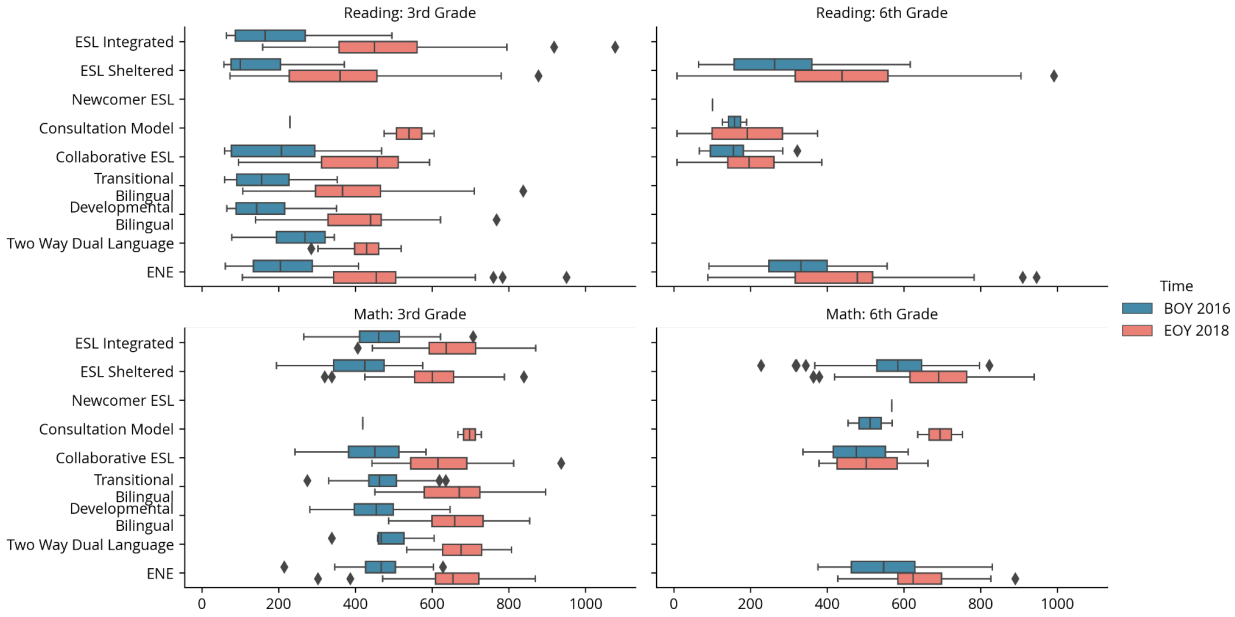
Across programs with at least 20 students enrolled (as with the kindergarten cohort, the number of students enrolled in each MLL program varied from program to program), third grade cohort students in the ESL Sheltered program had the lowest average Reading and Math scores in SY 2016-17 and SY 2018-19.

In the sixth grade cohort, we observe a 163.4 point increase in average STAR Reading test scores, and a 98.0 point increase in average STAR Math test scores. This is about two-thirds the increase in average Reading test scores and half the increase in average Math test scores that we saw with the third grade cohort. However, students in the sixth grade cohort had starting average test scores that were roughly 100 points higher than that of the third grade cohort in each subject, and it is possible that the higher average test scores in SY 2016-17 had an impact on the rate of increase in average test scores.

Table 15. Descriptive statistics of STAR Reading and Mathematics scores for the third grade and sixth grade cohorts, by program, Fall 2016-17 vs. Spring 2018-19.

		Reading				Math			
		BOY 2016		EOY 2018		BOY 2016		EOY 2018	
		n	Mean (SD)	n	Mean (SD)	n	Mean (SD)	n	Mean (SD)
3rd Grade	Cohort	416	177.19 (95.16)	369	409.43 (153.77)	425	449.68 (81.46)	374	643.46 (102.47)
	ESL Integrated	67	185.24 (110.90)	56	461.39 (177.35)	67	460.75 (88.79)	59	652.75 (105.14)
	ESL Sheltered	95	141.12 (79.32)	78	351.50 (166.51)	101	410.94 (85.10)	91	604.35 (89.89)
	Consultation Model	*	*	*	*	*	*	*	*
	Collaborative ESL	36	194.92 (112.10)	34	409.94 (133.11)	35	439.54 (91.41)	36	626.36 (115.67)
	Transitional Bilingual	71	161.94 (76.24)	67	386.25 (139.40)	74	467.27 (67.51)	51	663.37 (117.46)
	Developmental Bilingual	41	159.85 (82.06)	39	414.79 (116.02)	42	452.02 (74.97)	38	664.13 (91.34)
	Two Way Dual Language	*	*	*	*	*	*	*	*
	ENE	94	211.97 (93.61)	83	440.75 (149.53)	94	468.07 (68.55)	87	659.44 (96.12)
6th Grade	Cohort	203	260.20 (131.31)	195	423.64 (214.04)	198	569.77 (105.63)	180	667.79 (119.77)
	ESL Sheltered	163	263.25 (132.40)	155	444.71 (205.77)	159	581.30 (102.44)	145	685.23 (112.94)
	Newcomer ESL	N/A	N/A	*	*	N/A	N/A	*	*
	Consultation Model	*	*	*	*	*	*	*	*
	Collaborative ESL	*	*	15	199.00 (98.07)	*	*	11	500.09 (96.31)
	ENE	22	322.68 (116.18)	22	464.09 (228.01)	22	554.41 (112.00)	21	637.43 (111.17)

Figure 5. Comparison of STAR Reading and Mathematics scores for the third and sixth grade cohorts, by program, SY 2016-17 (Fall) vs. SY 2018-19 (Spring).



4.1.3 PARCC and RICAS Scores

Between SY 2016-17 and SY 2018-19, PPSD switched its summative assessment from the PARCC to the RICAS. Both assessments are aligned with the Common Core State Standards (CCSS). To our knowledge, there has been no linking study conducted to connect these two assessments; hence, we are unable to psychometrically link these two assessments in a way that illustrates change in ELA and mathematics test scores over time.

Table 16 summarizes the average PARCC and RICAS scores by year, cohort, and program for students with a set of test scores that met the complete case definition described earlier in this report (see Section 2.2.3 Complete Cases). Students in our kindergarten cohort did not take these summative assessments.

Comparing SY 2017-18 and SY 2018-19 RICAS ELA and Math[37] scores, we observe overall minimal movement or change, regardless of age or MLL program type.

[37] Students in ninth grade took the Algebra exam.

Table 16. Descriptive statistics of state standardized test scores (PARCC and RICAS), by cohort, program, and assessment year.

		2016		2017		2018		2016		2017		2018	
		PARCC ELA		RICAS ELA		RICAS ELA		PARCC Math		RICAS Math		RICAS Math	
		n	Mean (SD)	n	Mean (SD)	n	Mean (SD)	n	Mean (SD)	n	Mean (SD)	n	Mean (SD)
3rd Grade	Cohort	437	693.14 (25.79)	435	471.42 (15.21)	437	474.14 (16.06)	437	710.49 (29.18)	435	467.60 (17.30)	437	471.45 (16.75)
	ESL Integrated	68	700.29 (27.34)	67	475.28 (13.13)	68	478.29 (15.25)	68	718.69 (32.31)	67	471.33 (15.64)	68	475.03 (17.38)
	ESL Sheltered	106	686.95 (24.44)	105	466.51 (15.04)	106	466.98 (16.41)	106	703.95 (27.28)	105	461.90 (16.87)	106	467.58 (13.14)
	Consultation Model	*	*	*	*	*	*	*	*	*	*	*	*
	Collaborative ESL	37	692.03 (26.67)	37	469.24 (12.12)	37	474.76 (13.67)	37	708.41 (35.61)	37	465.00 (17.41)	37	466.32 (20.45)
	Transitional Bilingual	75	691.84 (25.76)	75	470.68 (13.29)	75	471.08 (14.62)	75	713.71 (29.10)	75	467.88 (18.69)	75	470.53 (16.51)
	Developmental Bilingual	42	687.12 (20.66)	42	473.10 (15.67)	42	480.71 (15.58)	42	705.76 (20.75)	42	466.88 (12.91)	42	473.64 (16.69)
	Two Way Dual Language	*	*	*	*	*	*	*	*	*	*	*	*
	ENE	96	697.99 (26.76)	96	475.24 (17.91)	96	477.41 (15.38)	96	711.69 (28.64)	96	471.75 (18.17)	96	474.59 (17.43)
6th Grade	Cohort	212	694.96 (15.88)	207	452.55 (11.41)	212	457.69 (14.36)	212	686.67 (21.56)	209	458.55 (12.37)	212	462.61 (12.23)
	ESL Sheltered	171	695.83 (15.54)	167	453.61 (11.37)	171	459.34 (14.89)	171	688.52 (20.93)	168	459.41 (12.51)	171	463.82 (12.13)
	Newcomer ESL	*	*	*	*	*	*	*	*	*	*	*	*
	Consultation Model	*	*	*	*	*	*	*	*	*	*	*	*
	Collaborative ESL	16	684.62 (16.07)	16	441.62 (1.36)	16	446.69 (6.76)	16	677.31 (15.19)	16	454.50 (9.09)	16	454.25 (9.14)
	ENE	22	696.50 (17.34)	21	453.19 (12.28)	22	453.95 (9.76)	22	682.45 (26.95)	22	454.36 (13.11)	22	459.77 (13.13)
9th Grade	Cohort	174	687.99 (19.11)	N/A	N/A	N/A	N/A	174	698.38 (19.29)	N/A	N/A	N/A	N/A
	ESL Sheltered	140	687.03 (17.65)	N/A	N/A	N/A	N/A	140	697.49 (19.22)	N/A	N/A	N/A	N/A
	Newcomer ESL	*	*	N/A	N/A	N/A	N/A	*	*	N/A	N/A	N/A	N/A
	Collaborative ESL	*	*	N/A	N/A	N/A	N/A	*	*	N/A	N/A	N/A	N/A
	ENE	29	696.34 (22.59)	N/A	N/A	N/A	N/A	29	703.28 (20.42)	N/A	N/A	N/A	N/A

4.1.4 SAT Scores

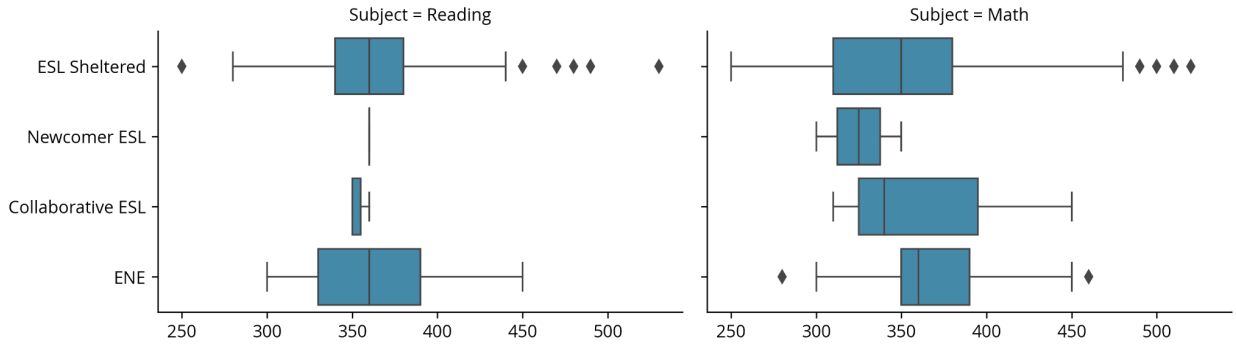
Of the 445 ninth grade students in our cohort, 174 sat for the SAT Reading and Math tests between SY 2016-17 and SY 2018-19. SAT scores were low across the board (Table 17), with students scoring an average of 361.5 and 357.5 out of a possible 800 each (and a floor of 200) for the SAT Reading and Math, respectively. We observe little variation in average SAT scores by MLL program and subject.

We also observe that there are a number of outlying students in a couple of MLL programs (Figure 6). These students did either much better or worse than the average of their peers. Notably, there was a student from the ESL Sheltered program who scored a 530 on the SAT Reading. Similarly, there were a few students from the ESL Sheltered program who scored between 500 and 520 on the SAT Math.

Table 17. Descriptive statistics of SAT scores for the ninth grade cohort by program.

		Reading		Math	
		n	Mean (SD)	n	Mean (SD)
9th Grade	ESL Sheltered	140	361.36 (39.13)	140	355.21 (54.79)
	Newcomer ESL	*	*	*	*
	Collaborative ESL	*	*	*	*
	ENE	29	363.10 (34.75)	29	369.66 (46.71)
All		174	361.49 (37.79)	174	357.47 (53.65)

Figure 6. Comparison of SAT scores for the ninth grade cohort by program.



4.2 MLL Program Exit Rates

At the end of each school year, MLL students who meet the criteria to exit the MLL program (Appendix D) are moved out of the program and into two years of Monitored status. PPSD monitors the academic performance of former MLL students twice a year, for two years after they exit the MLL program, by reviewing their standardized test scores, composite and domain scores on the English language proficiency test at the time of exit, and progress reports for grades, attendance, preparation, and behavior to determine if the student needs any academic support services (e.g. tutoring) or needs to be retested for possible reentry into the MLL Program. If a former MLL student fails to make academic progress and if an ESL-Certified Teacher, an administrator, and core-content teachers determine that this failure may be due to a lack of English proficiency, PPSD will notify the student's parents and offer MLL services, following which PPSD will provide the student with the services that the parents accept.

Table 18 summarizes the MLL program exit rates by cohort and program. We use SY 2019-20 MLL student status data that PPSD reports annually to the Rhode Island Department of Education (RIDE) to determine whether a student exited MLL services. The "Monitored Year 1" column indicates the number of students who moved out of receiving MLL services at the end of SY 2018-19. Correspondingly, students in the "Monitored Year 2" column exited the MLL program at the end of SY 2017-18, while students who were "no longer classified as MLL" (i.e. were no longer under Monitored status) exited the MLL program at the end of SY 2016-17 (assuming they completed two years under Monitored status). Additionally, there are a handful of students in each cohort for whom we do not know their MLL status ("unknown").

Overall, exit rates remain low across all cohorts except the third grade cohort, where exit rates appeared fairly high at the end of certain school years.

By the end of SY 2018-19, 286 out of 1837 students (11.2%) in our overall cohort exited MLL services. Among our cohorts, exit rates were highest for the third grade cohort (30.1%, n=205) and lowest for the ninth grade cohort (6.1%, n=27).

We note that the pattern of MLL program exits is such that it may be driven by certain milestones in a student's life. In particular, while exit rates for the third grade cohort were low at the end of SY 2018-19 (indicated by being in Monitored Year 1 in SY 2019-20, i.e. when students were in the sixth grade), these rates were much higher at the end of the two previous school years, when students would have been entering the fourth and fifth grades. At the end of SY 2016-17, 115 students (16.9% of the cohort) exited the MLL program (indicated by the column No Longer Classified as MLL) when they were likely in the third grade; another 82 students (12.0% of the cohort) exited the MLL program at the end of SY 2017-18 (indicated by the column Monitored Year 2), when they were likely in the fourth grade. In total, 197 students or 28.9% of the third grade cohort exited the MLL program between SY 2016-17 and SY 2017-18.

This aligns with conversations with PPSD, which is that when students do exit the MLL program, they tend to do so while in their middle elementary years.

Table 18. MLL status in SY 2019-20 by cohort and program.

		Classified as EL	Monitored Year 1	Monitored Year 2	No Longer Classified as EL	Unknown	Total
Kindergarten	Cohort	356 (100.0%)	25 (100.0%)	*	0 (0.0%)	20 (100.0%)	407 (100.0%)
	ESL Integrated	59 (16.6%)	*	0 (0.0%)	0 (0.0%)	*	66 (16.2%)
	ESL Sheltered	103 (28.9%)	*	0 (0.0%)	0 (0.0%)	*	116 (28.5%)
	Consultation Model	*	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	*
	Collaborative ESL	56 (15.7%)	*	*	0 (0.0%)	*	66 (16.2%)
	Transitional Bilingual	64 (18.0%)	*	*	0 (0.0%)	*	74 (18.2%)
	Developmental Bilingual	46 (12.9%)	0 (0.0%)	*	0 (0.0%)	*	53 (13.0%)
	Two Way Dual Language	*	*	0 (0.0%)	0 (0.0%)	0 (0.0%)	20 (4.9%)
	ENE	*	0 (0.0%)	0 (0.0%)	0 (0.0%)	*	*
3rd Grade	Cohort	423 (100.0%)	*	82 (100.0%)	115 (100.0%)	53 (100.0%)	681 (100.0%)
	ESL Integrated	62 (14.7%)	0 (0.0%)	15 (18.3%)	14 (12.2%)	*	100 (14.7%)
	ESL Sheltered	135 (31.9%)	*	13 (15.9%)	24 (20.9%)	13 (24.5%)	187 (27.5%)
	Consultation Model	*	0 (0.0%)	0 (0.0%)	0 (0.0%)	*	*
	Collaborative ESL	39 (9.2%)	0 (0.0%)	10 (12.2%)	13 (11.3%)	*	66 (9.7%)
	Transitional Bilingual	60 (14.2%)	*	15 (18.3%)	20 (17.4%)	*	104 (15.3%)
	Developmental Bilingual	39 (9.2%)	*	*	11 (9.6%)	*	60 (8.8%)
	Two Way Dual Language	*	0 (0.0%)	*	*	*	19 (2.8%)
	ENE	75 (17.7%)	*	20 (24.4%)	27 (23.5%)	14 (26.4%)	138 (20.3%)
6th Grade	Cohort	238 (100.0%)	*	*	*	43 (100.0%)	304 (100.0%)
	ESL Sheltered	193 (81.1%)	*	*	*	35 (81.4%)	251 (82.6%)
	Newcomer ESL	*	0 (0.0%)	0 (0.0%)	0 (0.0%)	*	*
	Consultation Model	*	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	*
	Collaborative ESL	17 (7.1%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	*	19 (6.2%)
	ENE	23 (9.7%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	*	28 (9.2%)
9th Grade	Cohort	339 (100.0%)	*	*	19 (100.0%)	79 (100.0%)	445 (100.0%)
	ESL Sheltered	234 (69.0%)	*	*	15 (78.9%)	51 (64.6%)	306 (68.8%)
	Newcomer ESL	63 (18.6%)	0 (0.0%)	0 (0.0%)	*	*	87 (19.6%)
	Consultation Model	*	0 (0.0%)	0 (0.0%)	0 (0.0%)	*	*
	Collaborative ESL	*	0 (0.0%)	0 (0.0%)	*	*	*
	ENE	35 (10.3%)	0 (0.0%)	*	*	*	40 (9.0%)
Total		1356 (100.0%)	44 (100.0%)	101 (100.0%)	141 (100.0%)	195 (100.0%)	1837 (100.0%)

4.3 Student Behavior

4.3.1 Student Attendance

To examine student attendance, we divide student attendance into the following five categories:

Table 19. Student attendance categories and definitions.

Absence Category	Definition[38]
Low	Missed <5% days
Moderate	Missed 5-10% days
Chronic	Missed 10-20% days
Excessive	Missed 20-50% days
Mostly Absent	Missed >50% days

Across all cohorts, absence rates were high (Figure 7, Table 20), with at least half of the students in each cohort having moderate or more severe absence levels. Overall, the third grade cohort (n=681) had the lowest absenteeism rates, with 54.0% of students having absence rates that fell into the moderate to severe categories. This cohort also had the lowest proportion (19.5%) of students with chronic and excessive absences. Within programs that had over 20 enrolled students, chronic and excessive absences were highest for ENE students (23.2%), students in the Collaborative ESL program (27.2%), and students in the ESL Integrated program (23.0%).

Absence levels were similar but slightly higher for students in the sixth grade cohort (n=304), most of whom participated in the ESL Sheltered program (n=251). 55.3% of students had moderate to excessive absence rates, including over one in four (26.3%) students with chronic or excessive absences. Within the ESL Sheltered program, 24.7% of students were considered chronically or excessively absent.

Students in the kindergarten cohort had absenteeism rates on par with that of the ninth grade cohort, which had the highest absenteeism rates across all four cohorts. However, in the kindergarten cohort, these rates were driven more by moderate absenteeism levels (39.3%, compared to the ninth grade cohort's 31.9). In contrast, the ninth grade cohort had the highest proportion of chronic and excessive absences across all cohorts (35.3%, compared to 27.8% in

[38] We use the definitions presented by Schwartz and Bisht (2020) for the absence categories of low, moderate, chronic, and excessive.

the kindergarten cohort and the overall cohort average of 25.9%). The ninth grade cohort was also the only cohort with students that had mostly absent levels of absenteeism; four of these six students were enrolled in the ESL Sheltered program. Finally, within the kindergarten cohort, in programs with over 20 students enrolled, the Collaborative (36.3%), ESL Sheltered (33.6%), and ESL Integrated (28.7%) programs had the highest rates of chronic and excessive absenteeism. Within the ninth grade cohort, these rates were 42.5% for ENE students, 34.5% in the Newcomer program, and 34.0% in the ESL Sheltered program i.e. above 30% for all programs in which over 20 students were enrolled.

We note that these levels of absence are similar or better compared to those experienced by all enrollees in PPSD during SY 2016-17 (Table 21). Indeed, in the elementary and high school years, students in our cohort were much less likely to be chronically absent or worse compared to all students who were enrolled in PPSD for at least 1 day (27.8% versus 35.1% for kindergarten, 19.5% versus 25.8% for third grade, and 35.3% versus 48.8% for ninth grade). However, students in the sixth grade cohort were slightly more likely to be chronically absent than their peers (26.3% versus 28.2%).

Figure 7. Breakdown of SY 2016-17 student attendance by cohort and program.

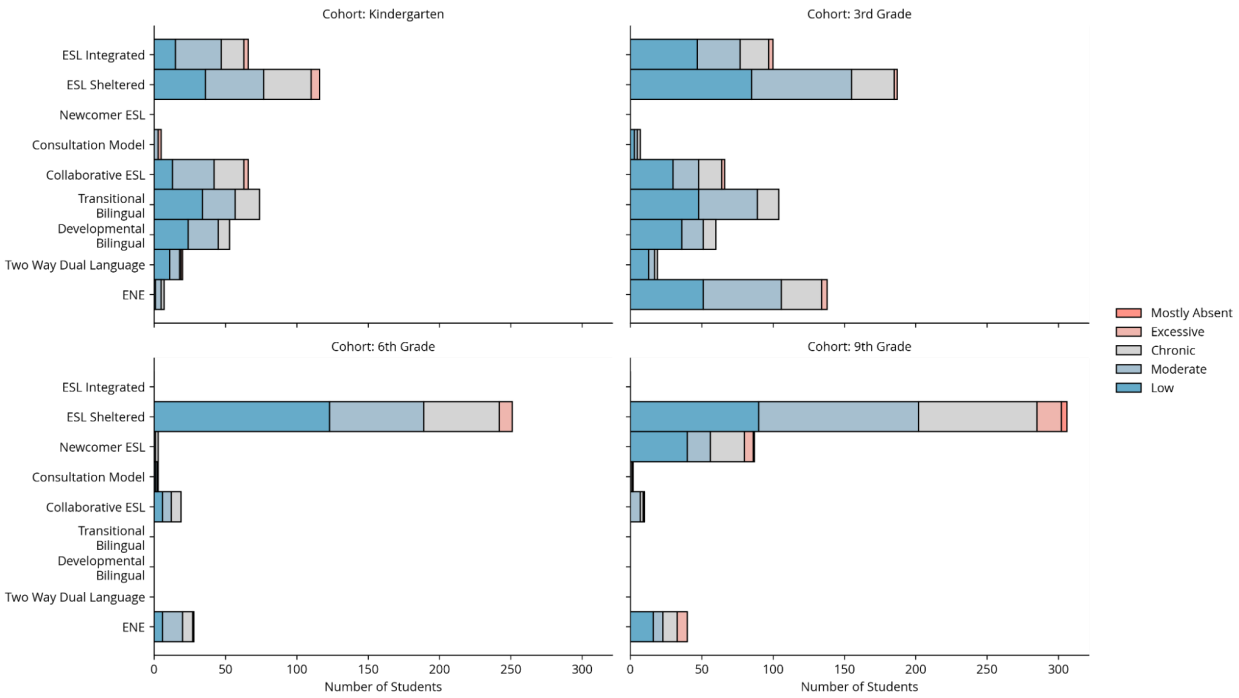


Table 20. Breakdown of SY 2016-17 student attendance by cohort and program.

		Mostly Absent	Excessive	Chronic	Moderate	Low	Total
Kindergarten	ESL Integrated	0 (0.0%)	*	16 (24.2%)	32 (48.5%)	15 (22.7%)	66 (100.0%)
	ESL Sheltered	0 (0.0%)	*	33 (28.4%)	41 (35.3%)	36 (31.0%)	116 (100.0%)
	Consultation Model	0 (0.0%)	*	0 (0.0%)	*	0 (0.0%)	*
	Collaborative ESL	0 (0.0%)	*	21 (31.8%)	29 (43.9%)	13 (19.7%)	66 (100.0%)
	Transitional Bilingual	0 (0.0%)	0 (0.0%)	17 (23.0%)	23 (31.1%)	34 (45.9%)	74 (100.0%)
	Developmental Bilingual	0 (0.0%)	0 (0.0%)	*	*	24 (45.3%)	53 (100.0%)
	Two Way Dual Language	0 (0.0%)	*	*	*	11 (55.0%)	20 (100.0%)
	ENE	0 (0.0%)	0 (0.0%)	*	*	*	*
	Not Cohort	*	*	21 (22.1%)	25 (26.3%)	29 (30.5%)	95 (100.0%)
3rd Grade	ESL Integrated	0 (0.0%)	*	*	30 (30.0%)	47 (47.0%)	100 (100.0%)
	ESL Sheltered	0 (0.0%)	*	*	70 (37.4%)	85 (45.5%)	187 (100.0%)
	Consultation Model	0 (0.0%)	0 (0.0%)	*	*	*	*
	Collaborative ESL	0 (0.0%)	*	*	18 (27.3%)	30 (45.5%)	66 (100.0%)
	Transitional Bilingual	0 (0.0%)	0 (0.0%)	15 (14.4%)	41 (39.4%)	48 (46.2%)	104 (100.0%)
	Developmental Bilingual	0 (0.0%)	0 (0.0%)	*	*	36 (60.0%)	60 (100.0%)
	Two Way Dual Language	0 (0.0%)	0 (0.0%)	*	*	13 (68.4%)	19 (100.0%)
	ENE	0 (0.0%)	*	*	55 (39.9%)	51 (37.0%)	138 (100.0%)
	Not Cohort	*	*	24 (23.3%)	26 (25.2%)	37 (35.9%)	103 (100.0%)
6th Grade	ESL Sheltered	0 (0.0%)	*	*	66 (26.3%)	123 (49.0%)	251 (100.0%)
	Newcomer ESL	0 (0.0%)	0 (0.0%)	*	*	0 (0.0%)	*
	Consultation Model	0 (0.0%)	0 (0.0%)	*	*	*	*
	Collaborative ESL	0 (0.0%)	0 (0.0%)	*	*	*	19 (100.0%)
	ENE	0 (0.0%)	*	*	14 (50.0%)	*	28 (100.0%)
	Not Cohort	0 (0.0%)	*	*	20 (33.3%)	27 (45.0%)	60 (100.0%)
9th Grade	ESL Sheltered	*	*	83 (27.1%)	112 (36.6%)	90 (29.4%)	306 (100.0%)
	Newcomer ESL	*	*	24 (27.6%)	16 (18.4%)	40 (46.0%)	87 (100.0%)
	Consultation Model	*	0 (0.0%)	*	0 (0.0%)	0 (0.0%)	*
	Collaborative ESL	0 (0.0%)	*	*	*	0 (0.0%)	10 (100.0%)
	ENE	0 (0.0%)	*	10 (25.0%)	*	16 (40.0%)	40 (100.0%)
	Not Cohort	77 (22.5%)	103 (30.1%)	86 (25.1%)	33 (9.6%)	43 (12.6%)	342 (100.0%)

Table 21. Breakdown of SY 2016-17 student attendance in PPSD by grade.

		Mostly Absent	Excessive	Chronic	Moderate	Low	Total
Kindergarten	**Cohort**	0 (0.0%)	15 (3.7%)	98 (24.1%)	160 (39.3%)	134 (32.9%)	407 (100.0%)
	All Students	13 (0.72%)	127 (7.03%)	494 (27.35%)	626 (34.66%)	546 (30.23%)	1806 (100.00%)
3rd Grade	**Cohort**	0 (0.0%)	11 (1.6%)	122 (17.9%)	235 (34.5%)	313 (46.0%)	681 (100.0%)
	All Students	*	*	449 (20.58%)	701 (32.13%)	919 (42.12%)	2182 (100.00%)
6th Grade	**Cohort**	0 (0.0%)	10 (3.3%)	70 (23.0%)	88 (28.9%)	136 (44.7%)	304 (100.0%)
	All Students	16 (0.86%)	111 (5.97%)	397 (21.34%)	558 (30.00%)	778 (41.83%)	1860 (100.00%)
9th Grade	**Cohort**	*	*	120 (27.0%)	142 (31.9%)	146 (32.8%)	445 (100.0%)
	All Students	188 (8.34%)	365 (16.19%)	547 (24.27%)	581 (25.78%)	573 (25.42%)	2254 (100.00%)

4.3.2 Enrollment in SPED Programs

The DOJ settlement listed enrollment in Special Education as an outcome. To avoid duplication, we discuss SPED status in Section 2.1.2 Student Demographics.

4.3.3 Enrollment in Enrichment Programs

PPSD offers the following academic enrichment programs: Advanced Academics, Honors courses, and Advanced Placement (AP) courses. The Gifted Program is not available at PPSD. In the following sections, we examine as outcomes the enrollment of ninth grade cohort students in Honors and AP courses. We do not examine enrollment in the Advanced Academics program due to inconsistencies with the process for identifying qualifying students during our study period.

Honors Classes

Across the entire MLL population in PPSD, less than ten students were enrolled in at least one Honors class[39] during the period of our study; all of these students are included in our cohort.

[39] Honors English 2 and Spanish 3

Notably, almost every one of these students was enrolled at Providence Career and Technical High School; the remaining students were ENE students and therefore not actively receiving MLL services.

Advanced Placement Classes

A total of 43 unique high school MLL students enrolled in at least one AP class during the period of our study, almost all of whom were included in our ninth grade cohort. Most of these students had enrolled in one AP course at some point between SY 2016-17 and SY 2018-19 (Table 22), with a handful of students having enrolled in two. Most of the students in our cohort who had enrolled in at least one AP course participated in the ESL Sheltered program (Table 23). No students from either the Consultation or Collaborative ESL programs enrolled in an AP course at any point during our study period, while only one Newcomer ESL student enrolled in an AP course during the same time period.

In comparison, out of 1797 non-MLL students with three years attendance in PPSD, 505 students (28.1%) enrolled in at least one AP course between SY 2016-17 and SY 2018-19.

Table 22. Numbers of students enrolled in AP classes, by number of AP classes (0-6) and cohort status.

	Cohort	Not Cohort	Total
0	404 (90.8%)	25 (92.6%)	429 (90.9%)
1	*	*	*
2	*	*	10 (2.1%)
3	*	0 (0.0%)	*
4	*	0 (0.0%)	*
6	*	0 (0.0%)	*
Total	445 (100.0%)	27 (100.0%)	472 (100.0%)

Table 23. Number of in-cohort students enrolled in AP classes, by program.

	Collaborative ESL	Consultation Model	ENE	ESL Sheltered	Newcomer ESL	Total
Enrolled in AP	0 (0.0%)	0 (0.0%)	*	36 (11.8%)	*	41 (9.2%)
Not Enrolled in AP	*	*	36 (90.0%)	270 (88.2%)	86 (98.9%)	404 (90.8%)

4.4.1 Grade Retention Rates (i.e. students held back a grade)

No students in our cohort repeated a grade between SY 2016-17 to SY 2018-19.[40]

4.4.2 Dropout and Graduation Rates

We examined student four-year dropout and graduation data using data from SY 2019-20 to provide a more complete sense of a student's high school experience, since students in the ninth grade cohort enter 12th grade at the start of SY 2019-20.

Four-year Dropout Rates

According to PPSD, as of SY 2019-20, the overall four-year dropout rate across all MLL programs for the ninth grade cohort was 19.9% (n=157). This rate varied across programs, with the Newcomer program having the highest dropout rate at 19.5%, compared with a 7.5% dropout rate for students in the Sheltered ESL program (Table 24).

Four-year Graduation Rates

According to PPSD, the four-year graduation rate across all MLL programs as of SY 2019-20 for the ninth grade cohort was 40.0% (n=315). Note that the four-year graduation rate for students who matriculated to ninth grade in PPSD during SY 2014-15, two years before our study began, was 74.9% (Rhode Island Department of Education, n.d.). As with dropout rates, graduation rates also varied across programs (Table 24), with the Sheltered ESL program seeing the highest four-year graduation rate (74.2%) of all MLL programs, followed by the Newcomer ESL program at 47.1%. Notably, graduation rates were highest for a fairly small number of ENE students (85%, n=34).

[40] Our data begins in SY 2016-17. Therefore, a student who was repeating e.g. ninth grade during SY 2016-17 would not appear in this statistic. Only grades repeated in SY 2017-18 and SY 2018-19 would be detected with this methodology.

Table 24. Four-year dropout and graduation rates for students in the ninth grade cohort, by program.

	Graduated in 4th Year	Graduated in 5th Year	Graduated in 6th Year	Completed GED/Other	Retained/ Still In School	Dropped Out	Reached Maximum Age	Exited - Unknown	Transferred Out	Not Tracked	Total
ESL Sheltered	227 (74.2%)	*	0 (0.0%)	0 (0.0%)	19 (6.2%)	23 (7.5%)	*	*	22 (7.2%)	0 (0.0%)	306 (100.0%)
Newcomer ESL	41 (47.1%)	*	0 (0.0%)	0 (0.0%)	10 (11.5%)	17 (19.5%)	*	*	*	0 (0.0%)	87 (100.0%)
Consultation Model	*	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	*	0 (0.0%)	0 (0.0%)	*
Collaborative ESL	*	0 (0.0%)	0 (0.0%)	0 (0.0%)	*	*	0 (0.0%)	*	0 (0.0%)	0 (0.0%)	10 (100.0%)
ENE	34 (85.0%)	*	0 (0.0%)	0 (0.0%)	*	0 (0.0%)	0 (0.0%)	*	0 (0.0%)	0 (0.0%)	40 (100.0%)
Not Cohort	*	*	*	*	21 (6.1%)	116 (33.9%)	17 (5.0%)	34 (9.9%)	137 (40.1%)	*	342 (100.0%)

5. Research Questions, Literature Review, and Quasi-experimental Analysis

In our pre-analysis plan, we laid out four general research questions that we set out to study for the purpose of this report.[41]

1. Do students in certain MLL programs have better language acquisition and educational outcomes (standardized tests, progress monitoring tests) than students in other MLL programs?
2. Do enrolled MLL students have different educational outcomes compared to eligible but never enrolled (ENE) students? Do these differences in educational outcomes vary across programs?
3. Do certain subpopulations of MLL students (e.g. students with certain home language types, students who enroll in an MLL program at a younger age, students with certain classifications, etc.) have better educational outcomes than other subpopulations? Do these differences in educational outcomes vary across programs?
4. Do enrolled MLL student outcomes vary by teacher certification status and certification type?

This section differs from the previous section in that we attempt to determine if a particular program *causes* better or worse outcomes, whereas the previous section attempted to answer whether the students in a particular program happened to have better or worse outcomes. Findings from this section attempt to answer the question of whether PPSD should, for instance, move to expand Two Way Dual Language programming or whether it should revisit the structure of the ESL Integrated program.

To do this, we describe the specific hypotheses of interest under each of these research questions, provide a brief review of the literature on the topic, and perform the analysis described in Section 2.2.4 Quasi-experimental Analyses to attempt to tease out whether differences observed in the previous section are due to, for example, selection effects or to the programs themselves.

Due to data limitations, we were unable to conduct the analyses for Research Questions 3 and 4 as intended and stated in the pre-analysis plan. For the literature review and additional details, please see Appendix J.

[41] Our pre-analysis plan also considers two further questions related to subpopulations of MLL students and the effects of teacher certification. These are not relevant for the purposes of responding to the DOJ's request, and so for brevity we only include the analyses for the first two questions here.

5.1 Research Question 1

Do students in certain MLL programs have better language acquisition and educational outcomes (standardized tests, progress monitoring tests) than students in other MLL programs?

Under the heading of this research question, we set out to explore the following questions:

- Do students in elementary Bilingual and Dual Language programs progress differentially in math, reading/ELA, and language acquisition than those in other programs?
- Do students in elementary schools with higher percentages of non-MLL students in the classroom (i.e. Integrated ESL and Collaborative) have higher English language acquisition assessment scores than students in programs with a lower composition of non-MLLs (i.e. Newcomer ESL and ESL Sheltered)?
- Do we observe differences in outcomes among secondary school students enrolled in different programs?

5.1.1 Literature review

Elementary Bilingual and Dual Language Learners vs. Other Elementary MLL Programs

We examined three commonly-used instructional programs for MLL students: English-Immersion, Bilingual, and Dual-Immersion. Though there is substantial variation across localities in how schools implement and even label these programs, we define these programs according to PPSD's definitions. English-Immersion programs provide instruction to MLL students primarily in the English language. Bilingual programs provide instruction to MLL students in both English and their native language. Instruction often starts out primarily in the native language, which is phased out over the course of several years as English is phased in. Dual-Immersion programs provide instruction to both MLL and non-MLL students in the same classroom. 50% of instruction is provided in MLL students' native language and 50% of instruction is provided in English.

Generally speaking, the literature shows that while Dual-Immersion programs fall behind Bilingual programs in short-term achievement scores (Anderberg & Ruby, 2013; Valentino & Reardon, 2015), in the long-term students in Dual-Immersion programs achieve better or comparable results in ELA while results vary in math (López & Tashakkori, 2006; Marian et al., 2013; Valentino & Reardon, 2015).

Secondary (Middle and High) MLL Programs

At the secondary level, the District offers three MLL programs to students – Newcomer ESL, Sheltered ESL, and English Learner Collaborative (ELC) – which differ by the incoming English proficiency of the students. Much of the literature focusing on secondary MLL education offered little comparison in outcomes between programs; further, there tended to be reduced program offerings at the middle and high school level as compared to at the elementary school level. Given that these secondary programs rely on proficiency to differentiate programming, we hypothesized that students in programs requiring higher language thresholds would score higher than those students in programs with lower language thresholds across subject assessment, language acquisition, and progress monitoring.

Program Classroom Composition

We find mixed results in the literature regarding the impact of classroom composition. While some studies show that a higher proportion of MLL students has positive effects on MLL outcomes (Bui, 2014), others show no effect (Ahn & Jepsen, 2015; Bui, 2014) or negative effects (Ahn & Jepsen, 2015; Estrada et al., 2020). There does seem to be more support, however, for the notion that a higher proportion of MLL students in a classroom worsens MLL outcomes. This is likely because concentrating MLLs into classrooms separate from non-MLLs leads to teacher burnout, lowers teacher expectations, and prevents MLLs from engaging with non-MLL peers (Ahn & Jepsen, 2015). Supporters of segregated classrooms, meanwhile, argue that teachers in classrooms with predominantly MLL students can deliver a more focused instruction (Bui, 2014).

5.1.2 Our findings

In accordance with the matching and regression procedure described in Section 2.2.4 Quasi-experimental Analyses, we looked at outcomes for students compared to similarly situated students across eight of PPSD's nine programs. (The final program, ENE, is covered by Research Question 2.) Specifically, we look at the outcome variables described in Table 4.[42] Detailed regression outputs may be found in Appendix H. Here, we summarize the main takeaways.

[42] In our pre-analysis plan, we noted that we would attempt to use STAR exams as baseline scores for some outcomes as well as performance on STAR exams in SY 2018-19 as separate outcomes. However, as discussed in Section 2.2.4 Quasi-experimental Analyses, STAR coverage amongst cohort students was not robust, and we therefore decided to focus instead on state testing outcomes.

Matching Failures

Our analysis can only work when a sufficient number of students can be matched to similarly situated students. This was not always possible. In particular, our matching required that students' baseline test scores and ACCESS scores be within ¼ standard deviation of each other. In some cases, this made finding comparable students impossible. In Appendix H, we document for which target outcomes our matching failed by marking the row with all dashes (-). This lack of matching may indicate that another method, e.g. a regression discontinuity design, may be more appropriate to study some cohorts or programs.

Cohort 1: Kindergarten students in SY 2016-17

In our analyses, we found only one statistically significant finding in our youngest cohort. We note that many of our matchings fail and so we do not report any effect estimates for these comparisons. This is unsurprising, though, as our descriptive analyses showed wide variation especially in STAR scores among students in this cohort by program type.

Two Way Dual Language

Under our distance criteria, we could match all 20 students in our cohort who were enrolled in the Two Way Dual Language program to other students, even controlling for their STAR Early Literacy scores. These students seemed to show significantly better scores on their second grade STAR reading exams than their matched peers. On their second grade EOY STAR Reading scores, students scored 191.30 points higher (standard error 31.43, p-value < 0.0001) than their matched peers.

In terms of robustness, the balance tests do not find much difference in the matched groups, including extremely similar baseline STAR Early Literacy scores, and the overall χ^2 test does not indicate significant differences (p-value 0.69). Moreover, the Wilcoxon-Mann-Whitney rank test also indicates that the finding is robust ($Z = 3.55$, p-value < 0.0001).

This result is extraordinarily large, potentially unbelievably so. Such statistically significant gains were not observed in the ACCESS exam scores (effect 9.15, standard error 5.78, p-value 0.13). We caution that given the small sample size and the limited availability of Two Way Dual Language programming for kindergartners at only two schools at PPSD, these results should be taken with a grain of salt. Indeed, from our descriptive analyses, we know that several programs have clumped scores on the STAR Reading exam near the minimum score, which indicates potential issues in administration, and Two Way Dual Language program itself has a long rightward tail of scores, indicating potentially localized effects.

Still, the size and robustness of this result merits further exploration from PPSD, such as by potentially expanding access to Two Way Dual Language instruction through additional seats or relaxed entrance criteria. If such an endeavor is undertaken, we recommend that PPSD strategically design and closely monitor the rollout in order to observe its effects relative to other programs.

Cohort 2: Third grade students in SY 2016-17

Among all the regressions we performed for third grade students, one group of students stood out: those who were enrolled in the Developmental Bilingual program.

Developmental Bilingual

Under our distance criteria, we could match 41 (for the PARCC/RICAS math assessments) or 42 (for PARCC/RICAS ELA and ACCESS assessments) of 42 total students who were enrolled in the Developmental Bilingual program. These students seemed to show significantly better improvement on standardized math and ELA tests than their matched peers. In ELA, students showed 8.67 points higher (standard error 2.41, p-value < 0.001) scores on the 2019 RICAS exam than their matched peers, whereas in math, students showed 6.22 points (standard error 2.64, p-value 0.02) higher scores. On the ACCESS, students also showed higher gains of 7.69 points with a standard error of 3.50 and a p-value of 0.03. Under the Bonferroni correction, only the ELA p-value is significant; however, the fact that all three exams have low p-values is quite interesting.

In terms of robustness, the balance tests do not find much difference in the matched groups (except potentially for the math result). Additionally, the Wilcoxon test yields p-values that are consistently on the order of the main p-values.

Notably, this collection of students consisted almost entirely of Hispanic/Latino students, and almost all students explicitly described their home language as Spanish. This makes sense as the Developmental Bilingual program at PPSD is explicitly taught in English and Spanish. Overall, these results indicate that PPSD should explore why these 42 students are seeing such significant gains, potentially including experimenting with expanding Developmental Bilingual programs where possible.

Cohort 3: Sixth grade students in SY 2016-17

Among all the regressions we performed for sixth grade students, no groups of students showed any significant effects, although one small group of students, those in the Collaborative ESL

program, came close. However, we caution up front that this effect is on a very small group of students and that the attendant *p*-values are relatively large given the number of regressions we have performed.

Collaborative ESL

Under our distance criteria, we could match 0 (for the PARCC/RICAS math and ELA assessments) or 16 (for the ACCESS assessment) of 16 total students who were enrolled in Collaborative ESL. These students seemed to show *almost* significantly less improvement on their ACCESS tests than their matched peers. On these tests, students showed 15.94 scale score points lower (standard error 6.82, *p*-value 0.03) on the SY 2018-19 ACCESS assessment than their matched peers. This *p*-value would be significant if not for our Bonferroni correction.

In terms of robustness, the balance tests find some difference in the matched groups, notably larger proportions of SPED students among those students enrolled in the Collaborative ESL program. Overall, Hansen and Bowers's χ^2 test indicates that these imbalances aren't particularly significant overall (*p*-value 0.19)

Notably, the Wilcoxon-Mann-Whitney test does not disagree with the potential significance of the finding for the ACCESS exam, giving it a *p*-value of 0.04.

Cohort 4: Ninth grade students in SY 2016-17

We found no significant results among this cohort.

5.2 Research Question 2

Do enrolled MLL students have different educational outcomes compared to eligible but never enrolled (ENE) students? Do these differences in educational outcomes vary across programs?

Under the heading of this research question, we set out to explore the following question:

- Do ENE students progress differentially in math, reading/ELA, and language acquisition than those in other programs?

5.2.1 Literature review

There was little in the way of research on the outcomes of those students who were eligible for MLL programming but elected not to enroll (i.e. ENE students). Thus, we chose to test the hypothesis that, among those students eligible for MLL services, receiving MLL instruction would produce better educational outcomes.

5.2.2 Our findings

In accordance with the matching and regression procedure described in Section 2.2.4 Quasi-experimental Analyses, we looked at outcomes for ENE students compared to similarly situated students across eight of PPSD's nine programs. Specifically, we look at the outcome variables described in Table 4.[43] Detailed regression outputs may be found in Appendix H. Here we summarize the main takeaways.

Matching Failures

Our analysis can only work when a sufficient number of students can be matched to similarly situated students. This was not always possible. In particular, our matching required that students' baseline test scores and ACCESS scores be within ¼ standard deviation of each other. In some cases, this made finding comparable students impossible. In Appendix H, we document for which target outcomes our matching failed by marking the row with all dashes (-). However, in the case of ENE students, we note that only two of our 11 attempts at matching failed: attempting to match for third-grade PARCC/RICAS math scores and ninth-grade PARCC/RICAS ELA scores. Given the size of these ENE cohorts, this result is somewhat surprising, and again points to perhaps these groups being amenable to study via a different mechanism than the matching design we have prescribed in this study.

Cohort 1: Kindergarten students in SY 2016-17

No significant differences were found between ENE students in the kindergarten cohort and their matched peers.

[43] In our pre-analysis plan, we noted that we would attempt to use STAR exams as baseline scores for some outcomes as well as performance on STAR exams in SY 2018-19 as separate outcomes. However, as discussed in the Quasi-experimental Analyses section, STAR coverage amongst cohort students was not robust, and so we have decided to focus instead on state testing outcomes.

Cohort 2: Third grade students in SY 2016-17

No significant differences were found between ENE students in the third grade cohort and their matched peers.

Cohort 3: Sixth grade students in SY 2016-17

Under our distance criteria, we could match 18 (for PARCC/RICAS math assessment), 17 (for the PARCC/RICAS ELA assessment), and 22 (for the ACCESS test) of the 22 total ENE students. These students seemed to show significantly less improvement on their RICAS ELA tests in SY 2018-2019 than their matched peers. On these tests, students showed 9.65 points lower scores (standard error 3.84, *p*-value 0.02) on the 2019 RICAS exam than their matched peers. While the point estimates from ACCESS tests were also negative (-9.50), the standard errors were quite large (8.93), resulting in a high *p*-value (0.30).

In terms of robustness, the balance tests found no significant differences in the matched groups, with Hansen and Bowers's overall χ^2 test yielding *p*-values of 0.16, 0.25, and 0.30 for the RICAS math, RICAS ELA, and ACCESS regressions, respectively. Moreover, the Wilcoxon-Mann-Whitney test agrees with the significance of the finding for the ELA exam, giving it a *p*-value of 0.04.

Importantly, this is a small number of students whose parents opted out of receiving MLL services. We did not find similar sizes of effects in ENE students in other cohorts, so this particular result may be idiosyncratic.

Cohort 4: Ninth grade students in SY 2016-17

No significant differences were found between ENE students in the ninth grade cohort and their matched peers.

6. Limitations and Implications

In addition to the three caveats mentioned in 2.2.4 Quasi-experimental Analyses, we note the following limitations to this report:

Timeline. PPSD signed the DOJ settlement in Aug 2018. According to PPSD, because of how close the timeline was to the start of the school year, students' schedules were already set in place. Therefore, the district was not able to implement many programmatic, staffing or scheduling changes until the following school year, SY 2019-20. As such, much of this report should ultimately be read as providing a relatively comprehensive overview of MLL education programming in PPSD *prior* to the institution of the DOJ settlement, thus affording both the DOJ and PPSD the opportunity to conduct longitudinal or year-over-year studies with this report as a baseline and/or template.

Limited Timeframe. Almost all of the data considered in this study came from the SY 2016-17 through SY 2018-19 time period. As such, we lack some facts about students who either entered PPSD *before* that time period or remained in PPSD *after* that time period. For instance, if a student in our sixth grade cohort entered PPSD in SY 2014-15, we do not have access to their initial enrollment information such as their initial WIDA screener scores. This means that we cannot directly answer questions such as how student outcomes differ by initial English language ability.

Unique Student Identifier. We were provided with two unique identifiers for each student: RIDE's State Assigned Student Identifier (SASID) and PPSD's Locally Assigned Student Identifier (LASID). Throughout the course of our analysis, we identified multiple cases in which SASIDs and LASIDs were mismatched, such as SASIDs that matched with multiple LASIDs, and vice versa. To our knowledge, these mismatched identifiers did not ultimately impact the students in our cohort; however, these issues do raise the possibility that there may be incorrectly identified students in our cohort. Such problems are common in data analyses and are extremely difficult to correct retrospectively as doing so requires significant staff time to hand review records from multiple agencies. Prospectively, this problem is likely to persist as long as there are two "sources of truth" for student identity. We would suggest RIDE, who has a broader view of student movement within the state, handle identity management moving forward to ensure there is one source for student identification.

Student Types. To identify students' MLL program enrollment, we were provided with the Student Types field in our dataset. Over the course of our data cleaning process, we identified multiple instances in which the Student Type was incorrectly labelled. These included instances where schools or grades were omitted from the classification; once, students in a school were incorrectly labelled as being enrolled in an MLL program type when no such program was offered in that school. We worked extensively with PPSD staff to trace and fix all identified data discrepancies; however, because our data starts from five years ago and because of changes to

PPSD MLL program staff over the last five years, we cannot guarantee that all errors were caught and accurately corrected.

Eligible Never Enrolled. In certain instances, students might appear as ENE students on paper, but be receiving MLL services in practice. One such instance is with the Two Way Dual Language program. Students in this program are identified as "Spanish" or "English" as part of their program code, with MLL students identified as "Spanish", since the program is designed such that half the class consists of MLL students and half of General Education students.

Teacher Demographic Data. As mentioned above, we did not receive sufficient data on teacher years of experience and demographics in time for this report, and so were unable to control for these factors in the analyses. However, a teacher's experience, efficacy, and demographics, and not mere program design, is a major contributor to a student's learning and growth. The analysis is therefore incomplete as we were unable to account for these factors.

School Level Effects. Not all schools offer all MLL programs, and indeed, some MLL programs like Two Way Dual Language are offered at only one or two schools. Thus, it is difficult to disentangle *school* effects versus *program* effects. In particular, our quasi-experimental analysis suffers from being unable to match students to similar students *within* their schools. As such, we recommend reading all the results presented with this caveat in mind.

ALT ACCESS Scores. PPSD provided 19,763 ACCESS scores and 203 ALT ACCESS scores for SY 2016-17 through SY 2018-19. These ALT ACCESS exams were taken by 118 different students. Of these students, 29 were in kindergarten, third grade, sixth grade, or ninth grade in SY 2016-17, all of whom are in our cohort (i.e. remained enrolled in PPSD for all three years of the study). Almost all of these 29 students were enrolled in special education all three years and the remaining were enrolled in special education for two of the three years.

Between SY 2016-17 and SY 2018-19, 306 MLL students were ever enrolled in special education classes, of which 194 were in our cohort. By excluding from our quasi-experimental analyses the ALT ACCESS scores of the 27 SPED students in our cohort whose ALT ACCESS scores met the "complete cases" definition, we exclude this data point for 12.6% of SPED students in our cohort, or 1.5% of students in our cohort overall.

Complete Cases. By looking at students who have a set of test scores that satisfy the "complete cases" criteria, it is possible that we examined a unique subset of students with certain observable or unobservable characteristics. Put differently, it is possible that certain groups of students, such as those with low attendance or those whose living situations are more transient, might be underrepresented in or excluded from the analysis. Readers should examine the results with this caveat in mind.

7. Key Takeaways

7.1 Data Storage and Availability

7.1.1 This analysis and its conclusions were limited by the availability and cleanliness of student and educator data.

Compiling this data was a time-consuming process largely because there was no centralized source of data. For instance, MLL census data and information about teacher certification were stored with RIDE, while most HR data was stored within a system used by the city; these data were not easily accessible by PPSD's research specialist except upon request. **This was particularly troubling when there was no key to link different datasets, or there were inconsistencies with the key that existed** (such as the issues with the unique student identifier [see 6. Limitations and Implications] linking students across the PPSD and RIDE data systems). This resulted in manual cross-referencing and informed inferences, which was both tedious and increased the likelihood of error.

At times, the required data was not available or was not available in a format that could be reasonably manipulated and merged for analysis. As a result, **we were unable to analyze and account for certain key factors, such as teacher demographics, years of experience, and certification status.**

Additionally, there were questions about the accuracy and reliability of certain data, either due to incorrect data entry or inconsistent data entry practices. There were several reasons for this, such as staff turnover resulting in inconsistent practices or practices that were not comprehensively documented. In some cases, this may have been due to vendor systems' interfaces not providing a sufficiently usable interface for data entry.

There also appeared to be a lack of clear documentation explaining how changes in policies might affect the interpretation of data. Staff who were at PPSD when those policy changes were made were adept at explaining its effect on the data, but a lack of centralized documentation does put PPSD at risk when people leave and institutional knowledge is lost, as we encountered through this process.

Finally, the availability of test data varied by student cohort. This issue was particularly problematic for the ninth grade cohort. For instance, SY 2016-17 and SY 2018-19 ACCESS scores were mostly available for the kindergarten cohort (89.9%), decreasing to 64.2% and 69.7% for the third and sixth grade cohorts, respectively. Within the ninth grade cohort, the proportion of students with a complete set of ACCESS scores was less than half that of the kindergarten cohort, at 39.1%. Based on conversations with PPSD, a range of factors may be causing the low

rates of ACCESS test completion for older students, including logistical testing challenges and high rates of absenteeism. Relatedly, we had to omit STAR test scores for the ninth grade cohort from analysis due to low levels of test data availability.

7.2 Program Enrollment

7.2.1 In practice, the intended grouping of students into MLL programs based on English language proficiency levels did not appear to be occurring with consistency during the years of this study.

Based on MLL program entry requirements, it appears that a goal of the MLL program design is to group students with similar English proficiency levels together, and to design scaffolds and interventions that meet students where they are linguistically. However, based on the wide distribution of ACCESS scores from SY 2018-19 (Figure 3), this did not appear to occur in practice during the years of this study. We do note, however, that according to PPSD, there has been a push since SY 2018-19 to place students into programs according to their English language proficiency level (see Figure 1 for examples of student movement across programs). As such, the current numbers might be very different from what is in this report, and could benefit from updated analyses.

7.3 Academic Achievement

7.3.1 On average, elementary school students made greater progress academically and in English language proficiency compared to secondary level students, where progress was, on average, minimal.

ACCESS. On average, all students made progress on English language proficiency (as measured by the ACCESS test). Kindergarten students, in particular, showed a lot of growth: a 75.1% percentage change in average ACCESS scores, the largest across all cohorts. However, growth was minimal at the secondary level, with both the sixth and ninth grade cohorts each seeing a percentage change in mean ACCESS scores of less than 0.5%.

STAR. On average, we observe gains on the STAR assessment for both the third and sixth grade cohorts, although the gains for the sixth grade cohort was about two-thirds the increase in average Reading test scores and half the increase in average Math test scores that we saw with the third grade cohort. However, students in the sixth grade cohort had starting average test scores that were roughly 100 points higher than that of the third grade cohort in each subject,

and it is possible that the higher average test scores in SY 2016-17 had an impact on the increase in average test scores.

PARCC/RICAS. Comparing SY 2017-18 and SY 2018-19 RICAS ELA and Math scores, we observe overall minimal movement or change, regardless of age or MLL program type.

7.3.2 Student academic performance varied across MLL programs, with students in the Two Way Dual Language and Bilingual programs at the elementary level outperforming the other programs.

Within the kindergarten cohort, students in the Developmental and Transitional Bilingual programs more than doubled their average ACCESS scores over the course of the study, while students in the Two Way Dual Language program increased their average ACCESS scores by 81.3%. In the third grade cohort, students in the Developmental Bilingual and ESL Sheltered programs both saw an increase in average ACCESS by over 30%.

We also observe surprisingly large gains on the STAR assessment for kindergarten cohort students enrolled in the Two Way Dual Language program and on the PARCC/RICAS state assessments for third grade cohort students enrolled in the Developmental Bilingual program. Although it is possible that these results are due to selection or school effects, the effects are large enough that we recommend PPSD explore increasing access to these types of programs where possible and to do so in a way that will allow rigorous evaluation of these program effects.

7.3.3 About a third of students in the third grade cohort exited MLL services in third and fourth grade, as anticipated. However, exit rates for secondary students were low overall.

By the end of SY 2018-19, 286 out of 1837 students (11.2%) in our overall cohort exited MLL services. Among our cohorts, exit rates were highest for the third grade cohort (30.1%, n=205) and lowest for the ninth grade cohort (6.1%, n=27).

While it is true that MLL program exit rates were generally low across cohorts, the story is more nuanced for the third grade cohort. We found that exit rates were higher within this cohort at the end of SY 2016-17 (when most students would have been in the third grade) and SY 2017-18 (when most students would have been in the fourth grade), with 115 students (16.9% of the cohort) and 82 students (12.0% of the cohort) exiting MLL services, respectively. In total, 197 students (28.9% of the third grade cohort) exited the MLL program between SY 2016-17 and SY

2017-18. This aligns with what we heard from PPSD, which is that when students do exit the MLL program, they tend to do so while in their middle elementary years.

7.4 Student Behavior

7.4.1 Absenteeism rates were objectively high, though low compared to the broader PPSD population.

At least half of the students in each cohort and across most MLL programs had moderate or more severe absence levels. Students in the kindergarten and ninth grade cohorts had the highest absenteeism rates. The ninth grade cohort had the highest proportion of chronic and excessive absences across all cohorts, and was also the only cohort with students that had mostly absent levels of absenteeism; in contrast, absenteeism rates in the kindergarten cohort were mostly driven by moderate levels of absenteeism. Overall, students in the third grade cohort had the lowest proportion of students with chronic and excessive absences.

However, **students in all but the sixth grade cohort were less likely to have chronic or more severe absence levels than PPSD as a whole**.

7.4.2 Enrollment in AP courses was low, as were the number of students who took the SAT. When students did take the SAT, most of them performed poorly.

AP Courses. Over the course of the study period, 41 students enrolled in at least one AP course. Most of the students in our cohort who had enrolled in at least one AP course participated in the ESL Sheltered program. No students from either the Consultation or Collaborative ESL programs enrolled in an AP course at any point during our study period, while almost no Newcomer ESL students enrolled in an AP course during the same time period.

SAT. Slightly more than a third of students in the ninth grade cohort took the SAT, with average scores of less than 370 for both the Reading and Math, regardless of program.

7.4.3 No students in any of the cohorts were held back during the course of the study. However, the four-year graduation rate was low and the four-year dropout rate high, although these rates varied across MLL programs.

Retention rates. While no students in any of the study cohorts repeated a grade during the course of the study, some MLL students who did not remain in PPSD for all three years of the study were retained.

Graduation rates. The overall four-year graduation rate across all MLL programs as of SY 2019-20 for the ninth grade cohort was 40.0% (n=315). This ranged from 74.2% within the Sheltered ESL program to 47.1% within the Newcomer ESL program. Notably, graduation rates were highest for a fairly small number of ENE students (85%, n=34). We further note that we could not study the five- and six-year graduation rates of MLL students in our cohort as the data was unavailable at the time this report was prepared.

Dropout rates. The overall four-year dropout rate across all MLL programs for the ninth grade cohort was 19.9%. This ranged from 19.5% within the Newcomer program to 7.5% within the Sheltered ESL program.

8. Next Steps

Based on the findings of this report, The Policy Lab and PPSD are planning to work together to improve four areas of MLL instruction in the district. In this section we briefly outline these next steps.

8.1 Can we improve attendance, especially at exams, among MLL students?

In this report, we found that there was a relationship between low attendance rates and also not completing standardized assessments. This was true even for the ACCESS exam, which is required to exit the English Language learner program.

Since the time period studied in this report, PPSD has worked to improve exam completion rates among MLL students. However, there is still room for improvement. The Policy Lab and PPSD will explore ways to further increase attendance and exam completion rates.

8.2 Can we build an alternate socioeconomic status indicator?

Since the time period studied by this report, PPSD has become a community eligibility district, meaning that every student in the district receives free lunch regardless of their family income. This means that the Free and Reduced Price Lunch (FRPL) indicator we relied on to indicate a a student's socioeconomic status is rapidly losing its utility as a proxy.

The Policy Lab and PPSD will work on creating an alternative metric of socioeconomic status that PPSD can use moving forward to track equitable access to and outcomes of its programs.

8.3 Can we measure the efficacy of the expansion of dual language instruction?

This report found evidence that PPSD's Two Way Dual Language and Developmental Bilingual programs at the elementary level had significant, positive effects on the outcomes of MLL students. Knowing of these effects, PPSD has been working to expand the number of seats in these programs and has recently begun offering similar curricula at the middle school level.

The Policy Lab and PPSD will work together to understand whether it is possible to study the efficacy of these programs, and where feasible, implement strategies to monitor their ongoing effectiveness.

8.4 Can we better understand the effect of teacher certification on student outcomes?

This report originally set out to understand whether or not teachers who became certified in MLL instruction saw better outcomes for their MLL students. However, data issues prevented us from answering this question in the reporting time frame.

While we could not address the question directly using PPSD data, our literature review indicated that teacher certification may improve student outcomes for new teachers, but there is mixed evidence of its efficacy for experienced teachers. The Policy Lab and PPSD remain interested in this issue and will explore data quality improvements as well as other sources of historical data that might shed light on this question.

Bibliography

Ahn, T., & Jepsen, C. (2015). The effect of sharing a mother tongue with peers: evidence from North Carolina middle schools. *IZA Journal of Migration*, *4*(1), 1-21.

Anderberg, A., & Ruby, M. M. (2013). Preschool bilingual learners' receptive vocabulary development in school readiness programs. *NABE Journal of Research and Practice*, *4*(1), 34-59.

Blair, G., Cooper, J., Coppock, A., Humphreys, M., and Sonnet, L. (n.d.). estimatr: Fast Estimators for Design-Based Inference. Part of the DeclareDesign package. Available at https://github.com/DeclareDesign/estimatr/

Borg, L. (2020, June 5). Providence high school to close, 30 jobs eliminated as new superintendent reorganizes. *The Providence Journal*. https://www.providencejournal.com/story/news/education/2020/06/05/providence-high-school-to-close-30-jobs-eliminated-as-new-superintendent-reorganizes/42400919/

Bui, S. A. (2014). How Do Limited English Proficient Students Affect Each Other's Educational Outcomes? Evidence from Student Panel Data. Cornell University, Working Paper. https://docplayer.net/27445831-How-do-limited-english-proficient-students-affect-each-other-s-educational-outcomes-evidence-from-student-panel-data.html

Chetty, R., J. N. Friedman, and J. E. Rockoff (2014). Measuring the impacts of teachers ii: Teacher value-added and student outcomes in adulthood. The American Economic Review *104*(9), 2633–2679

Collier, V. P. (1987). Age and rate of acquisition of second language for academic purposes. *TESOL Quarterly*, *21*(4), 617-641.

Conger, D. (2009). Testing, time limits, and English learners: Does age of school entry affect how quickly students can learn English? *Social Science Research*, *38*(2), 383-396.

Connecticut State Department of Education (2015). Data suppression guidelines. https://edsight.ct.gov/relatedreports/BDCRE%20Data%20Suppression%20Rules.pdf

De Brey, C., et al. (2019). *Status and Trends in the Education of Racial and Ethnic Groups 2018.* National Center for Education Statistics.

Dixon, L. Q., Zhao, J., Shin, J. Y., Wu, S., Su, J. H., Burgess-Brigham, R., Gezer M. U., & Snow, C. (2012). What we know about second language acquisition: A synthesis from four perspectives. *Review of Educational Research*, 82(1), 5-60.

Elder, C., & Davies, A. (1998). Performance on ESL examinations: Is there a language distance effect?. *Language and education*, *12*(1), 1-17.

Estrada, P., Wang, H., & Farkas, T. (2020). Elementary English Learner Classroom Composition and Academic Achievement: The Role of Classroom-Level Segregation, Number of English Proficiency Levels, and Opportunity to Learn. *American Educational Research Journal*, *57*(4), 1791-1836.

Fry, R. (2007). English Language Learners: How far behind in math and reading are English Language Learners? Pew Hispanic Center. https://www.researchgate.net/publication/241381784_How_Far_Behind_in_Math_and_Reading_are_English_Language_Learners

Gómez, G. E. R. (2009). *The role of morphological awareness in bilingual children's first and* second *language vocabulary and reading*. University of Toronto.

Grant, A., Gottardo, A., & Geva, E. (2012). Measures of reading comprehension: do they measure different skills for children learning English as a second language?. *Reading and Writing*, *25*(8), 1899-1928.

Greenberg Motamedi, J. (2015). Time to Reclassification: How Long Does It Take English Learner Students in Washington Road Map Districts to Develop English Proficiency? REL 2015-092. *Regional Educational Laboratory Northwest*.

Hakuta, K., Butler, Y. G., & Witt, D. (2000). How Long Does It Take English Learners to Attain Proficiency? *The University of California Linguistic Minority Research Institute Policy Report 2000- 1.*

Hansen, B. B., & Bowers, J. (2008). Covariate balance in simple, stratified and clustered comparative studies. *Statistical Science*, 219-236.

Hansen, B. B., & Klopfer, S. O. (2006). Optimal full matching and related designs via network flows. *Journal of computational and Graphical Statistics*, *15*(3), 609-627.

Hayes, K., & Salazar, J. J. (2001, March 1). Evaluation of the Structured English Immersion Program. Final Report: Year I. Los Angeles Unified School District. https://files.eric.ed.gov/fulltext/ED462009.pdf

Hickey, P. J., & Lewis, T. (2013). The Common Core, English Learners, and Morphology 101: Unpacking LS. 4 for ELLs. *Language and Literacy Spectrum*, 23, 69-84.

Hothorn, T., Hornik, K., van de Wiel, M. A., and Zeileis, A. (2008). Implementing a Class of Permutation Tests: The coin Package. *Journal of Statistical Software*, *28*(8), 1–23.

Kane, T. J., Rockoff, J. E., & Staiger, D. O. (2006). What does certification tell us about teacher effectiveness? Evidence from New York City. *Economics of Education review*, *27*(6), 615-631.

Kanno, Y., & Varghese, M. M. (2010). Immigrant and refugee ESL students' challenges to accessing four-year college education: From language policy to educational policy. *Journal of Language, Identity & Education, 9(5),* 310-328.

Kentucky Department of Education. (2020, November 25). *Early Warning and Persistence to* Graduation *Data Tools*. https://education.ky.gov/educational/int/Pages/EarlyWarningAndPersistenceToGraduation.aspx

Kini, T., & Podolsky, A. (2016). Does teaching experience increase teacher effectiveness? A review of the research. *Learning Policy Institute*. https://learningpolicyinstitute.org/sites/default/files/productfiles/Teaching_Experience_Report_June_2016.pdf

Krashen, S. D., Long, M. A., & Scarcella, R. C. (1979). Age, rate and eventual attainment in second language acquisition. *TESOL quarterly*, 573-582.

List, M. (2019, Oct. 9) Providence expands free meals to all students. *The Providence Journal*. https://www.providencejournal.com/news/20191009/providence-expands-free-meals-to-all-students

López, F., Scanlan, M., & Gundrum, B. (2013). Preparing teachers of English language learners: Empirical evidence and policy implications. *education policy analysis archives*, *21*, 20.

López, M. G., & Tashakkori, A. (2006). Differential outcomes of two bilingual education programs on English language learners. *Bilingual Research Journal*, *30*(1), 123-145.

MacSwan, J., & Pray, L. (2005). Learning English bilingually: Age of onset of exposure and rate of acquisition among English language learners in a bilingual education program. *Bilingual Research Journal*, 29(3), 653-678.

Marian, V., Shook, A., & Schroeder, S. R. (2013). Bilingual two-way immersion programs benefit academic achievement. *Bilingual Research Journal*, *36*(2), 167-186.

Marinova-Todd, S. H., Siegel, L. S., & Mazabel, S. (2013). The association between morphological awareness and literacy in English language learners from different language backgrounds. *Topics in Language Disorders*, *33*(1), 93-107.

Master, B., Loeb, S., Whitney, C., & Wyckoff, J. (2016). Different skills? Identifying differentially effective teachers of English language learners. *The Elementary School Journal*, 117(2), 261-284.

Massachusetts Department of Elementary and Secondary Education. (2015, July 16). Redefining Low Income - A New Metric for K-12 Education. https://www.doe.mass.edu/infoservices/data/ed.html

McGowan, D. (2015, Jan. 15). RI has no formal policy for opting out of new standardized test. WPRI 12. https://www.wpri.com/news/ri-has-no-formal-policy-for-opting-out-of-new-standardized-test

Morgan, S., & Winship, C. (2014). Counterfactuals and Causal Inference: Methods and Principles for Social Research (2nd ed., Analytical Methods for Social Research). *Cambridge: Cambridge University Press*. doi:10.1017/CBO9781107587991

National Education Association. (2020). English Language Learners: How far behind in math and reading are English Language Learners? https://www.nea.org/resource-library/english-language-learners

Office of English Language Acquisition. (2020). English learners: Demographic trends. https://ncela.ed.gov/files/fast_facts/19-0193_Del4.4_ELDemographicTrends_021220_508.pdf

Renaissance Learning, Inc. (2018). Using Star Early Literacy Scores to Match Books to Early Readers. https://doc.renlearn.com/kmnet/r001146420gb7513.pdf

Rhode Island Department of Education. (n.d.) *School Year 2016-17 Community Eligibility Provision*. Data available at https://www.ride.ri.gov/CNP/NutritionPrograms/NationalSchoolLunchProgram.aspx#21381431-community-eligibility-provision

Rhode Island Department of Education. (n.d.) Rhode Island Department of Education's 2017-18 Graduation Dashboard [Data Dashboard]. https://tableau.ride.ri.gov/t/Public/views/GRADRATES/GraduationStory?:iid=1&:isGuestRedirectFromVizportal=y&:embed=y

Rhode Island KidsCount (2018). *2018 Rhode Island KidsCount Factbook*. https://www.rikidscount.org/Portals/0/Uploads/Documents/Factbook%202018/2018%20Factbook.pdf

Ripley, B., & Venables, W. (2021). Feed-Forward Neural Networks and Multinomial Log-Linear Models. GitHub. https://github.com/cran/nnet

Rosenbaum, P. R. (2010). *Design of Observational Studies*. NY: Springer.

Staiger, D. O., & Rockoff, J. E. (2010). Searching for effective teachers with imperfect information. *Journal of Economic perspectives*, *24*(3), 97-118.

Soto, A.G.R., et al. (2015). Top languages spoken by English Language Learners nationally and by state. Migration Policy Institute. https://www.migrationpolicy.org/research/top-languages-spoken-english-language-learners-nationally-and-state

Tracy, A. M. (2009). Analysis of ESL teacher endorsement effects on English language learners' student achievement and English language acquisition. *University of Kentucky Doctoral Dissertations*. 706. https://uknowledge.uky.edu/gradschool_diss/706

The Lab @ DC. (2019, May 28). *Form-a-Palooza* [Prezi Presentation]. https://formapalooza.dc.gov/

Thompson, K. D. (2017). English learners' time to reclassification: An analysis. *Educational Policy*, *31*(3), 330-363.

Valentino, R. A., & Reardon, S. F. (2015). Effectiveness of four instructional programs designed to serve English learners: Variation by ethnicity and initial English proficiency. *Educational Evaluation and Policy Analysis*, *37*(4), 612-637.

World-Class Instructional Design and Assessment Consortium (WIDA). (2021a, March 1). *Assessing Comprehension and Communication in English State-to-State for English Language Learners, ACCESS for ELLs: Interpretative guide for score reports.* https://wida.wisc.edu/sites/default/files/resource/Interpretive-Guide.pdf

World-Class Instructional Design and Assessment Consortium (WIDA). (2021b). *Alternate ACCESS for ELLs*. https://wida.wisc.edu/assess/alt-access

Zinth, J.D. (2013). English Language Learners: A growing – yet underserved – student population. Education Commission of the States, 14 (6), http://www.ecs.org/clearinghouse/01/10/20/11020.pdf

Appendices

Appendix A. Program Descriptions

Elementary Bilingual/Dual Language Programs

	Transitional Bilingual Program	**Developmental Bilingual Program**	**Two Way Dual Language Program**
Program Model Goals	College and career-readiness, grade level proficiency in English.	College and career-readiness, grade level proficiency in English and Spanish. Develop bilingualism, biliteracy and multicultural competency.	College and career-readiness, grade level proficiency in English and Spanish. Develop bilingualism, biliteracy and multicultural competency.
Grade Levels	K-5	K-5	K-5
Eligible Students	Grades K-1: Multilingual Learners with Spanish listed as their home language Grades 2-5: Multilingual Learners with Spanish listed as their home language and who are proficient in Spanish Literacy.	Grades K-1: Multilingual Learners with Spanish listed as their home language Grades 2-5: Multilingual Learners with Spanish listed as their home language and who are proficient in Spanish Literacy	Grades K-1 50% Multilingual Learners with Spanish listed as their home language and 50% General Education students. Grades 2-5: 50% Multilingual Learners with Spanish listed as their home language and who are proficient in Spanish Literacy and 50% General Education students.
Program Structure	100% of the class is made up of Multilingual Learners. Students are taught in Spanish while they learn English. As their English proficiency develops, instruction in English gradually increases.	100% of the class is made up of Multilingual Learners or former Multilingual Learners. Students are taught in English and Spanish. Instruction is scaffolded based on standards and language progression for each respective language.	Up to 50% of the class is made up of Multilingual Learners and 50% General Education students. Students are taught using English and Spanish. Instruction is scaffolded based on standards and language progression for each respective language.
Teacher Certifications	All core instruction is provided by a Bilingual (Spanish) certified teacher.	All core instruction is provided by a Bilingual (Spanish) certified teacher.	English core instruction is taught by a Bilingual certified teacher or a teacher certified

<table>
<tr><td></td><td></td><td></td><td>in ESL. Spanish Core instruction is taught by a Bilingual (Spanish) certified teacher.</td></tr>
<tr><td>Language of Instruction</td><td>Academic instruction is delivered in the native language (Spanish) while they learn English. As their English proficiency develops, instruction in English gradually increases.</td><td>In K, instruction is mostly delivered in the native language and 10% in English. By Grade 4, the language allocation should reach the target goal of 50 percent in English and 50 percent in Spanish, continuing on through grade 5.

<table><tr><th>Grade</th><th>English</th><th>Spanish</th></tr><tr><td>K</td><td>10%</td><td>90%</td></tr><tr><td>1</td><td>20%</td><td>80%</td></tr><tr><td>2</td><td>30%</td><td>70%</td></tr><tr><td>3</td><td>40%</td><td>60%</td></tr><tr><td>4</td><td>50%</td><td>50%</td></tr><tr><td>5</td><td>50%</td><td>50%</td></tr></table></td><td>Starting in K, half of the instructional time is delivered in Spanish and half in English. This 50/50 distribution on the language of instruction would remain the same up through Grade 5.

<table><tr><th>Grade</th><th>English</th><th>Spanish</th></tr><tr><td>K</td><td>50%</td><td>50%</td></tr><tr><td>1</td><td>50%</td><td>50%</td></tr><tr><td>2</td><td>50%</td><td>50%</td></tr><tr><td>3</td><td>50%</td><td>50%</td></tr><tr><td>4</td><td>50%</td><td>50%</td></tr><tr><td>5</td><td>50%</td><td>50%</td></tr></table></td></tr>
<tr><td>Exiting from MLL Services</td><td>Multilingual Learners exit the program when they reach grade level proficiency in English.</td><td>All students are encouraged to stay enrolled in the bilingual program through grade 5, including Multilingual Learners who reach grade level proficiency in English.</td><td>All students are encouraged to stay enrolled in the Dual Language program through grade 5, including Multilingual Learners who reach grade level proficiency in English.</td></tr>
</table>

Elementary MLL Programs

	Newcomer ESL Program[44]	**Sheltered ESL Program**	**Integrated ESL Program**	**English Learner Collaborative (ELC) Program**
Program Model Goals	College and career-readiness and grade level proficiency in English.	College and career-readiness and grade level proficiency in English.	College and career-readiness and grade level proficiency in English.	College and career-readiness and grade level proficiency in English.
Grade Levels	2-5	K-5	K-5	K-5
Eligible Students	Multilingual Learners with at least two years of limited or interrupted formal schooling and with limited or no literacy skills in their native language who score less than a 2 out of 6 possible points on the state approved English Language Development assessment.	Multilingual Learners who score less than 3 out of 6 possible points on the state approved English Language Development assessment.	Multilingual Learners who score a 3 or greater out of 6 possible points on the state approved English Language Development assessment	Multilingual Learners who score a 3 or greater out of 6 possible points on the state approved English Language Development assessment
Program Structure	100% of the class is made up of Multilingual Learners. Students are taught in English throughout the day, using effective instructional strategies.	100% of the class is made up of Multilingual Learners. Students are taught in English throughout the day, using effective instructional strategies.	Up to 50% of the class is made up of Multilingual Learners. The rest are general education or special education peers. Students are taught in English throughout the day, using effective instructional strategies.	Up to 25% of the class is made up of Multilingual Learners. The rest are General Education or Special Education peers. Students are taught in English throughout the day, using effective instructional strategies.
Teacher Certifications	ESL certified teachers provide core instruction and English Language Development.	ESL certified teachers provide core instruction and English Language Development.	ESL certified teachers provide core instruction and English Language Development.	General Education or Special Education teachers provide core instruction. ESL certified teachers provide English

[44] Designed to typically be a 2 year program

				Language Development.
Language of Instruction	English is the primary language of instruction.	English is the primary language of instruction.	English is the primary language of instruction.	English is the primary language of instruction.
Exiting from MLL Services	Multilingual Learners exit the program when they reach grade level proficiency in English.	Multilingual Learners exit the program when they reach grade level proficiency in English.	Multilingual Learners exit the program when they reach grade level proficiency in English.	Multilingual Learners exit the program when they reach grade level proficiency in English.

Secondary MLL Programs

	Newcomer ESL Program	**Sheltered ESL Program**	**English Learner Collaborative (ELC) Program**
Program Model Goals	College and career- readiness and grade level proficiency in English.	College and career- readiness and grade level proficiency in English.	College and career- readiness and grade level proficiency in English.
Grade Levels	6-10	6-12	6-12
Eligible Students	Multilingual Learners with at least two years of limited or interrupted formal schooling and with limited or no literacy skills in their native language who score less than a 2 out of 6 possible points on the state approved English Language Development assessment.	Multilingual Learners who score less than 3 out of 6 possible points on the state approved English Language Development assessment.	Multilingual Learners who score a 3 or greater out of 6 possible points on the state approved English Language Development assessment
Program Structure	100% of core classes are made up of Multilingual Learners who meet the Newcomer criteria stated above	100% of English Language Development class is made up of Multilingual Learners who scored less than a 3 on the English Language Development assessment. All other courses may be integrated with General Education/ Special Education peers.	100% of English Language Development class is made up of Multilingual Learners who scored a 3 or higher on the English Language Development assessment. All other courses are integrated with General Education/ Special Education peers.
Teacher Certifications	ESL certified teachers provide English Language development and core instruction.	ESL certified teachers provide English Language Arts and English Language Development. All other courses may be taught by ESL Certified teachers or General/ Special Education teachers who have received training on Sheltered Content Instruction.	ESL certified teachers provide English Language Arts and/or English Language Development. All other courses may be taught by ESL Certified teachers or General/ Special Education teachers who have received training on Sheltered Content Instruction.
Language of Instruction	English is the primary language of instruction.	English is the primary language of instruction.	English is the primary language of instruction.

Exiting from MLL Services	Multilingual Learners exit the program when they reach grade level proficiency in English.	Multilingual Learners exit the program when they reach grade level proficiency in English.	Multilingual Learners exit the program when they reach grade level proficiency in English.

Appendix B. Programs No Longer Taught in PPSD

Program Name	Definition
Transitional Bilingual Model	MLLs receive academic instruction in Spanish while they learn English. As their English proficiency develops, instruction in English gradually increases. Students will exit the program when they meet the exit criteria or move on to middle school and are placed in ESL. See Appendix A for additional details.
Collaboration/ Consultation Model	The Collaboration/Consultation Model requires that the ESL- certified teacher, known as the "Collaborative Teacher," "Provide direct instruction, 30-60 minutes daily, of English Language Development (ELD) to all WIDA Literacy Proficiency levels 1.0 – 2.9 students who are in regular education." If an MLL student is in levels 2.9 and above, the Collaboration/Consultation Model does not require any direct instruction time to the student by the Collaborative Teacher. The Collaboration/Consultation Model further requires that the Collaborative Teacher consult and collaborate with non-ESL certified teachers (i.e., general and/or special education teachers) of MLLs. Collaborative Teachers are required to fill out a Consultation Log every time they consult with the teacher of a student they are servicing."Consultations must take place at a minimum of every 8 weeks." No minimum time per student for the consultation is specified.

Appendix C. MLL Programs Offered by School Year

Program Names	School Years				
	2016-2017	2017-2018	2018-2019	2019-2020	2020-2021
Sheltered ESL					
Integrated ESL					
ESL Push-In					
ESL Newcomer					
Collaborative ESL					
Transitional Bilingual					
Developmental Bilingual					
Two Way Dual Language					
Consultation Model					

Appendix D. MLL Exit Criteria

Dates in Effect	Exit Criteria
March 2019 - Present	1. Student is in grades 1-12 AND 2. ACCESS for MLLs 2.0 Overall score ≥ 4.8 Considerations for students with disabilities: ● Student is in grades 1-12 ● ACCESS for MLLs 2.0 Overall score ≥ 4.8 **or** ● Alternate ACCESS overall score of P2 or P3 for two consecutive years
May 2018 - March 2019	1. Student is in grades 1-12 2. ACCESS Scores: ○ ACCESS for MLLs 2.0 Literacy Score ≥ 4.5 **and** ○ ACCESS for MLLs 2.0 Comprehension Score ≥ 5.0 **and** ○ ACCESS for MLLs 2.0 Speaking Score ≥ 3.0 3. Any three of the following: ○ Passing grades in all core content classes (as reflected on mid or end of year report card attached). ○ ESL/ Bilingual Education Teacher Recommendation ○ At least 2 General Education Teacher Recommendations ○ At least 3 writing samples demonstrating skill not more than one year below grade level ○ Score on district reading assessment not more than one year below grade level. **OR** exit based on Special Education Criteria: An MLL with a disability shall be eligible to exit the English language instructional program if: ● The student has an IEP, **and** ● The student has been continuously enrolled in an ESL/bilingual education program for more than five years, **and** ● The student's overall composite language proficiency score on the ACCESS for MLLs® has not increased more than 10% total over the most recent three testing cycles, **and** ● The IEP team, with input from an ESL/bilingual education professional,

	recommends exit.
June 2016 - May 2018	1. Student is in grades 1-12 2. ACCESS Scores: ○ ACCESS for MLLs 2.0 Literacy Score ≥ 4.5 **and** ○ ACCESS for MLLs 2.0 Comprehension Score ≥ 5.0 **and** ○ ACCESS for MLLs 2.0 Speaking Score ≥ 4.0 **OR** ● PARCC reading ≥ 4.5 **and** ● ACCESS for MLLs 2.0 Overall Composite Score Set by District 3. Any three of the following: ○ Passing grades in all core content classes (as reflected on mid or end of year report card attached). ○ ESL/ Bilingual Education Teacher Recommendation ○ At least 2 General Education Teacher Recommendations ○ At least 3 writing samples demonstrating skill not more than one year below grade level ○ Score on district reading assessment not more than one year below grade level. **OR** exit based on Special Education Criteria: An MLL with a disability shall be eligible to exit the English language instructional program if: ● The student has an IEP, **and** ● The student has been continuously enrolled in an ESL/bilingual education program for more than five years, **and** ● The student's overall composite language proficiency score on the ACCESS for MLLs® has not increased more than 10% total over the most recent three testing cycles, **and** ● The IEP team, with input from an ESL/bilingual education professional, recommends exit.

Appendix E. MLL Program Assignment Business Rules

PPSD's historical data on which specific programs individual students were enrolled in during SY 2016-17 was incomplete. As such, PPSD and TPL compiled a set of rules to determine which program students were enrolled in. These rules relied on a few fields from student demographic information, as well as the school and classrooms the student was enrolled in. Finally, some manual fixes were performed for individual students when the rules resulted in a student being enrolled in a program which was not available at their school. For instance, Feinstein Elementary at Broad Street did not offer ESL Sheltered instruction during SY 2016-17, but our rules resulted in some Feinstein students being marked as ESL Sheltered students. These were corrected by hand.

In this section, we outline the rules we followed to assign students to programs. For more details, please reference the Github repository containing our code.

The "Student Type" field

The primary piece of demographic information we used was the "Student Type" field. It could take on several different values, which we detail below.

Student Type = ESL. During the years of the study, if the Student Type field was ESL, a student could either be in ESL Sheltered or ESL Integrated. For secondary students, there were no ESL Integrated classrooms, so we label them ESL Sheltered. For elementary students, we matched students to the classrooms in which they were enrolled. If the classroom had only MLL students (indicated by the "Current Status Description" flag), then they were labeled as ESL Sheltered. Else, they were labeled as ESL Integrated.

Student Type = ESL Newcomer. These students were all labeled as being in the Newcomer ESL program.

Student Type = SSC & ELC, Sped Self Cont, or ELC. After assigning Students with any of these flags were labeled as being in the Collaborative ESL Model.

Student Type = Bilingual or Bilingual Exit. These students were either part of the Transitional Bilingual or Developmental Bilingual program. Which one depended on which school a student was enrolled in during which year as different programs were offered at different schools.

Student Type = Dual Lang Eng; Dual Lang Span. These students were part of the Two Way Dual Language program. MLL students were coded as "Dual Lang Span"; while General Education students were coded as "Dual Lang Eng". Notably, there were instances where MLL students opted for the ENE option if all "Dual Lang Span" seats were filled, enrolling instead as "Dual

Lang Eng" students. In these instances, although these students were considered ENE, in practice, they were receiving the same MLL services as their MLL peers.

An aside: The Consultation Model. Participation in the Consultation model was separately tracked. PPSD provided TPL with a file enumerating all students in the Consultation Model. These students were labeled as being in the Consultation Model, no matter their Student Type.

Appendix F. Breakdown of Home Languages of Students in the Cohort, SY 2016-17.

	Kindergarten	3rd Grade	6th Grade	9th Grade	Total
Spanish	205	614	255	345	1419
Mayan languages	*	*	*	21	31
Swahili	*	*	*	10	22
Arabic	*	*	*	10	19
English	*	*	*	*	14
Other languages	*	*	*	*	14
Khmer	0	*	*	*	14
X	*	0	*	*	12
Lao	0	*	*	*	*
Portuguese	0	*	*	*	*
Kinyarwanda	*	*	*	*	*
Burmese	*	*	*	*	*
Haitian Creole	0	*	0	*	*
Hmong	0	*	0	*	*
Yoruba	*	*	0	0	*
Afro-Asiatic (Other)	*	0	0	0	*
French	0	*	0	*	*
Creoles and pidgins, Frenchbased (Other)	0	*	0	*	*
Creoles and pidgins, Portuguese-based (Other)	0	*	0	0	*
Somali	0	0	0	*	*
Indic (Other)	0	*	0	0	*
Urdu	0	0	0	1	*
Wolof	0	*	0	0	*
Vietnamese	0	*	0	0	*
Uzbek	*	0	0	0	*
Multiple languages	0	*	0	0	*
Tigrinya	0	0	0	*	*
Japanese	0	*	0	0	*
Samoan	0	0	*	0	*
Quechua	0	0	*	0	*
Iranian (Other)	0	0	0	*	*
Philippine (Other)	0	0	0	*	*
Nepali	0	*	0	0	*

Appendix G. Deriving our Study Population

The following table gives a high-level overview of the steps taken to derive our final cohort population from the population of all students who attended PPSD during our study period (SY 2016-17 to SY 2018-19). In particular, the "Step" column highlights the key steps in the process, the "Category" column provides a short description of the logic implemented, the "Number" column the counts of unique LASIDs that adhere to the specified logic, and the "Notebook" column specifies the location of the code where this logic is implemented. Specifically, please refer to the Jupyter notebook cells tagged "population".

Step	Category	Number	Notebook
	Students that have attended PPSD from SY 2016-17 to SY 2018-19	34,726	210
	Outplaced Students	893	210
Remove Outplace Students.	**Students that have attended PPSD from SY 2016-17 to SY 2018-19 after removing Outplaced Students.**	**33,833**	**210**
	Students with 0 Enrollment	2	210
	Students with 0 Attendance	380	210
Remove students with 0 Enrollment and Attendance.	**Students that remain after removing students with 0 Enrollment and Attendance.**	**33,451**	**210**
	Students with Correct Starting Grade (kindergarten, 3rd, 6th, 9th) in SY 2016-17	8,229	230
Keep only students with the correct starting grade who also have active MLL status for the final cohort population.	**Students with Correct Starting Grade with active MLL status.**	**2,472**	**230**

Appendix H. Regression Tables for Quasi-experimental Analyses

Results from each regression from the quasi-experimental analysis section. These tables represent the top-line numbers from the regression, including the measured effect size and its standard error, *t* value, and *p* value, the results of the omnibus balance test (its χ^2 value and its *p* value), and the Wilcoxon test for robustness (its *z* value and its *p* value). Each cohort appears on a separate page. A dash (-) in a row means that a match was impossible.

Cohort 1: Kindergarten Students in SY 2016-17

There are 7 successful matches for the ACCESS exam and 6 for the STAR Reading exam. As such, according to our Bonferroni correction, an ACCESS *p*-value will be significant if it is at most 0.05/7 = 0.007 and a Reading *p*-value will be significant if it is at most 0.05/6 = 0.008. Such *p*-values are starred.

Table 25: Quasi-experimental analysis results for the kindergarten cohort.

Program	Outcome	Measured Effect	Measured Std. Err.	Measured t value	Measured p value	Balance χ^2 value	Balance p value	Wilcoxon z value	Wilcoxon p value
ESL Integrated	ACCESS	– 8.13	3.71	–2.19	0.03	15.69	0.11	– 1.80	0.07
	Reading	-	-	-	-	-	-	-	-
ESL Sheltered	ACCESS	– 5.28	2.83	– 1.86	0.07	21.27	0.03	– 2.10	0.04
	Reading	-	-	-	-	-	-	-	-
Consultation Model	ACCESS	-	-	-	-	-	-	-	-
	Reading	-	-	-	-	-	-	-	-
Collaborative ESL	ACCESS	4.77	3.68	1.30	0.20	12.95	0.16	1.74	0.08
	Reading	6.62	18.59	0.36	0.72	13.28	0.21	0.62	0.55
Transitional Bilingual	ACCESS	8.55	4.30	1.99	0.05	3.58	0.61	2.42	0.02
	Reading	– 17.96	19.65	– 0.91	0.36	5.24	0.73	– 0.34	0.74
Dev. Bilingual	ACCESS	2.93	6.32	0.46	0.64	10.60	0.10	0.24	0.80
	Reading	– 10.45	28.09	– 0.37	0.71	3.80	0.70	– 0.54	0.60
Two Way D.L.	ACCESS	9.15	5.78	1.58	0.13	5.88	0.12	1.37	0.18
	Reading	**191.30**	**31.43**	**6.09**	**0.00***	**4.73**	**0.69**	**3.55**	**0.00**
ENE	ACCESS	16.17	15.36	1.05	0.34	5.91	0.21	1.03	0.34
	Reading	– 38.67	55.72	– 0.69	0.52	3.94	0.56	– 0.98	0.50

Cohort 2: Third Grade Students in SY 2016-17

There are 7 successful matches for the ACCESS exam, 7 for the RICAS ELA exam, and 5 for the RICAS Math exam. According to our Bonferroni correction, an ACCESS or ELA *p*-value will be significant if it is at most 0.05/7 = 0.007 and a Math *p*-value will be significant if it is at most 0.05/5 = 0.01. Such *p*-values are starred.

Table 26: Quasi-experimental analysis results for the third grade cohort.

Program	Outcome	Measured Effect	Measured Std. Err.	Measured t value	Measured p value	Balance χ^2 value	Balance p value	Wilcoxon z value	Wilcoxon p value
ESL Integrated	ACCESS	1.21	3.05	0.40	0.69	13.34	0.27	0.31	0.76
	ELA	2.39	2.08	1.15	0.26	19.26	0.08	1.15	0.25
	Math	– 0.23	1.81	– 0.13	0.90	23.35	0.04	– 0.35	0.74
ESL Sheltered	ACCESS	-	-	-	-	-	-	-	-
	ELA	-	-	-	-	-	-	-	-
	Math	-	-	-	-	-	-	-	-
Consultation	ACCESS	5.00	33.00	0.15	0.90	1.00	0.32	– 0.00	1.00
	ELA	9.50	11.50	0.83	0.56	2.00	0.37	0.63	1.00
	Math	– 17.50	9.50	– 1.84	0.32	2.00	0.37	– 1.41	0.50
Collaborative ESL	ACCESS	– 1.22	4.06	– 0.30	0.77	7.29	0.30	0.30	0.77
	ELA	2.92	2.22	1.31	0.20	16.05	0.10	1.17	0.25
	Math	– 6.86	2.83	– 2.42	0.02	12.18	0.14	– 2.36	0.01
Transitional	ACCESS	0.73	3.40	0.21	0.83	5.54	0.14	0.73	0.46
	ELA	– 5.00	2.05	– 2.44	0.02	14.24	0.08	– 1.96	0.05
	Math	-	-	-	-	-	-	-	-
Developmental	ACCESS	7.69	3.50	2.20	0.03	4.07	0.40	1.78	0.07
	ELA	**8.67**	**2.41**	**3.59**	**0.00***	**6.72**	**0.35**	**3.24**	**0.00**
	Math	6.22	2.64	2.36	0.02	12.72	0.03	2.26	0.02
Two Way D.L.	ACCESS	7.64	9.49	0.80	0.44	0.62	0.43	1.34	0.19
	ELA	– 5.00	3.72	– 1.34	0.21	3.75	0.44	– 1.08	0.30
	Math	2.45	5.20	0.47	0.65	3.44	0.33	1.05	0.32
ENE	ACCESS	– 2.14	2.89	– 0.74	0.46	10.48	0.16	– 1.15	0.25
	ELA	1.27	1.81	0.70	0.48	24.23	0.01	0.35	0.73

Math	-	-	-	-	-	-	-	-	-

Cohort 3: Sixth Grade Students in SY 2016-17

There are 3 successful matches for the ACCESS exam, 2 for the RICAS ELA exam, and 2 for the RICAS Math exam. According to our Bonferroni correction, an ACCESS *p*-value will be significant if it is at most 0.05/3 = 0.017 and an ELA or Math *p*-value will be significant if it is at most 0.05/2 = 0.025. Such *p*-values are starred.

Table 27: Quasi-experimental analysis results for the sixth grade cohort.

Program	Outcome	Measured Effect	Measured Std. Err.	Measured t value	Measured p value	Balance χ^2 value	Balance p value	Wilcoxon z value	Wilcoxon p value
ESL Sheltered	ACCESS	-	-	-	-	-	-	-	-
	ELA	-	-	-	-	-	-	-	-
	Math	-	-	-	-	-	-	-	-
Consultation	ACCESS	6.00	48.00	0.12	0.92	1.00	0.32	– 0.00	1.00
	ELA	10.00	9.00	1.11	0.47	2.00	0.37	1.41	0.50
	Math	– 6.00	3.00	– 2.00	0.30	2.00	0.37	– 1.41	0.50
Collaborative ESL	ACCESS	– 15.94	6.82	– 2.34	0.03	7.44	0.19	– 1.97	0.04
	ELA	-	-	-	-	-	-	-	-
	Math	-	-	-	-	-	-	-	-
ENE	ACCESS	– 9.50	8.93	– 1.06	0.30	3.68	0.30	– 1.14	0.26
	ELA	– 9.65	3.84	– 2.52	**0.02***	6.58	0.25	– 2.02	0.04
	Math	– 3.89	3.22	– 1.21	0.24	9.32	0.16	– 1.29	0.21

Cohort 4: Ninth Grade Students in SY 2016-17

There are 2 successful matches for the ACCESS exam, 1 for the RICAS ELA exam, and 2 for the RICAS Math exam. According to our Bonferroni correction, an ACCESS or Math *p*-value will be significant if it is at most 0.05/2 = 0.025 and a Reading *p*-value will be significant if it is at most 0.05/1 = 0.025. Such *p*-values are starred.

Table 28: Quasi-experimental analysis results for the ninth grade cohort.

Program	Outcome	Measured Effect	Measured Std. Err.	Measured t value	Measured p value	Balance χ^2 value	Balance p value	Wilcoxon z value	Wilcoxon p value
ESL Sheltered	ACCESS	-	-	-	-	-	-	-	-
	ELA	-	-	-	-	-	-	-	-
	Math	-	-	-	-	-	-	-	-
Collaborative ESL	ACCESS	– 26.67	16.42	– 1.62	0.25	3.00	0.39	– 1.09	0.50
	ELA	– 25.00	15.00	– 1.67	0.34	2.00	0.37	– 1.41	0.50
	Math	36.67	50.44	0.73	0.54	3.00	0.39	0.43	1.00
ENE	ACCESS	8.07	5.94	1.36	0.18	5.92	0.43	1.31	0.19
	ELA	-	-	-	-	-	-	-	-
	Math	16.67	13.46	1.24	0.23	9.14	0.33	1.27	0.21

Appendix I. Cohort vs. MLL Student Population Comparisons

In the previous sections, we provided some descriptive statistics about the overall MLL population in PPSD and the MLL population who make up our cohort. In this section we ask the question of whether there are any statistically significant differences between the two populations. In order to make this comparison, we perform a propensity score analysis.

We attempt to answer three questions:

- Do demographic factors predict whether a student stays enrolled in PPSD for the three years of the study, i.e. that they remain in our cohort?
- Do demographic factors predict which program a student enrolls in?
- Do demographic and behavioral factors (e.g. attendance rate) predict whether a student will complete all the exams described in Section 2.2.3 Complete Cases?

Throughout this section, we note that many of the implicit subgroups studied by our regressions are very small. This leads to potential interpretability problems for some regression coefficients, and so we caution that the results of this section should be interpreted mostly as whether or not our ultimate study population is representative of the broader MLL population in PPSD, and statistically significant coefficients should indicate to the reader potential limitations on interpreting those coefficients.

I.1 Do demographic factors predict sustained enrollment?

In order to answer this question, for each of kindergarten, third grade, sixth grade, and ninth grade, we run a logistic regression predicting a binary outcome variable of whether a particular MLL student who enrolled in that grade in SY 2016-17 remained enrolled in PPSD for all three years of our study. On the right-hand side of that regression, we place indicators for a student's:

- gender
- race
- FRPL status
- SPED status
- SIFE status
- home language
- school

For this section, we collapsed a student's FRPL status to either "Free or Reduced" or "Paid". We also collapsed students' home language to "Spanish," "Other," or "Missing" (see Appendix F for

specific breakdown of student home languages in SY 2016-17). Finally, we include an indicator for which school a student attended.

The output of this regression analysis can be found in Table 31 in Appendix K. The main takeaways, however, can be found in Table 29. We note that a student missing their home language data frequently appears as a risk factor for exiting PPSD within the three years of our study. The consistency of this factor leads us to believe that it is not that a student's home language is missing that is *per se* reducing their propensity to stay in PPSD, but that it likely indicates some other effect is being captured by this missingness. This is especially true as the home language survey is ostensibly completed by all students in the district, even though the data from it may not end up in the district's data warehouse.

We note the worrying fact that in ninth grade, male students are particularly at risk of leaving PPSD versus their female peers, as evinced by the large negative coefficient in Table 31, Column (9), Row "Gender = Male". Indeed, looking at raw numbers (Table 30), 66% of female students stay in the ninth grade cohort for the entire study period whereas only 51% of male students do.[45]

We also note that a small number of schools showed significant effects in retaining students across the cohort time period. Specifically, attending kindergarten at George J. West Elementary School and Feinstein (Broad St.) Elementary School appears correlated with a likelihood to remain enrolled in PPSD, as does attending Providence Career and Technical High School in the ninth grade. While these correlations were highly statistically significant and had ostensibly large effects, we note that the number of students attending any given school is small. Still, it may be desirable to perform some qualitative outreach as to why these schools in particular may be retaining students.

Of the remaining factors associated with staying in the cohort or leaving PPSD, we note that there are less than ten multiracial kindergarten students, and therefore this may be a small numbers phenomenon. On the other hand, there are 35 Asian students in the potential kindergarten cohort and 97 in the potential third grade cohort, so it is more troubling that these are risk factors.

Similarly, SIFE students in the ninth grade are significantly less likely to remain in the cohort, though not at the differential rates observed between male and female students.

[45] In discussions with PPSD staff about these findings, they pointed out that there is a broader concern about Newcomer male students in high school leaving school to find work to support their families. Qualitative follow up research may help elucidate the specific needs of this population.

I.2 Do demographic factors predict students' programs?

In order to answer this question, for each of kindergarten, third grade, sixth grade, and ninth grade, we run several logistic regressions predicting which program a student enrolled in during SY 2016-17. On the right hand side of the regression, we use the same regressors as in the previous section, except we do *not* control for school. This is because it is unclear to us whether a school might drive program enrollment (e.g. the nearest school to a student's home doesn't offer a particular program) or whether a desire to enroll in a particular program might drive a particular school enrollment (e.g. only two schools offered Two Way Dual Language instruction).

Table 29. (Relative) protective and risk factors (non-school) for remaining in the cohort.

Cohort	Protective Factors	Risk Factors
Kindergarten		race = Asian or Multiracial home language = Missing lunch status = Paid
Third Grade		race = Asian home language = Missing
Sixth Grade		home language = Missing
Ninth Grade	race = Asian	gender = Male SIFE = True lunch status = Paid

A protective factor is a demographic factor that predicts a student will stay enrolled in the cohort. A risk factor is one that predicts they will not remain enrolled in the cohort.

Table 30. Ninth grade MLL students in our cohort and not in our cohort, broken down by gender.

	Female	Male	Total
In Cohort	197	248	445
Not in Cohort	102	240	342
Total	299	488	787

To start, we ran a full multinomial logistic regression[46] (with ESL Sheltered as the reference group) for each cohort, attempting to predict which program a particular student ultimately enrolled in based on their demographics (see Tables 32-35 in Appendix K). For the kindergarten and third grade cohorts, these regression models are likely underpowered, as evidenced by the large size of many coefficients. This is likely due to the fact that given the large number of programs and the large number of demographics relative to the number of students in our cohorts, there was not enough data to get good precision on many of these estimates. There are also some known multicollinearities. For instance, Two Way Dual Language instruction was only available for the third grade cohort at Leviton Dual Language Academy, which was enrolled in the Community Eligibility Program for Free and Reduced Price Lunch (Rhode Island Department of Education, n.d.). Thus, it is unclear whether FRPL status is downstream of the program enrollment or the other way around.

Still, from these multinomial regressions, we do find certain trends in coefficients that are of potential interest as they reiterate several of the observations from Section 2.2.2 Descriptive Analyses. Specifically, we note that SPED students seem more likely to enroll in the Consultation and Collaborative ESL programs instead of programs such as ESL Sheltered.[47] Moreover, perhaps unsurprisingly, students who had demographic characteristics less associated with being Hispanic (e.g. having race Asian or Black, having a home language other than Spanish, etc.) were less likely in the third grade cohort to be enrolled in programs which are targeted at Spanish speaking students (such as the Developmental Bilingual and Two Way Dual Language programs). As we saw in Section 5.1 Research Question 1, these programs in particular appear to perform very well on some metrics. As such, PPSD may wish to more deeply investigate their program offerings for non-Spanish speaking students.

Similarly, for the sixth and ninth grade cohorts, while we have fewer programs, the vast majority of students are enrolled in the ESL Sheltered or Newcomer programs (and for students in the sixth grade cohort, the vast majority are enrolled in ESL Sheltered alone). As such, Tables 34 and 35 display an implausibly large number of significant coefficients.

Given how thin this stretches the data, it is perhaps better to look at a single program. In particular, if we regress whether a student is enrolled in ESL Sheltered on their demographics, we get Table 36. Here, two trends are much clearer: SIFE students are much less likely to be enrolled in ESL Sheltered than all other students in the sixth and ninth grade cohorts. This is as expected, and is therefore a positive finding. Similarly, SPED students are much *less* likely to be enrolled in ESL Sheltered across all cohorts, which is consistent with the discussion in the previous paragraph. Additionally, while male students in the sixth grade cohort are less likely to be enrolled in ESL Sheltered, this is isolated to this single cohort.

[46] We utilize the multinom function in the nnet R package (version 7.3-16) (Ripley & Venables, 2021).
[47] In conversations with PPSD staff, we learned that this may be due to the fact that SPED teachers are not dual-certified as ESL teachers and so may not be able to provide ESL services, a problem the Consultation model was built to overcome. As such, this is an unsurprising result.

I.3 Do demographic factors predict whether students complete all the exams in our study?

As discussed in Section 2.2.3 Complete Cases, our quasi-experimental analyses relies on students who completed all of the assessments that we are studying. These concerns are also present in the descriptive analyses of typical test scores of students in our cohort. As such, for each of kindergarten, third grade, sixth grade, and ninth grade, we run a logistic regression predicting whether a student in our cohort (i.e. a student who remains enrolled in PPSD for all three years) completes all exams described in Table 4 over the course of the study. On the right-hand side of the regression, we use the same regressors as in Section I.1, except we do *not* control for school and we *do* control for the student's absenteeism rate as described in Section 4.3.1 Student Attendance.

The output of this regression analysis can be found in Table 37 in Appendix K. Most notably (and perhaps unsurprisingly), sixth and ninth grade SIFE students seem to be much less likely to complete all the exams necessary to be considered complete cases.

Of particular interest, however, are the coefficients on various absenteeism regressors and on the student gender indicator. Here we see that being *excessively* absent (i.e. absent between 20-50% of school days) is highly correlated with not completing exams across the board and is statistically significant in the kindergarten and sixth grade cohorts. On the other hand, being *chronically absent* (i.e. absent between 10-20% of school days) is sometimes *positively* correlated with completing all exams (e.g. in the third grade cohort). This is surprising, though this may be due to selection effects of those students who are chronically absent being more likely to leave the cohort entirely.

A similar phenomenon can be seen among male students, who in the third and sixth grade cohorts are significantly more likely to complete all exams. From the previous section, we know that male students are more likely to leave the cohort entirely, so this may simply be due to selection effects.

Appendix J. Research Questions 3 and 4

Research Question 3

Do certain subpopulations of MLL students (e.g. students with certain home language types, students who enroll in an MLL program at a younger age, students with certain classifications, etc.) have better educational outcomes than other subpopulations? Do these differences in educational outcomes vary across programs?

Under the heading of this research question, we set out to explore the following question:

- Do MLL students who enter MLL services at younger ages exit earlier than those who enter later?
- Do MLL students who speak Romance languages at home according to the Home Language Survey perform better on standardized tests?

Literature review

MLL Program Exit Rates

The expected duration of time in MLL programs varies by a number of factors, including the testing requirements to reach reclassification status. Generally, the literature uses a measure of parity with native English speakers on oral and written tasks. The average length of time in MLL programs was between three and four years (Greenberg Motamedi, 2015; MacSwan & Pray, 2005), with listening and speaking proficiency achieved before reading and writing proficiency (Hakuta et al., 2000; Thompson, 2017). While different studies state different expected durations, boys, students eligible for special education, and students whose parents had attained lower levels of education generally took longer to exit their MLL programs (Greenberg Motamedi, 2015; Thompson, 2017).

Age of Entry

The general consensus in the literature is that older (approximately ages 10-12, depending on the study) students' rate of language acquisition is faster than younger learners, but that students exposed to English at an earlier age will ultimately achieve a higher level of fluency (Collier, 1987; Dixon, 2012; Krashen, 1979; MacSwan, 2005). However, because younger entrants into MLL programs are also more likely to have lived in the US longer, observational

studies often find a faster rate of language acquisition with this population (Conger, 2009; Greenberg Motamedi, 2015).

Student Home Language

Theoretically, MLLs whose home language is a romance language should be able to leverage the languages' shared Latin roots in order to learn English more quickly. However, this may only be relevant if students are made aware of the relationship between their home language and English (Hickey & Lewis, 2013), making the impact of knowledge of romance home languages on English acquisition hard to measure (Elder & Davies, 1998; Grant et al., 2012). Still, observational studies suggest that there is some evidence of a beneficial relationship (Gómez, 2009; Marinova-Todd et al., 2013).

Our Findings

Unfortunately, data availability and study restrictions hampered our ability to rigorously answer these questions beyond the descriptive analyses above.

Age of Entry. We were not able to obtain information on the age of student entry into MLL programs in time for producing this report. However, we do note that language acquisition is slower in older MLL students at PPSD based on changes to their ACCESS scores. Of course, older MLL students likely have different characteristics than younger children, which may explain the slower rate of language acquisition.

One such group may be students who have remained in the MLL program for an extended period of time. Given their many years struggling to acquire English language proficiency, then, it would be unsurprising that their linguistic and academic achievement remained low throughout high school. More could be done to understand how many of such students exist in PPSD, and the barriers to their learning that the district could help mitigate.

Another group might be students who are recent immigrants to America (Newcomers), and who may also have experienced interruptions to their education (SIFE). Since Newcomer and SIFE students in our study are primarily concentrated in the ninth grade cohort[48] (Table 7), this seems a likely scenario. Although Newcomer and SIFE students of all ages may face unique challenges, older students have different or additional difficulties compared to what younger MLL students might experience, such as having to master English at a much higher level before they can access the material from other subjects and having more years of learning to catch up on.

[48] Across our cohorts, 100% of Newcomers are in the ninth grade cohort, 56.6% of SIFE students are in the ninth grade cohort, and 96.4% of students who are classified as both Newcomer and SIFE are in the ninth grade cohort (Table 7).

Students in our ninth grade cohort also have lower attendance rates on average, which may impact their learning. Relatedly, older students likely have greater familial responsibilities, which may then take time away from school and learning.

Student Home Language. The student home language data seems to contain suspicious missingness patterns that make us wary of using it for these sorts of studies. While on the whole we are missing data on home language for about 15% of students, over 40% of students in our kindergarten cohort lack home language data. Specifically, students who are enrolled in programs which are targeted at Spanish-speaking students (such as the Developmental Bilingual program) almost always have their home language data filled in (and it is almost always Spanish). This raises the concern that a filled-in home language survey is downstream of participation in particular programs, whereas we would want participation in particular programs to be downstream of actual home language.

Research Question 4

Do enrolled MLL student outcomes vary by teacher certification status and certification type?

Literature Review

There are mixed results regarding the impact of teacher certification on MLL academic outcomes. While some studies show that teacher certification has positive effects (López et al., 2013; Hayes & Salazar, 2001), others show no effect at all (Master et al., 2012; Tracy, 2009). Once teacher years of experience is taken into account, however, teacher certification seems to only have positive effects for novice teachers (Master et al., 2012). In addition, other teacher characteristics, such as years of experience teaching MLL students and preference to teach MLL students may be more important for improving MLL outcomes (Master et al., 2012) than teacher certification itself.

Unfortunately, as mentioned earlier in this report, due to a lack of complete data regarding teacher years of experience and demographics, we were unable to conduct statistical analyses to answer this research question. However, the question of whether and the degree to which teacher certification status affects students outcomes, and if so, whether a particular certification type matters, remains a pertinent one. We discuss this further in Section 8. Next Steps.

Appendix K. Propensity Analysis Regression Tables

In this section we provide the detailed tables discussed in Appendix I. Cohort vs. MLL Student Population Comparisons.

Table 31. Results of propensity analysis in Section I.1: Cohort inclusion as the outcome.

	Dependent variable:			
	Included in Cohort			
(Cohort)	(K)	(3)	(6)	(9)
Gender = Male	-0.129	0.001	0.319	-0.595***
	(0.264)	(0.245)	(0.325)	(0.175)
Race = Asian	-1.631**	-1.453**	1.517	1.389**
	(0.718)	(0.646)	(1.166)	(0.652)
Race = Black	-0.566	-1.092*	-0.457	0.463
	(0.502)	(0.653)	(0.618)	(0.327)
Race = Multiracial	-3.449***	17.971	0.030	-0.279
	(1.336)	(2,722.134)	(1.123)	(0.710)
Race = Native American		-1.422	13.462	0.175
		(1.259)	(1,455.398)	(0.832)
Race = Pacific Islander	-1.906	-0.912		13.649
	(1.489)	(1.828)		(882.743)
Race = White	0.733	-0.802	0.892	0.771
	(0.618)	(0.696)	(0.940)	(0.480)
SIFE	-0.822	1.000	1.118	-0.471**
	(1.454)	(0.835)	(1.111)	(0.213)
Special Education	-0.317	-0.261	0.810	0.709*
	(0.468)	(0.388)	(0.525)	(0.424)
Lunch Status = Paid	-1.111***	0.487	-0.624	-0.556**
	(0.349)	(0.432)	(0.418)	(0.247)
School = Alan Shawn Feinstein Elementary School (Broad)	1.727**	-0.289		
	(0.871)	(0.598)		
School = Alfred Lima Elementary School		0.022		
		(0.547)		
School = Anthony Carnevale Elementary School	0.890	0.776		
	(0.905)	(0.768)		
School = Asa Messer Elementary School	-0.351	0.039		
	(0.481)	(0.548)		
School = Carl G. Lauro Elementary School	17.084			
	(711.819)			
School = Charles Fortes Elementary School	0.896			
	(0.646)			
School = Dr. Martin Luther King Jr. Elementary School	20.018	0.216		
	(3,956.181)	(0.562)		
School = Frank Spaziano Elementary School	0.673	16.652		
	(0.627)	(910.417)		
School = George J. West Elementary School	1.520**	1.682*		
	(0.738)	(0.870)		

School = Harry Kizirian Elementary School	-0.174	-0.019		
	(0.525)	(0.626)		
School = Leviton Dual Language	0.469	-0.151		
	(0.670)	(0.660)		
School = Lillian Feinstein Elementary School (Sackett)	0.701	0.411		
	(0.681)	(0.727)		
School = Mary Fogarty Elementary School	0.762	-0.215		
	(0.709)	(0.498)		
School = Pleasant View Elementary School	0.392	0.633		
	(1.154)	(0.958)		
School = Reservoir Avenue Elementary School	-0.376	0.671		
	(0.589)	(1.111)		
School = Robert F. Kennedy Elementary School	1.342	0.291		
	(0.912)	(0.911)		
School = Robert L. Bailey Elementary School	16.339	-1.161		
	(1,232.004)	(0.807)		
School = Del Sesto Middle School			0.705	
			(0.603)	
School = Esek Hopkins Middle School			-0.329	
			(0.543)	
School = Nathan Bishop Middle School			-0.367	
			(0.561)	
School = Nathanael Greene Middle School			0.653	
			(0.555)	
School = Roger Williams Middle School			0.429	
			(0.547)	
School = 360 High School				-0.465
				(0.459)
School = Academy for Career Exploration				-0.583
				(0.709)
School = Central High School				-0.308
				(0.268)
School = Classical High School				13.145
				(882.743)
School = E Cubed Academy				0.224
				(0.560)
School = Evolutions High School				-0.824*
				(0.468)
School = Hope High School				0.156
				(0.318)
School = JSEC				-0.391
				(0.298)
School = Mount Pleasant High School				-0.289
				(0.261)
School = Providence Career and Technical High School				1.666**
				(0.831)
School = Times2	16.095	16.603	12.971	13.635
	(1,539.694)	(2,153.936)	(1,455.398)	(620.085)
School = Vartan Gregorian Elementary School	-1.932	0.647		
	(1.302)	(1.036)		
School = Veazie Street Elementary School	-0.475	0.862		
	(0.698)	(0.860)		

School = Webster Avenue Elementary School	0.679	1.223		
	(0.745)	(1.169)		
School = William D'Abate Elementary School	0.341	1.574		
	(0.620)	(1.102)		
School = Woods/Young Elementary School		0.107		
		(0.473)		
School = West Broadway Middle School			-0.291	
			(0.652)	
Home Language = Missing	-0.864***	-2.678***	-2.427***	-2.707***
	(0.293)	(0.340)	(0.431)	(0.282)
Home Language = Other	0.137	0.152	-0.984	-0.026
	(0.618)	(0.614)	(0.602)	(0.282)
Constant	1.861***	2.132***	1.785***	1.308***
	(0.457)	(0.359)	(0.438)	(0.249)
Observations	502	784	364	787
Log Likelihood	-200.882	-245.303	-131.010	-436.189
Akaike Inf. Crit.	467.764	558.607	300.020	920.378

Note: *p<0.1 **p<0.05 ***p<0.01

Table 32. Results of propensity analysis in Section I.2 with program designation as the outcome, kindergarten cohort.

	Dependent variable:						
	Collaborative	Consultation	Dev Bilingual	ENE	Integrated ESL	Trans Bilingual	Two Way Dual Language
Gender = Male	-0.433	-1.566	-0.459	-0.219	-0.187	-0.510	-0.326
	(0.335)	(1.159)	(0.348)	(0.864)	(0.322)	(0.311)	(0.498)
Race = Asian	-0.497	-7.315	-118.413	-38.416	0.990	-125.306***	-129.688***
	(1.245)	(432.912)			(0.864)	(0.000)	(0.000)
Race = Black	-1.444*	1.807	-1.223	1.334	-0.897	-2.084*	-107.500***
	(0.874)	(1.761)	(0.891)	(1.262)	(0.725)	(1.098)	(0.000)
Race = Multiracial	133.693***	-0.651	-47.668***	156.022***	-56.233***	133.452***	-35.043***
	(0.723)		(0.000)	(0.000)	(0.000)	(0.723)	(0.000)
Race = Native American	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Race = Pacific Islander	-53.801	-14.413***	-89.830	-6.351	-102.300***	-84.258***	-2.357***
		(0.00000)			(0.000)	(0.000)	(0.000)
Race = White	0.455	-50.464***	-87.509***	-58.368***	0.509	-0.154	0.457
	(0.608)	(0.000)	(0.000)	(0.000)	(0.574)	(0.660)	(0.863)
SIFE	-114.190***	-12.762***	-95.902***	-0.113***	-106.535***	-120.862***	-80.820***
	(0.000)	(0.00002)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Special Education	1.929***	4.874***	0.316	-60.868	0.796	-0.987	-0.379
	(0.502)	(1.391)	(0.593)		(0.553)	(0.822)	(1.119)
Lunch Status = Paid	-0.979	2.439**	-0.566	-86.486	0.286	-0.877	-71.743***
	(0.678)	(1.237)	(0.609)		(0.463)	(0.601)	(0.000)
Home Language = Missing	0.672*	0.802	-0.767**	-69.273	0.275	-0.071	0.100
	(0.360)	(1.314)	(0.377)		(0.341)	(0.318)	(0.513)
Home Language = Other	0.355	-80.946	-0.369	-131.324	-0.019	-0.894	1.007
	(0.846)		(0.940)		(0.725)	(1.154)	(1.229)
Constant	-0.832**	-5.305***	-0.007	-1.829***	-0.695**	0.152	-1.356***
	(0.344)	(1.630)	(0.298)	(0.646)	(0.324)	(0.281)	(0.458)
Akaike Inf. Crit.	1,484.692	1,484.692	1,484.692	1,484.692	1,484.692	1,484.692	1,484.692

Note: *p<0.1 **p<0.05 ***p<0.01

Table 33. Results of propensity analysis in Section I.2 with program designation as the outcome, third grade cohort.

	Dependent variable:						
	Collaborative	Consultation	Dev Bilingual	ENE	Integrated ESL	Trans Bilingual	Two Way Dual Language
Gender = Male	-0.630**	-0.544	-0.149	0.166	-0.107	-0.524**	-0.612
	(0.301)	(0.840)	(0.305)	(0.238)	(0.259)	(0.255)	(0.495)
Race = Asian	0.860	3.260**	-14.837***	0.217	0.913	-17.510***	-11.632***
	(0.966)	(1.583)	(0.00000)	(0.792)	(0.781)	(0.00000)	(0.00001)
Race = Black	0.479	3.917**	0.880	-0.054	1.424*	-17.203***	-11.985***
	(1.057)	(1.721)	(1.220)	(0.825)	(0.752)	(0.00000)	(0.00000)
Race = Multiracial	-5.285***	0.566	-3.109***	-4.916***	21.960***	21.034***	-0.591***
	(0.000)		(0.000)	(0.000)	(0.650)	(0.650)	(0.000)
Race = Native American	-23.059***	-19.379***	-19.329***	-25.148	-22.920***	-21.733***	-19.067***
	(0.000)	(0.00000)	(0.000)		(0.000)	(0.000)	(0.00000)
Race = Pacific Islander	-2.782	-0.064***	-2.802	-3.724	24.330***	-4.802***	-1.248***
		(0.000)			(0.000)	(0.000)	(0.000)
Race = White	-0.036	-9.147	-0.884	-19.701***	-20.749***	-1.435	-18.645***
	(0.829)	(421.945)	(1.088)	(0.00000)	(0.00000)	(1.085)	(0.00000)
SIFE	-1.604	2.277	-20.464***	-19.013***	-1.935*	-20.895***	-17.561***
	(1.182)	(1.526)	(0.000)	(0.00000)	(1.081)	(0.000)	(0.000)
Special Education	1.759***	4.030***	-1.019	1.543***	0.952*	-0.377	1.064
	(0.488)	(1.008)	(1.077)	(0.439)	(0.501)	(0.697)	(0.845)
Lunch Status = Paid	-0.613	-22.834	-0.131	-0.504	0.092	-0.363	-23.121***
	(0.541)		(0.535)	(0.414)	(0.394)	(0.473)	(0.000)
Home Language = Missing	0.576	-29.259	-0.496	-13.933	-0.244	-1.204	-19.777***
	(0.555)		(0.812)	(393.362)	(0.616)	(0.828)	(0.00000)
Home Language = Other	-0.909	-3.437*	-24.043***	0.216	-0.681	-1.302	-17.336***
	(0.913)	(1.954)	(0.000)	(0.698)	(0.701)	(1.183)	(0.00000)
Constant	-0.825***	-4.214***	-0.795***	-0.363*	-0.634***	-0.044	-1.684***
	(0.222)	(0.867)	(0.230)	(0.193)	(0.207)	(0.183)	(0.332)
Akaike Inf. Crit.	2,485.874	2,485.874	2,485.874	2,485.874	2,485.874	2,485.874	2,485.874

Note: *p<0.1 **p<0.05 ***p<0.01

Table 34. Results of propensity analysis in Section I.2 with program designation as the outcome, sixth grade cohort.

	Dependent variable:			
	Collaborative	Consultation	ENE	Newcomer
Gender = Male	1.140*	0.204	1.042**	-44.479***
	(0.683)	(1.554)	(0.496)	(0.471)
Race = Asian	5.151	-13.091***	-9.511***	-19.581***
		(0.000)	(0.000)	(0.000)
Race = Black	27.386***	-4.386***	-26.303***	-24.545***
	(0.000)	(0.000)	(0.000)	(0.000)
Race = Multiracial	-2.971	3.364***	-50.311***	0.362***
		(0.000)	(0.000)	(0.000)
Race = Native American	-0.623***	0.141***	-30.428***	-0.056***
	(0.000)	(0.000)	(0.000)	(0.000)
Race = Pacific Islander	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Race = White	-79.852	-70.201***	-57.694***	-7.674***
		(0.000)	(0.000)	(0.000)
SIFE	-7.584***	12.899***	-24.426***	90.177***
	(0.000)	(0.000)	(0.000)	(0.471)
Special Education	75.295***	64.280***	2.197***	-0.461***
	(0.287)	(0.666)	(0.458)	(0.000)
Lunch Status = Paid	1.256	-10.415	-0.713	-51.130***
	(1.531)	(1,005.072)	(1.068)	(0.000)
Home Language = Missing	0.198	3.799**	-53.839***	5.763***
	(1.480)	(1.613)	(0.000)	(0.000)
Home Language = Other	10.677***	18.649***	-20.988***	-61.210***
	(0.000)	(0.000)	(0.000)	(0.000)
Constant	-76.389***	-67.529***	-3.227***	-45.699***
	(0.287)	(0.666)	(0.492)	(0.471)
Akaike Inf. Crit.	320.396	320.396	320.396	320.396

Note: *p<0.1 **p<0.05 ***p<0.01

Table 35. Results of propensity analysis in Section I.2 with program designation as the outcome, ninth grade cohort.

	Dependent variable:			
	Collaborative (1)	Consultation (2)	ENE (3)	Newcomer (4)
Gender = Male	0.417	-8.271	0.609*	-0.060
	(0.815)	(20.049)	(0.364)	(0.429)
Race = Asian	-30.541***	13.688***	0.161	0.033
	(0.771)	(0.00000)	(1.415)	(1.155)
Race = Black	-25.221***	29.141	-19.256***	-0.194
	(0.000)	(148.685)	(0.00000)	(0.749)
Race = Multiracial	-55.218***	0.878***	-54.321***	1.225
	(0.000)	(0.00000)	(0.000)	(2.187)
Race = Native American	2.555***	4.935***	-51.766***	-0.800
	(0.000)	(0.000)	(0.000)	(1.420)
Race = Pacific Islander	1.132	1.226	74.650	-3.248***
				(0.000)
Race = White	-101.140***	-1.409***	0.145	-0.597
	(0.000)	(0.000)	(0.845)	(1.119)
SIFE	-28.374***	22.710	-25.919***	4.685***
	(0.000)	(237.456)	(0.000)	(0.492)
Special Education	46.937***	59.343	1.557***	0.064
	(0.449)	(47.675)	(0.503)	(1.382)
Lunch Status = Paid	0.078	44.020	-0.021	-0.165
	(31.064)	(392.440)	(0.657)	(0.591)
Home Language = Missing	46.091***	-30.414***	-40.297***	-128.417***
	(0.569)	(0.000)	(0.000)	(0.000)
Home Language = Other	32.015***	-38.926***	0.022	0.991*
	(0.771)	(1.050)	(1.171)	(0.582)
Constant	-48.003***	-80.197	-2.303***	-3.892***
	(0.411)	(216.404)	(0.311)	(0.503)
Akaike Inf. Crit.	539.501	539.501	539.501	539.501

Note: *p<0.1 **p<0.05 ***p<0.01

Table 36. Results of propensity analysis in Section I.2 with whether a student is included in the Sheltered ESL program as the outcome.

	Dependent variable: Placed in Sheltered ESL			
(Cohort)	**(K)**	**(3)**	**(6)**	**(9)**
Gender = Male	0.414*	0.224	-0.817**	-0.260
	(0.231)	(0.181)	(0.412)	(0.261)
Race = Asian	0.355	-0.444	15.158	0.252
	(0.794)	(0.619)	(1,763.921)	(0.745)
Race = Black	1.239**	-0.551	0.480	0.622
	(0.484)	(0.614)	(1.829)	(0.516)
Race = Multiracial	-14.483	-14.996	15.852	0.114
	(835.726)	(716.733)	(2,437.681)	(1.370)
Race = Native American		17.230	16.245	1.000
		(1,010.371)	(6,522.639)	(1.334)
Race = Pacific Islander	15.914	-14.389		-16.626
	(1,455.398)	(1,455.398)		(882.743)
Race = White	-0.080	1.532**	17.428	0.472
	(0.472)	(0.630)	(1,690.517)	(0.690)
SIFE	16.793	2.192***	-2.685***	-2.750***
	(1,455.398)	(0.681)	(0.900)	(0.292)
Special Education	-0.800*	-1.090***	-3.116***	-2.139***
	(0.421)	(0.399)	(0.406)	(0.394)
Lunch Status = Paid	0.472	0.346	0.747	-0.004
	(0.373)	(0.296)	(0.847)	(0.429)
Home Language = Missing	0.007	0.587	0.139	-0.262
	(0.241)	(0.449)	(0.857)	(0.634)
Home Language = Other	0.300	0.727	1.487	-0.701
	(0.541)	(0.531)	(1.925)	(0.463)
Constant	-1.234***	-1.177***	3.138***	2.060***
	(0.221)	(0.141)	(0.430)	(0.242)
Observations	407	681	304	445
Log Likelihood	-230.076	-375.851	-89.668	-200.300
Akaike Inf. Crit.	484.153	777.702	203.336	426.599

Note: *p<0.1 **p<0.05 ***p<0.01

Table 37. Results of propensity analysis in Section I.3 with whether a student completed all exams as the outcome.

	Dependent variable:			
	Has Complete Case			
(Cohort)	(K)	(3)	(6)	(9)
Gender = Male	0.360	0.513***	0.996***	0.092
	(0.377)	(0.171)	(0.279)	(0.228)
Race = Asian	-0.592	-0.689	-1.351	-0.085
	(1.086)	(0.594)	(0.832)	(0.714)
Race = Black	0.129	1.235*	-0.675	0.021
	(0.821)	(0.661)	(0.667)	(0.498)
Race = Multiracial	16.728	-2.008	-0.148	2.685
	(3,559.441)	(1.269)	(0.879)	(1.711)
Race = Native American		-0.496	12.961	-0.087
		(1.502)	(882.743)	(1.284)
Race = Pacific Islander	16.463	12.709		-17.863
	(6,522.639)	(535.411)		(2,399.545)
Race = White	16.127	-0.754	-1.494**	-0.374
	(1,167.372)	(0.627)	(0.745)	(0.543)
SIFE	14.793	-0.622	-1.570*	-1.485***
	(6,522.639)	(0.610)	(0.866)	(0.482)
Special Education	-0.304	0.100	0.086	-0.025
	(0.564)	(0.306)	(0.428)	(0.455)
Lunch Status = Paid	-0.810	-0.784***	-0.809**	-0.358
	(0.561)	(0.284)	(0.402)	(0.394)
Home Language = Missing	1.220***	0.122	-0.309	-0.219
	(0.470)	(0.447)	(0.601)	(0.567)
Home Language = Other	-0.912	-0.187	0.600	-0.213
	(0.758)	(0.526)	(0.655)	(0.506)
Program = Collaborative	-0.185	-0.050	0.396	-0.711
	(0.603)	(0.306)	(0.770)	(0.814)
Program = Consultation	-3.883***	-1.476*	-0.413	-15.591
	(1.388)	(0.892)	(1.409)	(1,419.839)
Program = Developmental Bilingual	-0.501	0.587*		

	(0.566)	(0.332)		
Program = ENE	-0.793	0.425*	0.088	1.218***
	(1.210)	(0.254)	(0.551)	(0.414)
Program = Integrated ESl	0.064	0.493*		
	(0.605)	(0.281)		
Program = Transitional Bilingual	-0.371	0.703**		
	(0.558)	(0.279)		
Programs = Two Way Dual Language	15.657	0.060		
	(1,364.695)	(0.502)		
Program = Newcomer ESL			-0.176	-2.513***
			(1.560)	(0.825)
Absenteeism = Chronic	-0.846*	0.635***	0.165	-0.092
	(0.465)	(0.243)	(0.375)	(0.302)
Absenteeism = Excessive	-1.299	-0.252	-1.574**	-0.901*
	(0.930)	(0.685)	(0.725)	(0.502)
Absenteeism = Moderate	0.347	0.492**	-0.128	0.003
	(0.488)	(0.192)	(0.326)	(0.283)
Absenteeism = Mostly Absent				-15.938
				(854.427)
Constant	2.206***	-0.128	0.609**	0.078
	(0.545)	(0.211)	(0.252)	(0.260)
Observations	407	681	304	445
Log Likelihood	-108.072	-413.111	-167.677	-235.587
Akaike Inf. Crit.	260.144	872.222	373.353	513.174

Note: *p<0.1 **p<0.05 ***p<0.01