IES Institute of Education Sciences

Predicting Early Fall Student Enrollment in the School District of Philadelphia

Regional Educational Laboratory Mid-Atlantic

At Mathematica

REL 2022–124
U.S. DEPARTMENT OF EDUCATION

A Publication of the National Center for Education Evaluation and Regional Assistance at IES

Predicting Early Fall Student Enrollment in the School District of Philadelphia

Sean Tanner, Jenna Terrell, Emily Vislosky, Jonathan Gellar, and Brian Gill

October 2021

Predicting incoming enrollment is an ongoing concern for the School District of Philadelphia (SDP) and similar districts with school choice systems, substantial student mobility, or both. Inaccurate predictions can disrupt learning as districts adjust to enrollment fluctuations by reshuffling teachers and students well into the fall semester. This study compared the accuracy of four statistical techniques for predicting fall enrollment at the school-by-grade level, using data from prior years, to assess which approach might be the most useful for planning school staffing in SDP. The predictions differ little in accuracy: predicted cohort size differs from actual cohort size by roughly six students across all methods The statistical techniques leave much student mobility unaccounted for. Even under the best prediction approach, students and teachers in 22 percent of incoming grade levels within schools might have to be reassigned because of unexpected student mobility and district rules on maximum class size. Predictive accuracy is not meaningfully different in schools with larger proportions of Black students, economically disadvantaged students, or English learner students. Of the 259 predictors analyzed, 4 stand out as the most important: prior cohort sizes, in-school suspensions, out-of-school suspensions, and absences.

Why this study?

Fluctuations in enrollment are a perennial issue for school districts (Hussar & Bailey, 2017). Unpredictable enrollment patterns can bedevil school systems when a mismatch between predicted and actual cohort sizes (defined in this study as school-by-year-by-grade units) in the fall disrupts teacher and student classroom assignment. For example, in a district with a class size limit of 30 students, a school that planned for 100 students entering a particular grade but has only 90 students actually attend in the fall would likely have to cut back from four to three classrooms, reassigning the fourth teacher to another grade or school and the students in that teacher's would-be class to the other three classrooms.

The core problem for most schools is that successive incoming cohorts of students differ in size (Sweeney & Middleton, 2005). Beyond changes in initial cohort sizes, many schools could face further variation in enrollment because of school choice systems (which allow students to change schools without a residential move) and student mobility, or both. In 2016, 112 districts serving 13 million students in 24,000 schools across the country had a choice system (Whitehurst, 2017). The Regional Educational Laboratory (REL) Mid-Atlantic region is home to at least nine of these districts: Washington, DC; Camden, New Jersey; Newark, New Jersey; Philadelphia, Pennsylvania; and five districts in Maryland[1] (Whitehurst, 2017). These districts serve more than 1 million students in 2,000 schools.

In the School District of Philadelphia (SDP), many schools experience such disruption (Schmitt, 2017). SDP reports that in the spring many students plan to attend their neighborhood public school in the fall but in the intervening months choose to attend a different school, be it a charter or private school. The district allocates resources in March for the upcoming school year and then reallocates resources as needed after October 1, once it has accounted for discrepancies between prospective spring enrollment

For additional information, including technical methods and supporting analyses, see the appendices for the report on the Institute of Education Sciences website.

1. The five districts in Maryland are Anne Arundel County, Baltimore City, Baltimore County, Montgomery County, and Prince George's County Public Schools.

and actual attendance in the fall (a process SDP refers to as leveling). This reallocation can entail eliminating classrooms, reassigning students to a different teacher, and reassigning teachers to a different school or grade level. Because of this late sorting, many students experience changes in classroom peers and teachers in October, which can have negative impacts on classroom engagement (Isernhagen & Bulkin, 2009) and performance on end-of-year standardized tests (Gibbons & Telhaj, 2011; Henry & Redding, 2020; Whitesell et al., 2016). Though some students move into neighborhood schools between the spring and the fall, cohorts in SDP neighborhood schools experience a net loss of students progressively from grade 1 to grade 8, such that each cohort is smaller in grade 8 than when it began in kindergarten, as students move to charter schools, repeat grades, or leave the district entirely.

SDP tries to account for such attrition, as well as new entrants to the school system, when allocating teachers to schools and grades. In recent years it has modeled enrollment based on broad trends from prior cohorts, the proximity of alternative schools to students' neighborhood schools, and other factors such as nearby school closures. The models are not formally fit on prior data but are based on assumptions about future behavior and the size of the cohorts that prospectively enroll in the spring.[2] SDP reports that unpredicted fluctuations still disrupt classroom stability. This is largely because the district tries to allocate teaching staff—its largest cost—so that the number of students per class is as close to the class size limit of 30 without going over. With more accurate predictions of cohort sizes, SDP could reduce the disruptions caused by attrition and (to a lesser extent) in-migration of new students.

Machine learning—a systematic approach to statistical prediction—might enhance predictions of student enrollment, which would enable SDP to prospectively allocate resources in a manner that results in less disruption in October. Traditional statistical methods such as ordinary least squares (OLS) regression and researcher practices such as sequentially testing increasing numbers of predictors might fit sample data well but perform less effectively on novel data (a phenomenon known as overfitting). Machine learning combines more recent statistical prediction tools with safeguards against overfitting in an attempt to yield more stable, accurate predictions (Mullainathan & Spiess, 2017). This report uses the term algorithm to refer to statistical methods of data analysis. A prediction algorithm can be either a simple regression or a more complex model that relies on machine learning.

Many states use complex administrative data and machine learning algorithms to improve services for segments of their populations. For example, in Connecticut and Missouri predictive algorithms identify high-risk people who use certain medical services more than others and match them with intensive care management teams (Bergh et al., 2018). Applying machine learning in predicting criminal activity has substantially increased in the past decade, to predict both who might be involved in the criminal activity and where criminal activity might take place (Berk, 2017; Chandler et al., 2011; Gerber, 2014; McClendon & Meghanathan, 2015; Rhodes, 2013; Rockoff et al., 2008; Toppireddy et al., 2018).

In education, researchers have applied machine learning more recently to estimate individuals' likelihood of achieving key outcomes of interest (Porter & Balu, 2016). In informing district determinations of which teachers to retain, Chalfin et al. (2016) compared the performance of machine learning with that of a traditional teacher value-added model in predicting student gains in the final year of a study that randomly assigned teachers to students. They found that using machine learning rather than value-added measures to identify the bottom 10 percent of teachers and replacing them with average quality teachers would increase student learning gains by 0.02 standard deviation in math and 0.01 standard deviation in language arts in the school system as a whole (the gains were an order of magnitude larger for directly affected students). Though these effects are small, the

2. The informal judgments involved in the models, coupled with staff turnover at SDP, made it impossible to replicate this approach so the study team included a simple OLS regression with a basic set of predictors as a comparison to the more sophisticated machine learning algorithms.

cost of using machine learning algorithms instead of traditional value-added modeling is likely trivial (though the authors do not provide a cost estimate), so switching to these algorithms might be cost-effective. A recent REL Mid-Atlantic study used machine learning to generate accurate predictions of academic performance for elementary, middle, and high school students using administrative data from education and child welfare agencies (Bruch et al., 2020). In short, machine learning algorithms appear to be a promising strategy for districts deciding about resource allocation, but it remains unclear whether they will markedly improve on traditional statistical prediction methods and other current forecasting practices.

The current study explored how machine learning algorithms might improve predictions of incoming cohort size in SDP. More accurate predictions could help SDP better anticipate incoming cohort sizes and, consequently, reduce instability among students and staff. If the new methods of projecting enrollment prove highly accurate, other districts in the REL Mid-Atlantic region and across the country might consider similar strategies to manage the burden that a choice system places on enrollment stability.

Research questions

This study addressed one primary research question and three subquestions:

1. How well do machine learning algorithms predict incoming cohort sizes for grade 1–8 in SDP's neighborhood schools?

    1a. How well does each algorithm predict incoming cohort sizes for the following fall semester?

    1b. Does the precision of each algorithm differ for cohorts with a larger proportion of Black students, economically disadvantaged students, or English learner students compared with cohorts overall?

    1c. Which administrative variables contribute most to the predictions?

The first two research tasks proceeded along similar lines. First, the study team built prediction models by applying four prediction algorithms—OLS, least absolute shrinkage and selection operator (LASSO), elastic net, and random forest—to retrospective data from the 2016/17 and 2017/18 school years. Those models were then used to predict incoming grade sizes in the 2018/19 school year. Performance was assessed by comparing predicted grade sizes to actual grade sizes. Box 1 summarizes the data sources, sample, and methods, and appendix A provides additional details.

Box 1. Data sources, sample, and methods

Data sources. The study used student-level administrative data provided by the School District of Philadelphia (SDP) for students enrolled in SDP neighborhood K–8 schools during the 2015/16–2018/19 school years. The data included attendance records, test score performance, demographic information, and geographic distance from neighborhood and alternative schools as well as a unique identifier to track students over time and across data files.

Sample. The study used data on students who were in grade 1–8 in 174 SDP neighborhood schools in any school year from 2015/16 to 2018/19. Each year had about 1,000 grade-by-year units. The analytic sample contained 149,154 unique students, 50 percent of whom were Black, 21 percent of whom were Hispanic, 13 percent of whom were English learners, 14 percent of whom were White, and 61 percent of whom were economically disadvantaged (see table A3 in appendix A). These proportions are similar to those for SDP as a whole but differ markedly from those for the overall U.S. student population.

Methodology. The analysis used observed patterns between grade-level enrollments and school-level administrative data in one period (the base year) to predict grade-level enrollments in a subsequent period (the target year) in each SDP school that enrolled

students in grades K–8. The predictor data are from February of the base year or of prior years (for lagged data), as that is the only information available to the district when it allocates staff for the target year. The outcomes are from fall of the target year, when grade sizes are realized. Six categories of data were used as predictors:

- *Grade size.* Although incoming grade size in the target year was the outcome variable, current and past grade sizes for the same cohort of students (in earlier grades) and other cohorts of students were included as predictors. The change in the current cohort's size from the previous grade to the current grade—for example, the number of students in grade 4 in the base year compared with the number of students in grade 3 in the prior year—and the change in grade size in a given school and grade from the prior year to the base year—for example, the number of students in grade 4 in the base year compared with the number of students in grade 4 in the prior year—were included. Additionally, the lagged attrition count for the current cohort in the target school-by-grade combination—for example, to predict the incoming cohort size in grade 4 in a given school in 2017, the number of students lost from grade 3 to grade 4 in that school for the current (2016) grade 4 cohort—was included.
- *Attendance and suspensions.* Attendance was calculated as the total number of absences across all students in a given year and grade. The study team created three indicator variables at the student level: 0–5 absences, 6–10 absences, and more than 10 absences.[1] The cohort-level aggregate versions of those indicators are the number of students in each category and were included as predictors along with total absences at the cohort level. For in-school and out-of-school suspensions the counts of students in the following categories for each indicator were included: less than 3 days, 3–5 days, and more than 5 days.
- *Test scores.* Student-level data included three measures of English language arts—the Reading Curriculum-Based Measurement, Nonsense Word Fluency, and Oral Reading Fluency—in grade 1 and 2 and English language arts and math performance on the Pennsylvania System of School Assessment in grade 3 and above. The study team calculated within-district percentile rankings by grade across all years for each student for each subject and test and then divided the percentile ranks into three equal categories (a process known as trichotomization): under 30, 30–60, and above 60. Missing data, including structurally missing scores for students in grades that do not receive a particular test, are not counted in the three categories. The counts of students in each performance category and the counts of students with missing values in each test at the cohort level were aggregated and included as predictors. Trichotomization provides more information on the distribution of achievement within cohorts than would be available with performance averages, especially with modest amounts of missing data.
- *Demographics.* Aggregate counts of students in each cohort for the following demographic variables were included as predictors: Asian, Black, Hispanic, Native Hawaiian/Pacific islander, White, Multiracial, unknown race/ethnicity, male, female, economically disadvantaged, English learner students, and students with an Individualized Education Program.
- *Geographic data.* The study team calculated the distances from students' home address to their neighborhood school and the nearest charter school. Each continuous distance measure was trichotomized into categories of less than 0.5 mile, 0.5–1 mile, and more than 1 mile. The counts of students in each category for each measure and the count of students with missing data at the cohort level were included as predictors. As with test scores, the desire for distributional information and the large number of students with missing addresses led to the use of trichotomous variables.
- *Structure.* The study team created indicator variables for each grade level and school.

The study used four prediction algorithms:

- *Ordinary least squares (OLS).* A widely used statistical model that serves as the baseline to which more sophisticated models are compared.
- *Least absolute shrinkage and selection operator (LASSO).* An extension to OLS that constrains the influence of weak predictors. The constraints might produce more accurate predictions of incoming cohort sizes by focusing on the strongest predictors.
- *Elastic net.* An algorithm that produces a weighted average of predictions from LASSO and an alternate set of constraints. The blending of constraints might enhance predictions of incoming cohort sizes if LASSO discards too many predictors.
- *Random forest.* A statistical routine that iteratively tests partitions of the data until predictions of subgroup outcomes can no longer be improved on. By selecting predictors that perform well across many subsets of the data, this algorithm might produce more reliable predictions of incoming cohort sizes.

Two common metrics (the median absolute deviation between predicted and actual cohort sizes and the root mean squared error) and one study-specific metric (the percentage of school-by-grade cohorts subject to reallocation) were used to evaluate the performance of the prediction algorithms. The median absolute deviation measures the number of students by which a prediction deviates (in a positive or negative direction) from the actual number of students in a typical target grade. The root mean squared error is the square root of the sum of squared differences between the predicted and actual number of students in target

grades; it is more sensitive to outliers, such as those that arise when predicted and observed grade sizes are markedly different. The percentage of school-by-grade cohorts subject to reallocation was based on the number of grades for which predicted and actual student counts were on opposite sides of a multiple of 30—the nominal limit on class size in SDP (under the assumption that schools are allocated the minimum number of classes to serve each grade without allowing class size to exceed 30). This is a measure of how often the discrepancy between predictions and realized data is likely to cause disruptive reallocation of teachers and students in the fall. However, this reallocation rule might not reflect actual district practices. Accordingly, reallocation patterns were also calculated with the alternate assumption that the district must abide by the class size limit of 30 in the springtime allocation of teachers but will tolerate observed class sizes of up to 35 in the fall. This alternative rule still assumes the district will use as few teachers as possible.

The accuracy measures were calculated for the analytic sample as a whole, as well as for cohorts in which the proportion of Black students is above the median for SDP, for cohorts in which the proportion of economically disadvantaged students is above the median, and for cohorts in which the proportion of English learner students is greater than 10 percent. The study team also investigated which variables provide the most information on incoming cohort size.

Before running each algorithm, the target years were partitioned into the modeling set (target years 2016/17 and 2017/18) and the extrapolation set (target year 2018/19).

Machine learning algorithms are built to maximize predictive performance on a randomly excluded testing set of data, but any use in the real world requires that predictions be extrapolated to new datasets in which the outcome has not yet been recorded. The 2018/19 data served as that dataset in this case and allowed for testing the generalizability of prior years' models to future cohort sizes. For transparency, accuracy measures are presented for the testing set and the extrapolation set, although the study team believes that predictive performance in the extrapolation set is the better indication of future performance.

Note

1. These thresholds were chosen because they are common in other datasets that the study team has worked with.

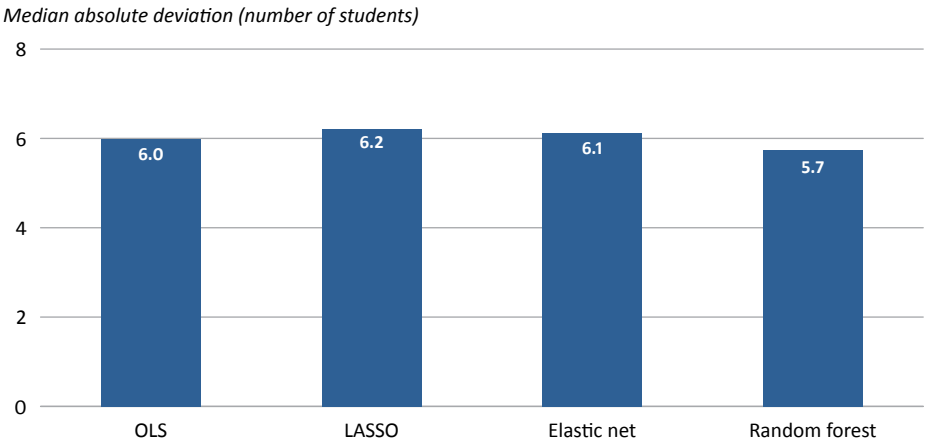
Findings

All four algorithms have similar predictive accuracy, with the random forest algorithm slightly outperforming the others

The median absolute deviation for each algorithm represents how close the predicted cohort sizes (in number of students) were to the observed cohort sizes in a typical cohort. The range of median absolute deviations is small, from 5.74 students per school by grade to 6.21 (figure 1 and table 1). This means that all four algorithms generated predictions that were within about six students of the observed cohort sizes. The random forest algorithm has the lowest median absolute deviation, 5.74, which is roughly 10 percent of the median cohort (school-by-grade) size in the extrapolation year (2018/19).

To assess the real-world consequences of error in predictions, the median absolute deviation of each algorithm was converted to the percentage of cohorts that would be subject to reallocation in the fall if the district relied on the algorithm for planning in the spring. Reallocation disrupts learning and is a chief concern for the district when predicting incoming cohort sizes. The percentage of cohorts that would be subject to reallocation as a result of prediction error is 22 percent for the random forest algorithm, 23 percent for the LASSO and elastic net algorithms, and 29 percent for the OLS algorithm (table 2). This means that between a fifth and a quarter of cohorts would have to be reshuffled in October—for example, being split from two classes into three or condensed from three classes to two.

Figure 1. Predictive accuracy for student enrollment in K–8 neighborhood schools in the School District of Philadelphia was similar across all four algorithms analyzed



OLS is ordinary least squares. LASSO is least absolute shrinkage and selection operator.

Note: Values are from the extrapolation set (2018/19 school year).

Source: Authors' calculations based on administrative data for 2015–19 provided by the School District of Philadelphia.

Table 1. Predictive accuracy for incoming cohort sizes in School District of Philadelphia neighborhood schools, by algorithm

Algorithm	Median absolute deviation (number of students)
OLS	5.99
LASSO	6.21
Elastic net	6.12
Random forest	5.74

OLS is ordinary least squares. LASSO is least absolute shrinkage and selection operator.

Note: Figures are from the extrapolation set (2018/19 school year).

Source: Authors' calculations based on administrative data for 2015–19 provided by the School District of Philadelphia.

Table 2. Percentage of school-by-grade cohorts subject to reallocation, by algorithm

Algorithm	Percent
OLS	29
LASSO	23
Elastic net	23
Random forest	22

OLS is ordinary least squares. LASSO is least absolute shrinkage and selection operator.

Note: Figures are from the extrapolation set (2018/19 school year).

Source: Authors' calculations based on administrative data for 2015–19 provided by the School District of Philadelphia.

Performance does not differ across cohort demographic composition

The random forest algorithm, highlighted here because it slightly outperformed the other algorithms, produces predictions with similar accuracy for cohorts in which the proportion of Black students is above the median for SDP, for cohorts in which the proportion of economically disadvantaged students is above the median, and for cohorts in which the proportion of English learner students is greater than 10 percent. The percentage of those cohorts that would be subject to reallocation ranges from 21 percent to 25 percent (table 3; see tables B4, B6, and B8 in appendix B for results from all four algorithms).

The alternate reallocation rule, which tolerates class sizes of up to 35, reduces disruptions by roughly half for three of the four models. About 11 percent of cohorts are subject to reallocation under the random forest predictions, and 12 percent are under the LASSO and elastic net predictions (table 4). This provides a likely lower bound on reallocation. It is improbable that the district would allow class sizes of 35 for schools with six or seven classes per grade, because there would be enough students above the 30 student cutoff in those classes to fill another class.

There are not clear patterns by grade level (table 5).[3]

Table 3. Percentage of school-by-grade cohorts subject to reallocation under the random forest algorithm, by cohort demographic

Cohort demographic	Percent
Proportion of Black students is above the median for SDP	21
Proportion of economically disadvantaged students is above the median for SDP	22
Proportion of English learner students is greater than 10 percent	25

SDP is School District of Philadelphia.

Note: Figures are from the extrapolation set (2018/19 school year).

Source: Authors' calculations based on administrative data for 2015–19 provided by the School District of Philadelphia.

Table 4. Percentage of school-by-grade cohorts subject to reallocation with higher class size tolerance, by algorithm

Algorithm	Percent
OLS	17
LASSO	12
Elastic net	12
Random forest	11

OLS is ordinary least squares. LASSO is least absolute shrinkage and selection operator.

Note: Figures are from the extrapolation set (2018/19 school year).

Source: Authors' calculations based on administrative data for 2015–19 provided by the School District of Philadelphia.

3. Because of the use of lagged information on grade sizes and cohort, as well as information on past patterns of attrition from cohorts, predictions can be made only for grade 2–7 rather than for grade 1–8.

Table 5. Percentage of school-by-grade cohorts subject to reallocation under the random forest algorithm, by grade level

Grade level	Percent
Grade 2	30
Grade 3	23
Grade 4	28
Grade 5	19
Grade 6	32
Grade 7	27

Note: Figures are from the extrapolation set (2018/19 school year).

Source: Authors' calculations based on administrative data for 2015–19 provided by the School District of Philadelphia.

Four predictors provide the most meaningful contribution to accurately predicting school-by-grade enrollment

Of the 259 predictors assessed in the models, only a handful meaningfully contribute to predictive accuracy. The random forest algorithm, again highlighted here because it slightly outperformed the other approaches, produces an importance score on a scale of 0 (least important) to 1 (most important) for each variable. The following discussion relies on these scores.

Not surprisingly, the number of students in a grade in one year is the best predictor of the number of students in the same grade of the same school the following year (see table 6 for the top four predictors). The three next best predictors are the number of students with more than five in-school suspensions, the number of students with more than five out-of-school suspensions, and the number of students with fewer than six absences. There is a large drop in variable importance scores after the first four predictors. Many of the predictors ranked 5th–15th are indicators of missingness on test scores (see table B10 in appendix B). These missing test indicators are likely picking up signal from grade-level indicators because much of the missingness is structural due to different standardized tests across grades. Beyond what can be explained by missing test scores, individual grade-level indicators have low importance scores (near .01 or lower). The 15th ranked predictor, with an importance score of .10, is the base-year enrollment in the grade before the prediction grade (that is, the spring enrollment of the same cohort of students whose fall enrollment is being predicted). In other words, a school's current enrollment for each grade (e.g., the number of grade 5 students this year) is a better predictor of enrollment size in that same grade the following year (e.g. the number of grade 5 students next year) than is the current size of the cohort that will enter that grade (e.g., the number of grade 4 students this year).

Table 6. Top four predictors of school-by-grade enrollment under the random forest algorithm

Predictive strength rank	Predictor	Predictor importance	Data category
1	Base-year enrollment (prior cohort) in same grade as prediction year	.97	Grade size
2	Number of students with more than five in-school suspensions	.91	Attendance and suspensions
3	Number of students with more than five out-of-school suspensions	.83	Attendance and suspensions
4	Number of students with fewer than six absences	.65	Attendance and suspensions

Source: Authors' calculations based on administrative data for 2015–19 provided by the School District of Philadelphia.

Limitations

This study has two main limitations. First, the precise procedures, reallocation rates, and prediction errors from the district's prior method of forecasting enrollment were not available for comparison. Without those benchmarks, it is unclear whether the use of a random forest algorithm (or any of the other three algorithms) would meaningfully reduce disruptive reallocation of students and teachers in the fall semester.

Second, and more important, the results might not be generalizable to future years. Though the analysis included an additional extrapolation exercise to explicitly account for changes in the structure of attrition and new entrants across years, the COVID-19 pandemic might have fundamentally altered patterns of attrition and new entrants in a way that models based on historic data are unable to capture. Even so, in such a scenario, this analysis is still useful for two reasons. First, although COVID-19 might change the association of incoming cohort sizes with school-level variables, the ordinal performance rank of algorithms might not change, in which case the approach with the highest predictive accuracy in this analysis would also be approach with the highest predictive accuracy in a post-COVID-19 environment. Second, even if the ordinal performance rank of algorithms changes because of COVID-19, the statistical code produced here is adaptable to future years so that the district can rerun the analysis to choose the best algorithm in future cohorts.

Implications

The study findings yield four main implications. First, if the goal is to increase predictive accuracy at the typical school, the choice of algorithm does not appear to be consequential. With the data used in this analysis, the algorithms had similar median absolute deviations in the extrapolation sets (and the test sets), with the random forest algorithm having a slightly lower median absolute deviation in the extrapolation set than the other methods. However, the difference in prediction error between the random forest algorithm and the other three algorithms is less than half a student. Because the median absolute deviation reflects error at the typical school, the similarity of performance across algorithms means that SDP could choose any of the four and achieve similar results. The root mean squared errors did not markedly differ in the extrapolation set either. These results suggest that if the district is concerned that machine learning algorithms are difficult to use and understand, it could rely on OLS with little loss in predictive accuracy.

Second, the performance of the algorithms does not raise equity concerns. Predictive accuracy does not dramatically differ across the demographic makeup of cohorts. Incoming cohorts in which the proportion of Black students or the proportion of economically disadvantaged students is above the median for SDP would not have been subject to different reallocation rates than cohorts in general (21–22 percent). Cohorts in which the proportion of English learner students is greater than 10 percent would have been subject to slightly more reallocation than cohorts as a whole—25 percent compared with 22 percent. This additional 3 percentage points translates into approximately 15 cohorts districtwide. Using predictions of incoming cohort sizes derived from these algorithms is unlikely to have inequitable effects on learning disruptions in the fall.

Third, large sets of variables are not needed to predict cohort sizes. Beyond a handful of variables, few predictors taken from the administrative data were helpful for predicting incoming cohort size. The nature of attrition and new entrants might frustrate attempts at systematic prediction, or the readily available administrative data used in this analysis might not have captured the underlying indicators of attrition and new entrants. Alternately, data on charter school applications or intent to remain in the district might provide much stronger predictors of incoming cohort sizes.

Fourth, SDP will need better predictors of incoming cohort sizes to successfully reduce reallocation in the fall. The data and algorithms used in this study do not eliminate reallocation. Overall, the best performing algorithm—random

forest—would have still exposed between a tenth and a quarter of cohorts to disruptive reallocation of teachers and students, depending on the district's tolerance for maximum class size. Using data from additional prior years is unlikely to improve predictions, as adding information on cohort sizes from two years prior did not meaningfully improve accuracy. Regardless of the methods used, improved accuracy is likely to require additional predictors that include stronger signals of incoming cohort sizes, data that might be available by June each year.

Despite these modest results, there are two reasons for optimism about the ability of algorithms, be they OLS or more sophisticated machine learning ones, to help districts accurately predict enrollment. First, SDP obtains data (such as registration forms) from February to July that are likely to be highly indicative of new entrants and attrition. If accurate predictions of the late August cohort sizes can be made in July, the district might be able to substantially reduce reallocation of students in the fall. Teachers would likely still have to be reallocated from the target year positions assigned to them in the base year, but such staff reshuffling would be far less disruptive in August than in October. Second, indicators of attrition might be obtainable through enrollment intention surveys that districts could field in the early spring. In a district such as SDP, with a large school choice system, most attrition is likely to be the result of a deliberate process of shopping for schools. Although new entrants to the district would not be addressed by such enrollment intention surveys, the surveys might predict attrition well enough to warrant further exploration of their use. Combining survey data with administrative records in the late spring and summer might substantially improve predictions of incoming fall cohort sizes.

References

Bergh, S., Davis, A., Ivey, A., Kitson, D., & Thornton, J. (2018). *How states use data to inform decisions*. Pew Charitable Trusts. https://www.pewtrusts.org/-/media/assets/2018/02/dasa_how_states_use_data_report_v5.pdf.

Berk, R. (2017). An impact assessment of machine learning risk forecasts on parole board decisions and recidivism. *Journal of Experimental Criminology*, *13*(1), 193–216. https://doi.org/10.1007/s11292-017-9286-2.

Bruch, J., Gellar, J., Cattell, L., Hotchkiss, J., & Killewald, P. (2020). *Using data from schools and child welfare agencies to predict near-term academic risks* (REL 2020–027). U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Mid-Atlantic.

Chalfin, A., Danieli, O., Hillis, A., Jelveh, Z., Luca, M., Ludwig, J., & Mullainathan, S. (2016). Productivity and selection of human capital with machine learning. *American Economic Review*, *106*(5), 124–127.

Chandler, B. D., Levitt, S. D., & List, J. A. (2011). Predicting and preventing shootings among at-risk youth. *American Economic Review, 101*(3), 288–292.

Gerber, M. S. (2014). Predicting crime using Twitter and kernel density estimation. *Decision Support Systems*, *61*(1), 115–125. https://doi.org/10.1016/j.dss.2014.02.003.

Gibbons, S., & Telhaj, S. (2011). Pupil mobility and school disruption. *Journal of Public Economics*, *95*(9–10), 1156–1167.

Henry, G. T., & Redding, C. (2020). The consequences of leaving school early: The effects of within-year and end-of-year teacher turnover. *Education Finance and Policy*, *15*(2), 332–356. https://www.mitpressjournals.org/doi/abs/10.1162/edfp_a_00274.

Hussar, W. J., & Bailey, T. M. (2017). *Projections of education statistics to 2025* (NCES No. 2017–019). U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics. http://eric.ed.gov/?id=ED576296.

Isernhagen, J. C., & Bulkin, N. (2009). The impact of mobility on student performance and teacher practice. *Journal of At-Risk Practices*, *16*(1), 17–24.

McClendon, L., & Meghanathan, N. (2015). Using machine learning algorithms to analyze crime data. *Machine Learning and Applications: An International Journal*, *2*(1), 1–12. https://doi.org/10.5121/mlaij.2015.2101.

Mullainathan, S., & Spiess, J. (2017). Machine learning: An applied econometric approach. *Journal of Economic Perspectives*, *31*(2), 87–106.

Porter, K. E., & Balu, R. (2016). *Predictive modeling of K-12 academic outcomes: A primer for researchers working with education data*. MDRC.

Rhodes, W. (2013). Machine learning approaches as a tool for effective offender risk prediction. *Criminology and Public Policy*, *12*(3), 507–510.

Rockoff, J., Jacob, B., Kane, T., & Staiger, D. (2008). Can you recognize an effective teacher when you recruit one? (Working Paper No. 14485). National Bureau of Economic Research.

Schmitt, M. (2017). *Getting into high school in Philadelphia: The workings of a complicated system*. Pew Charitable Trusts.

Sweeney, S. H., & Middleton, E. J. (2005). Multiregional cohort enrolment projections: Matching methods to enrolment policies. *Population, Space and Place*, *11*(5), 361–380.

Toppireddy, H. K. R., Saini, B., & Mahajan, G. (2018). Crime prediction & monitoring framework based on spatial analysis. *Procedia Computer Science*, *132*(1), 696–705. https://doi.org/10.1016/j.procs.2018.05.075.

Whitehurst, G. J. R. (2017). *Education choice and competition index 2016: Summary and commentary*. Center on Children and Families, Brookings Institution.

Whitesell, E. R., Stiefel, L., & Schwartz, A. E. (2016). Unexpected arrivals: The spillover effects of mid-year entry on stable student achievement in New York City. *Educational Evaluation and Policy Analysis*, *38*(4), 692–713. https://doi.org/10.3102/0162373716661242.

REL 2022–124

October 2021

This report was prepared for the Institute of Education Sciences (IES) under Contract ED-IES-17-C-0006 by the Regional Educational Laboratory Mid-Atlantic administered by Mathematica. The content of the publication does not necessarily reflect the views or policies of IES or the U.S. Department of Education, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government.

This REL report is in the public domain. While permission to reprint this publication is not necessary, it should be cited as:

Tanner, S., Terrell, J., Vislosky, E., Gellar, J., & Gill, B. (2021). *Predicting early fall student enrollment in the School District of Philadelphia* (REL 2022–124). U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Mid-Atlantic. https://ies.ed.gov/use-work/resource-library/reports/

This report is available on the Regional Educational Laboratory website at https://ies.ed.gov/use-work/resource-library/reports/.