Institute of Education Sciences

Using Data from Schools and Child Welfare Agencies to Predict Near-Term Academic Risks

Regional Educational Laboratory Mid-Atlantic

At Mathematica Policy Research

REL 2020–027
U.S. DEPARTMENT OF EDUCATION

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

Using Data from Schools and Child Welfare Agencies to Predict Near-Term Academic Risks

Julie Bruch, Jonathan Gellar, Lindsay Cattell, John Hotchkiss, and Phil Killewald **July 2020**

This report provides information for administrators, researchers, and student support staff in local education agencies who are interested in identifying students who are likely to have near-term academic problems such as absenteeism, suspensions, poor grades, and low performance on state tests. The report describes an approach for developing a predictive model and assesses how well the model identifies at-risk students using data from two local education agencies in Allegheny County, Pennsylvania: a large local education agency and a smaller charter school network. It also examines which types of predictors—in-school variables (performance, behavior, and consequences) and out-of-school variables (human services involvement and public benefit receipt)—are individually related to each type of near-term academic problem to better understand why the model might flag students as at risk and how best to support these students.

The study finds that predictive models using machine learning algorithms identify at-risk students with moderate to high accuracy. In-school variables drawing on school data are the strongest predictors across all outcomes, and predictive performance is not reduced much when out-of-school variables drawing on human services data are excluded and only school data are used. However, some out-of-school events and services—including child welfare involvement, emergency homeless services, and juvenile justice system involvement—are individually related to near-term academic problems. The models are more accurate for the large local education agency than for the smaller charter school network. The models are better at predicting low grade point average, course failure, and scores below the basic level on state tests in grades 3–8 than at predicting chronic absenteeism, suspensions, and scores below the basic level on high school end-of-course standardized tests. The findings suggest that many local education agencies could apply machine learning algorithms to existing school data to identify students who are at risk of near-term academic problems that are known to be precursors to school dropout.

Why this study?

Many school districts use early warning systems to identify students who are at risk of dropping out of high school. In the 2014/15 school year more than half of high schools used some kind of early warning system (U.S. Department of Education, 2016). Many of the systems track attendance, behavior, and course performance, which research has shown reliably identify at-risk students (Allensworth, Gwynne, Moore, & de la Torre, 2014; Balfanz, Herzog, & Mac Iver, 2007; Bowers, Sprott, & Taff, 2013). Educators can then target resources to the most at-risk students and intervene before students drop out (Bruce, Bridgeland, Hornig Fox, & Balfanz, 2011; Edmunds, Willse, Arshavsky, & Dallas, 2013).

Attendance, behavior, and course performance problems are widespread in Allegheny County. For example, a third of Pittsburgh Public Schools (PPS) students missed at least 10 percent of school days in 2018/19 (Pittsburgh Public Schools, 2020). In Pittsburgh, as in other communities, chronic absenteeism is associated with academic problems. In PPS chronic absenteeism is especially high among students receiving public benefits or mental health services and among students involved in the child welfare system. Nearly half of students in out-of-home child welfare placements were chronically absent in 2011/12 (Allegheny County Department of Human Services, 2015).

For additional information, including technical methods and supporting analyses, access the report appendices at https://go.usa.gov/xwGSq.

Out-of-school risk factors, such as homelessness, justice system involvement, and involvement with child welfare services, are not currently incorporated into early warning systems in Allegheny County or in most other school districts across the country. Among schools that used early warning systems in 2014/15, fewer than half included data on such out-of-school factors (U.S. Department of Education, 2016). Data on out-of-school factors in students' lives are typically used to provide context when determining appropriate interventions rather than as predictors of academic problems (National Forum on Education Statistics, 2018) despite the fact that homelessness, teenage pregnancy, child abuse, parental substance abuse, and unsafe living conditions are risk factors for negative academic outcomes such as absenteeism (Kearney, 2008).

PPS, Propel Schools (a charter school network in Allegheny County), and the Allegheny County Department of Human Services (DHS) requested a study to assess how well combinations of in-school and out-of-school longitudinal, event-level variables predict near-term academic problems. Stakeholders in Allegheny County can take advantage of richer data than are typically available to schools thanks to a unique data-sharing agreement between the Allegheny County DHS and local education agencies. DHS data are linked to school data and tracked at the daily or monthly level. This study used these data to assess the extent to which including out-of-school variables improved predictions of near-term academic problems.

PPS is the second-largest school district in Pennsylvania, operating 54 schools that serve a socioeconomically diverse population of about 24,000 students. Propel Schools is a public charter school network in Pittsburgh and surrounding communities, primarily in low-income neighborhoods. Established in 2003, the Propel Schools network has 13 locations and serves around 4,000 students, most of whom are socioeconomically disadvantaged. (See table A3 in appendix A for a demographic profile of students enrolled in PPS and Propel Schools during the study period.)

In recent years PPS has developed a positive behavioral interventions and support dashboard and a suspensions dashboard, which include attendance warning flags and some achievement measures. Propel has a student assistance program and a multitier system of support in place that uses attendance and behavior data to identify students for support. While these tools provide some information to educators about students' prior academic problems, they are not comprehensive systems for predicting the likelihood that specific types of academic problems will occur in the coming months.

Research questions

The goal of the study was to develop an approach to predict, on a periodic basis throughout a school year, whether students in any grade level are at risk for academic problems in the coming months (referred to as the outcome period; see box 1 for definitions of key terms). The study answered the following research questions:

1. Which types of in-school performance, behavior, and consequences are related to academic problems for students in kindergarten–grade 12? Which types of out-of-school human services involvement and public benefit receipt are related to academic problems for students in kindergarten–grade 12?

2. How well do combinations of in-school and out-of-school variables predict academic problems, and how does including out-of-school predictors affect model performance?

The study used 2014/15–2016/17 data on student demographics, enrollment, courses, state tests, attendance, and behavior from PPS and Propel Schools. It also used Allegheny County DHS data on student use of social services, justice system involvement, and public benefits.

Research question 1, the descriptive analysis, identifies patterns in the individual relationships between previously measured in-school performance, behavior, and consequences and out-of-school human services involvement

Box 1. Key terms

Academic outcomes. The near-term academic problems identified by the predictors in the preceding adjacent period. The negative academic outcomes—or academic problems—included in the study are chronic absenteeism, any suspension, course failure, low grade point average, and a score below the basic level on state tests.

Area under the receiver operating characteristic (ROC) curve. A single-number summary of the ROC curve (see below), which can take a value from 0 to 1, where 1 is perfect prediction and .5 is the performance of a "coin flip." The area under the receiver operating characteristic curve can be interpreted as the probability that the model will consider a randomly selected student with an academic problem at higher risk than a randomly selected student without an academic problem. Generally, values above .7 indicate a strong model fit in social science (Rice & Harris, 2005).

False positive rate. Probability that the model will incorrectly predict that a student without an academic problem will have such a problem.

Machine learning algorithms. A broad class of techniques in which computers identify patterns in data, with minimal user instructions. The analyses in this study use supervised machine learning, in which the machine learns a pattern that maps a set of predictor variables to an outcome.

Optimal cutoff for a risk score. The risk score cutoff used in the models to identify at-risk students. A common data-driven approach to define the optimal cutoff is to find the cutoff that maximizes the Youden statistic, which is equivalent to maximizing the difference between the true positive rate and the false positive rate (Youden, 1950). Education agencies should choose the cutoff that best aligns with their needs for balancing false positives and false negatives, which is not necessarily the same as the optimal cutoff.

Outcome period. The near term (upcoming term or year, depending on outcome and data collection periods) over which academic problems are defined and predicted. Both the descriptive and predictive analyses in this study define outcome periods over the two most recent available years (2015/16 and 2016/17).

Predictors. The in-school performance, behavior, and consequences and out-of-school human services involvement and public benefit receipt variables whose individual associations with outcomes are examined in the descriptive analysis. Predictors are measured in earlier time periods that are adjacent to those in which outcomes are measured. Time periods are generally adjacent academic terms (quarters in Pittsburgh Public Schools and trimesters in Propel Schools) or, when that is not possible, in adjacent two-month periods or academic years. The descriptive analysis refers to data from the earlier time period as predictors, and data from the later time period as outcomes. This terminology is meant to imply temporal relationships, not causality.

Receiver operating characteristic (ROC) curve. A plot of a model's sensitivity against its false positive rate for every possible decision threshold on the predicted probabilities. It is used to illustrate how well a model distinguishes between students who will have an academic problem and those who will not. When two sets of predictions—for example, for different outcomes or different samples of students—are compared, an ROC curve that is closer to the top left corner (maximizing sensitivity and minimizing the false positive rate) indicates more accurate predictions from the model.

Risk score. Indicates the predicted probability of each outcome occurring in the upcoming student-period (see below). Risk scores range from 0 percent to 100 percent.

Student-period. The level of observation for each outcome. The analyses include one observation for each student for each period (quarter, trimester, or year, depending on the outcome and local education agency) for which the student had an available outcome.

Sensitivity. Probability that the model will correctly predict that a student with an academic problem will have one.

Variable importance. A numerical measure associated with each predictor in a predictive model that expresses how sensitive predictions from that model are to changes in the value of the predictor. Variables (predictors) can be ranked by their variable importance to produce, for example, the 10 most important predictors of the outcome.

and benefit receipt (the predictors) and near-term academic problems (the outcomes). The predictors and outcomes are measured in adjacent time periods—generally in adjacent academic terms (quarters in PPS and trimesters in Propel Schools). For example, for the outcome of chronic absenteeism in PPS in the second nine-week quarter of the school year, the predictors are measured in the first nine-week quarter of that year. When measuring in adjacent terms is not possible, predictors are measured in the preceding two-month period or academic year. Throughout the discussion of the descriptive analysis, the terms *predictors* and *outcomes* are meant to imply only temporal relationships, not causality. The study used linear probability models to estimate the direction and strength of the relationships between predictors in one time period and outcomes in the following time period (see box 2 and appendix A for details on data, sample, and methodology).

Research question 2, the predictive analysis, uses predictive models based on machine learning algorithms to identify the probability that each student will experience each academic outcome in the following period. The models were trained on 2014/15–2015/16 data to determine how to most accurately predict outcomes and tested using 2016/17 data to compare predicted risk scores with actual outcomes.

Box 2. Data sources, samples, and methods

Data sources. The study used student data from three sources. Pittsburgh Public Schools (PPS) and Propel Schools provided a range of student academic data (see table A1 in appendix A). The Allegheny County Department of Human Services (DHS) provided student data on use of social services, justice system involvement, and public benefits receipt (box table 1).

Sample. The descriptive analysis included 28,719 unique PPS students and 4,614 unique Propel Schools students in kindergarten–grade 12. Each combination of outcome and local education agency was analyzed separately. For each analysis the study team first identified all available observations of that outcome for school year 2015/16 or 2016/17. The sample was then limited to outcomes that occurred during academic terms in which the student was enrolled for at least 50 percent of school days, which means that the model predicts risks only for students who met that enrollment threshold. While students who are enrolled for fewer days might also be at risk for academic problems, they are not included in the model. In PPS, 3.5 percent of students were excluded because they did not meet the enrollment threshold in any term in 2015/16 or 2016/17; in Propel Schools, 6 percent of students were excluded for this reason. Sample sizes and number of observations for each analysis are in table A2 in appendix A.

Descriptive analysis. The linear probability models in this study estimate the direction and strength of the relationships between predictors in one period and outcomes in the following period, by local education agency and grade span. All academic outcomes are binary (for example, any out-of-school suspensions or no out-of-school suspensions), as are many predictors. Linear probability models were used because they produced more stable estimates for continuous predictors with skewed or bimodal distributions than logistic regression models did. (See box table 1 and table A1 in appendix A for a high-level description of the predictor data included in the analysis and box table 2 for the definition and level of observation for each outcome.) The relationships in this analysis are unadjusted for other events occurring at the same time or for other student characteristics. Results are displayed in heat maps showing the direction and strength of the relationships (see box 3 for additional information on interpreting heat maps).

Predictive analysis. The predictive models identify the probability, or risk score (between 0 and 100 percent), that each student will experience each outcome in the following period. Three machine learning algorithms—random forest, elastic net logistic regression, and recurrent neural network—were considered, with the random forest model ultimately selected (see appendix A). The models were trained on 2014/15–2015/16 data and tested using 2016/17 data to compare predicted risk scores with actual outcomes. The model calculations summarize each prediction in a single number (the area under the receiver operating characteristic curve; see box 1). For each local education agency and outcome, performance is reported for the entire sample as well as for subgroups of interest, defined by grade span, race/ethnicity, and gender. The most important predictors in each model were examined based on variable importance lists (see box 1), as well as the extent to which model performance differs with and without out-of-school variables as predictors. Finally, the performance of the models with just in-school variables was compared with the performance of models with both in-school and out-of-school variables.

- Students starting or ending a placement were more than 40 percentage points more likely to have a low GPA than students who did not.

A similar pattern, but with smaller percentage point differences, is observed for middle school students experiencing child welfare transitions. That pattern is not observed at the elementary school level.

The findings indicate that ongoing placements of any kind—that is, a student is in a placement, but the placement did not start or end during the time period—might not be as disruptive as placement transitions (see figure B2 in appendix B). For example, PPS high school students with an ongoing placement were 23 percentage points more likely to be chronically absent than students without one, while students in middle school with an ongoing placement were 11 percentage points more likely. Almost no difference was found for elementary school students. An exception to this pattern is for state test scores below the basic level. Relative to child welfare transitions, ongoing placement is associated with below basic state test scores at similar or stronger levels (across grade spans).

Some other types of human services involvement are also associated with academic problems, although relationships are generally weaker than for student academic performance, behavior, and events. Among the Allegheny County DHS human services predictors, emergency shelter services and juvenile justice involvement had the strongest relationships with academic outcomes for PPS students (see figure 2 and table B6 in appendix B; see figure B4 for additional predictors[4]). Across grade levels PPS students with emergency shelter stays of seven days or more were 27–40 percentage points more likely to be chronically absent and 14–27 percentage points more likely to perform below the basic level on state tests than peers without such stays. PPS high school students with emergency shelter stays of seven days or more were 16 percentage points more likely to have a low GPA.

Involvement with the juvenile justice system is also generally associated with higher academic risk among PPS students (see figure 2; see figure B3 in appendix B for additional predictors). Across grade levels students who had an active juvenile justice case in the predictor period were 28–36 percentage points more likely to be chronically absent in the outcome period than students without active cases. Students with active juvenile justice cases were also more likely to experience the other academic problems examined, but the differences were generally smaller than they were for chronic absenteeism. Specific juvenile justice case outcomes—including having a case adjudicated delinquent or having a consent decree[5]—are also associated with increased academic risks. Findings on justice system predictors are similar for Propel Schools students, but the relationship patterns are less consistent (see figure B10).

Behavioral health services and active family court cases are generally associated with smaller differences in risk of academic problems in PPS and Propel Schools.

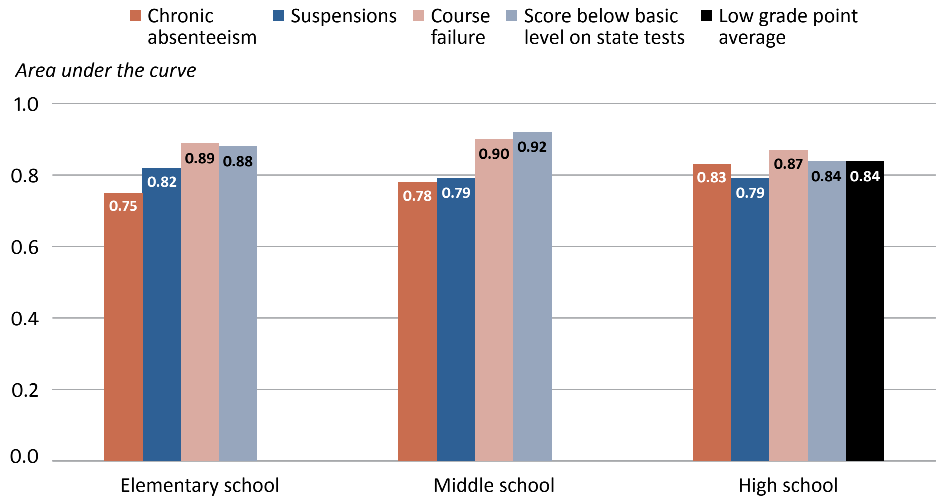
Overall, the predictive models effectively identified at-risk students

The area under the curve (AUC) is a metric used for assessing the strength of predictions; it can have values from 0 to 1, with 1 indicating perfect prediction and .7 and higher considered strong prediction (see box 1). The predictions are strong for PPS: all AUC statistics are above .7 (figure 3). Predictions are somewhat weaker for Propel Schools, although in most cases the AUC was above .7 (figure 4). This discrepancy is likely due to the much larger number of observations available for predictions for PPS than for Propel Schools (box 4). Across both local education agencies the models more accurately predict course and PSSA performance than other academic outcomes.

4. Figure B3 and table B7 (PPS) and figure B10 and table B11 (Propel Schools) in appendix B show additional findings for relationships between other justice system predictors and academic problems. Figure B4 and table B6 (PPS) and figure B11 and table B10 (Propel Schools) in appendix B show additional findings for relationships between other behavioral health and housing services and academic problems.

5. A case that is adjudicated "delinquent" is analogous to a "guilty" verdict in an adult case. A consent decree is an agreement between the court and the juvenile; it might include community service or nonplacement services.

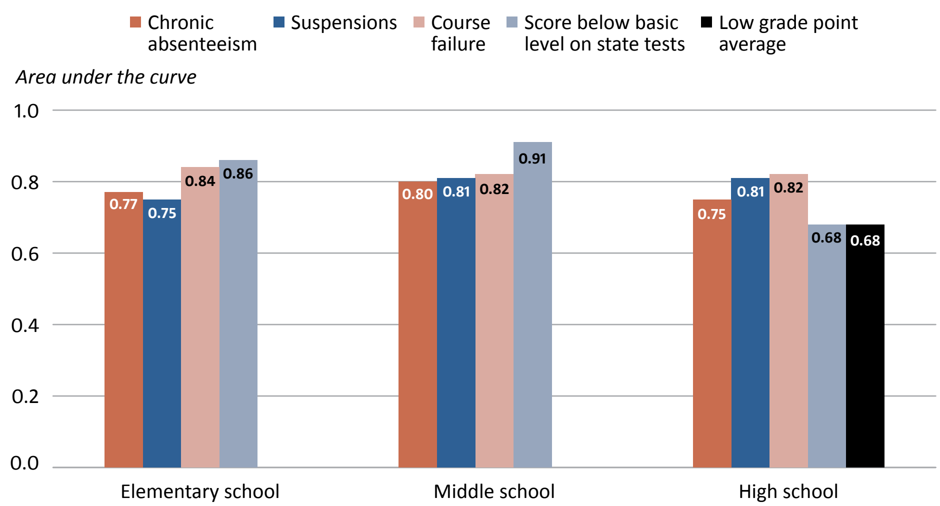
Figure 3. Model predictions are strong for the Pittsburgh Public Schools sample during the 2015/16 and 2016/17 school years, by grade level



Note: See table 2 for definitions of outcomes. The state tests examined are the Pennsylvania System of School Assessment for elementary and middle school and Keystone exams for high school.

Source: Authors' analysis of data from Pittsburgh Public Schools and the Allegheny County Department of Human Services for school years 2014/15–2016/17.

Figure 4. Model predictions are somewhat weaker for the Propel Schools sample than for the Pittsburgh Public Schools sample during the 2015/16 and 2016/17 school years but are still strong, by grade level



Note: See table 2 for definitions of outcomes. The state tests examined are the Pennsylvania System of School Assessment for elementary and middle school and Keystone exams for high school.

Source: Authors' analysis of data from Propel Schools and the Allegheny County Department of Human Services for school years 2014/15–2016/17.

When aggregating across grade levels, the model that makes the best predictions is the one that identifies PPS students at risk for low GPA (see figure B13 in appendix B). There is an 82 percent chance that the model will correctly identify PPS students who will have a low GPA in the outcome period. Only 15 percent of those identified as at risk will not actually have a low GPA.

Box 4. Relationship between the number of observations and model performance

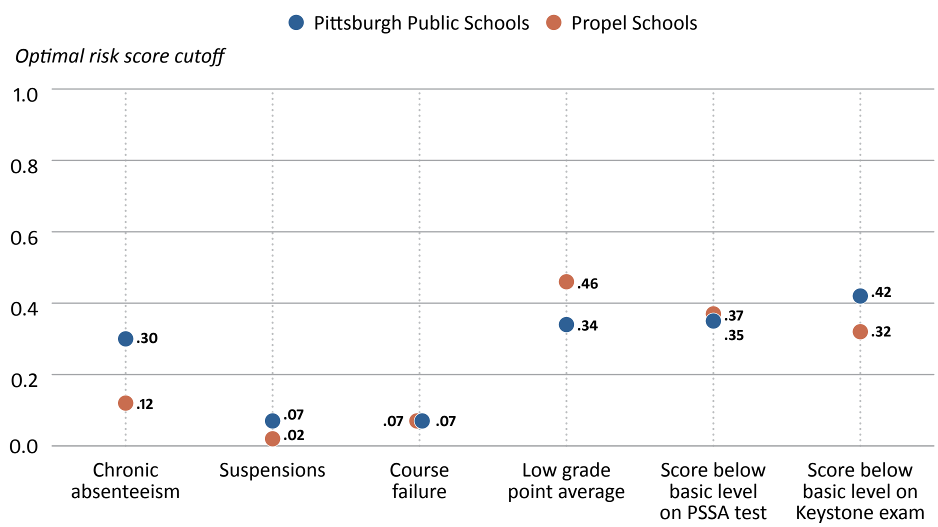
The number of observations is driven by the sample size of unique students—28,719 in Pittsburgh Public Schools (PPS) and 4,614 in Propel Schools—and by the number of outcome periods within the study timeframe. PPS generally has more outcome periods than Propel Schools (for example, PPS tracks some outcomes at the nine-week quarter rather than at the longer trimester tracked in Propel Schools).

The length of outcome periods is important for two reasons. First, shorter periods mean more student-periods to draw on for prediction, increasing the effective sample size. As sample size increases, more information is available to learn the true relationships between predictors and outcomes. These true relationships may include nonlinear effects and interactions that are not distinguishable from noise in small samples. When the sample is larger, the model can determine which relationships are more consistently observed and which are noise, leading to better predictive performance. Second, shorter periods mean that the model can use predictors that occur closer in time to the outcomes, often only weeks or months before the outcome. These more recent events are likely to better predict outcomes than events that occurred in previous years. The difference in study periods between PPS and Propel Schools is especially pronounced for chronic absenteeism. PPS tracks absences at the daily level and predicts them for each nine-week quarter, whereas Propel Schools stores absence data at the annual level.

The optimal risk score cutoff varies across outcomes and local education agencies. The predictive models calculate a risk score between 0 percent and 100 percent for each student for each outcome. To know how to use these numbers to guide support for students, local education agencies need to determine what level of risk score would indicate that students are at risk and thus eligible for intervention.

While many contextual factors go into determining a risk score cutoff for identifying at-risk students, this study employed a commonly used statistical approach known as the Youden statistic (Youden, 1950). The report refers to these cutoffs as "optimal risk score cutoffs." Each outcome and local education agency combination has a different optimal risk score cutoff (see figure 5 and table B1 in appendix B; see figure B13 for cutoffs on the receiver operating characteristic curves). The analyses of the number and characteristics of at-risk students identified by the models consider all students with risk scores above the cutoff for each outcome to be at risk.

Figure 5. Optimal risk score cutoffs by outcome and local education agency, 2014/15–2016/17



PSSA is Pennsylvania System of School Assessment.

Note: See table 2 for definitions of outcomes. The PSSA is administered in elementary and middle school and the Keystone exams are administered in high school.

Source: Authors' analysis using data from Pittsburgh Public Schools, Propel Schools, and the Allegheny County Department of Human Services for school years 2014/15–2016/17.

The number of students identified as at risk using the optimal risk score cutoffs varies by outcome, and for some outcomes more than half of students are considered at risk (table 3). Nearly 18,000 students (75 percent) across all grades were identified as at risk for at least one academic outcome at some point in 2016/17 in PPS, and more than 2,400 (65 percent) in Propel Schools. The most frequently predicted outcomes were chronic absenteeism and suspensions in PPS and suspensions and course failure in Propel Schools. Students can be flagged as at risk for multiple outcomes and in multiple periods. (See figures B15–B17 in appendix B for the characteristics of at-risk students for each outcome by local education agency.)

Prior academic problems and other student characteristics and services tracked in school data systems are the strongest predictors

Prior academic problems, including absences, course performance, and state test performance, are consistently in the variable importance lists of the strongest predictors. That these are among the top predictors is not surprising (see variable importance in box 1): previous absences would be expected to strongly predict current absences, and previous test performance to strongly predict current test performance. The only human services predictors that appear among the strongest predictors are outpatient behavioral health services and HealthChoices insurance, an indicator of Medicaid eligibility. Even human services predictors with a strong correlation with academic problems, such as a change in placement and a stay in a homeless shelter, were not among the top 10 predictors when all variables were included.

Excluding out-of-school predictors does not substantially change how well the models identify at-risk students. Excluding human services data as predictors did not substantially reduce the ability of the models to predict any of the outcomes in PPS or Propel Schools (see table B4 in appendix B). This pattern was consistent overall and for subgroups of students, including students in elementary school, middle school, and high school; students in kindergarten–grade 3, who do not have prior state test score data; and students with any social services involvement during the outcome period. Including out-of-school predictors did improve the predictions for students with no previous in-school data (including students in kindergarten and students transferring to the local education agency).

Table 3. Number of students identified as at risk at any point in 2016/17 using optimal risk score cutoffs, by outcome and local education agency

	Pittsburgh Public Schools				Propel Schools			
Outcome	Elementary school	Middle school	High school	Total	Elementary school	Middle school	High school	Total
Chronic absenteeism	6,677	3,236	5,379	15,292	687	287	242	1,216
Suspensions	2,896	2,921	4,396	10,213	356	474	489	1,319
Course failure	1,949	1,651	3,960	7,560	739	593	411	1,743
Low grade point average	na	na	3,663	3,663	na	na	325	325
Score below basic level on PSSA test	2,398	2,883	na	5,281	533	511	na	1,044
Score below basic level on Keystone exam	na	na	1,965	1,965	na	na	88	88
Any outcome[a]	7,837	4,193	5,948	17,978	1,199	715	532	2,446

PSSA is Pennsylvania System of School Assessment.

Note: See table 2 for definitions of outcomes. This table shows the number of unique students predicted as at risk for each outcome using the optimal cutoffs in any period during 2016/17.

a. The number of unique students predicted as at risk for one or more of the outcomes during 2016/17.

Source: Authors' calculations using data from Pittsburgh Public Schools, Propel Schools, and the Allegheny County Department of Human Services for school years 2014/15–2016/17.

Table 4. Most important categories of predictors in predictive analysis, by outcome and local education agency, 2016/17

	Prior academic problems predictors				Other predictors			
Outcome	Absences or tardiness	Suspensions	Course grades or grade point average	State test score level	Age or grade level	Special education or type of disability	HealthChoices	Outpatient behavioral health services
Pittsburgh Public Schools								
Chronic absenteeism	✔		✔	✔	✔		✔	✔
Suspensions	✔	✔	✔	✔	✔			
Course failure	✔		✔		✔			
Low grade point average	✔		✔	✔	✔			
Score below basic level on PSSA test	✔		✔	✔	✔	✔		
Score below basic level on Keystone exam	✔		✔	✔	✔			
Propel Schools								
Chronic absenteeism	✔		✔	✔	✔		✔	✔
Suspensions	✔	✔	✔		✔			
Course failure	✔		✔	✔	✔			
Low grade point average	✔		✔	✔	✔		✔	
Score below basic level on PSSA test	✔		✔	✔	✔	✔		
Score below basic level on Keystone exam	✔		✔		✔	✔		

PSSA is Pennsylvania System of School Assessment.

Note: See table 2 for definitions of outcomes. This table shows the categories of predictors in the top 10 variable importance list in each random forest model. Each category could represent more than one individual predictor. For example, there are separate predictors for total number of absences, number of excused absences, number of unexcused absences, chronic absenteeism, and number of days tardy. In addition to the predictor categories shown in this table, the subject of the test was in the top 10 list of predictors for the state test outcomes. Other predictors included in the models are not shown in this table because they did not appear in the top 10 variable importance list for any model.

Source: Authors' calculations using data from Pittsburgh Public Schools, Propel Schools, and the Allegheny County Department of Human Services for school years 2014/15–2016/17.

Limitations

An important limitation of the study is that the models are trained on a snapshot of historical data (drawn from 2014/15 and 2015/16) and based on relationships between predictors and outcomes in that timeframe. The underlying relationships between predictors and outcomes may change over time. For example, if behavior policies change and some types of behavior are no longer as likely to lead to out-of-school suspensions, the model might not predict which students are at risk for suspensions under the new policy as well as it did under the previous policy. Changes in how predictor data are measured would also affect the performance of the model on future data. To lessen this risk, the model should be periodically retrained using more recent data.

A key limitation of the descriptive findings for research question 1 is that the relationships are not adjusted for any other background characteristics or events and should not be considered causal. For example, one cannot conclude that having an ongoing child welfare placement is worse for a student—or causes future risks for the student—than not having a placement. Rather, this relationship implies that students in ongoing placements are at higher risk for some academic problems than students who are not. These findings are intended to demonstrate the strong correlations between some predictors and outcomes, which could provide insights for student support staff into why a student is flagged as at risk and how best to support the student.

Limitations of the data also need to be considered. The data are from administrative sources and are likely to have both random and systematic measurement errors. Although standardized test scores are determined at the state level and are likely to have a consistent and reliable meaning across students, schools, and local education agencies, schools may track other outcomes, such as absences, suspensions, and grades, differently from each other. In addition to differences in school policies, the data are subject to random human error and the subjectivity of individual teachers or administrators who determine the outcomes. These concerns are not unique to the local education agencies included in this study, and multiple studies across the country show that these outcomes —with similar levels of measurement error—are correlated with dropping out of school (Allensworth et al., 2014).

It is particularly important to identify any systematic bias in how outcomes are determined for particular subgroups of students. For example, if students of a specific racial/ethnic group or gender are more likely than other students to receive a suspension for the same type of behavior, the underlying data feeding into the model would be biased in measuring the true behavior of students. With the available data there is no way to assess whether a school's disciplinary system or grading system is biased. For that reason, users of the predictive model should recognize that the model is predicting the recorded outcomes rather than student behavior per se and that the correlation between outcomes and actual behavior might vary across racial/ethnic or gender subgroups.

Implications

The study examined how in-school and out-of-school events and services in students' lives are related to and can be used to predict near-term academic problems, including absenteeism, suspensions, course performance, and test performance. These outcomes are important warning flags for school dropout. If administrators and school staff can identify students likely to experience these outcomes in the near term, they can provide additional support before a problem worsens.

As a hypothetical example, a student might be chronically absent for much of grade 9 and then drop out at the end of the year. A typical early warning system using grade 9 attendance data to identify students at risk of dropout might not identify this student until after the first semester or perhaps not until the end of grade 9, too late to prevent the student from dropping out. An early warning system that predicts chronic absenteeism in the near term could flag this student as at risk early in the year based on a range of predictors from the first two months of school or the previous year. Educators could provide additional support before the student gets to the level of chronic absenteeism in grade 9 and drops out.

The predictive analysis discussed in this report indicates that the data on in-school and out-of-school events available in Allegheny County can be used to correctly identify most students who are at risk for academic problems. Models predicting outcomes in PPS identified at-risk students at a level that is generally considered strong in social science (Rice & Harris, 2005). Most models for Propel Schools also met this threshold, with the exception of suspensions, which are rare in Propel Schools and consequently more difficult to predict.

Moreover, the predictive ability of the models was approximately the same for models using only in-school data and for models that also incorporated out-of-school predictors. The only human services data on the lists of top 10 predictors in variable importance were outpatient behavioral health services and HealthChoices eligibility, and those were important for only two of the outcomes. This implies that while individual human services predictors are related to outcomes (as shown in the descriptive analysis), these predictors do not substantially improve the models' predictive ability beyond what is already predicted by school data.[6] For local education agencies

6. This is similar to findings in the literature comparing teacher value-added models, which indicate that predictions of student achievement growth are relatively insensitive to the inclusion of additional predictors beyond prior test performance (Ballou, Sanders, & Wright, 2014; Johnson, Lipscomb, & Gill, 2015).

outside Allegheny County that have school data similar to that of the local education agencies in this study, a similar predictive modeling approach could be used without human services data. Even in Allegheny County, users could consider whether to use out-of-school data for predictions, given the cost implications for processing the linked data.

Nonetheless, educators and human services staff may want to use out-of-school data to better understand some of the underlying challenges faced by students in their home lives. Involvement with human services can be an indicator of underlying challenges at home, and these problems likely play a role in academic problems. The descriptive analysis shows clear connections between a number of in-school and out-of-school variables and academic outcomes that can help in understanding why the model flags a student as at risk—or why a student has had prior academic problems. Even if students' human services event histories are not included in the predictive models, educators and human services staff may want to examine them to inform strategies to address academic risks. A number of out-of-school events—including child welfare services, juvenile justice involvement, and homeless services—are associated with greater likelihood of academic problems. Within these categories, there are specific types of events, such as transitions in and out of child welfare placements, that are more strongly associated with academic problems. Educators and human services staff might want to focus resources on students during those types of transitions.

Local education agency staff developing and using similar predictive models might want to consider the differences in predictive performance across local education agencies, outcomes, and student subgroups and the implications for practice. Staff collecting data and developing similar models might want to:

- *Maximize sample size and the number of outcome periods.* The better predictions for PPS than for Propel Schools are driven in part by the much larger sample of PPS students and by the greater number of outcome periods in the study timeframe. If this model is retrained on new data, both maximizing sample size and using data stored at the most granular level possible will likely lead to better predictions.
- *Develop capacity to do machine learning analysis.* Local education agencies will need staff who are familiar with random forest models or other more flexible machine learning approaches. Though the process of fitting such a model is relatively straightforward and generally requires less preprocessing and fewer modeling decisions than linear or logistic regression, it does require familiarity with the approach and a compatible software package.

Local education agency administrators and staff determining how to use the information produced by predictive models may want to:

- *Consider that some outcomes are easier for the models to predict than others when deciding where to devote resources.* The models predicting course-based and state test outcomes perform better because students' predictor histories are more strongly associated with future outcomes of these types. Chronic absenteeism and suspensions, in contrast, might be driven by events that are not reflected in students' predictor history. Local education agencies that want to maximize the number of students served who are actually at risk might choose to identify at-risk students based on outcomes that have better predictions.
- *Define risk score cutoffs separately for each outcome based on local data.* For both local education agencies, each outcome has a different optimal risk score cutoff for maximizing accurate predictions and minimizing false predictions. The optimal cutoffs for the same outcomes in PPS and Propel Schools differ considerably in some cases, implying that local education agencies should use their own data to determine optimal cutoffs.
- *Consider other local factors when defining risk score cutoffs.* While optimal cutoffs provide a starting point, users will also want to consider other local factors. For example, if the cost of intervention is high, users might set a cutoff that further reduces the false positive rate, even if that leads to missing some at-risk students. Multiple risk categories might also be useful if tiered interventions are available. For example, students in the highest risk category might receive case management and individualized support, while students in a middle risk category could receive group-based supports.

References

Allegheny County Department of Human Services. (2015). *Data brief: The impact of child welfare system involvement on school stability.* Allegheny, PA: Author.

Allensworth, E., Gwynne, J., Moore, P., & de la Torre, M. (2014). *Looking to High School and College: Middle grade indicators of readiness in Chicago Public Schools.* Chicago, IL: The University of Chicago Consortium on School Research. https://eric.ed.gov/?id=ED553149.

Balfanz, R., Herzog, L., & Mac Iver, D. (2007). Preventing student disengagement and keeping students on the graduation path in urban middle-grades schools: Early identification and effective interventions. *Educational Psychologist, 42*(4), 223–235. https://eric.ed.gov/?id=EJ780922.

Ballou, D., Sanders, W., & Wright, P. (2004). Controlling for student background in value-added assessments of teachers. *Journal of Educational and Behavioral Statistics*, *29*(1), 37–65. https://eric.ed.gov/?id=EJ727499.

Bowers, A. J., Sprott, R., & Taff, S. A. (2013). Do we know who will drop out? A review of the predictors of dropping out of high school: Precision, sensitivity, and specificity. *The High School Journal, 96*(2), 77–100. https://eric.ed.gov/?id=EJ995291.

Bruce, E., Bridgeland, J., Hornig Fox, J., & Balfanz, R. (2011). *On track for success: The use of early warning indicator and intervention systems to build a grad nation.* Washington, DC: Civic Enterprises. https://eric.ed.gov/?id=ED526421.

Edmunds, J., Willse, L., Arshavsky, N., & Dallas, A. (2013). Mandated engagement: The impact of early college high schools. *Teachers College Record, 115*(7), 1–31. https://eric.ed.gov/?id=EJ1020156.

Gelman, A. (2008). Scaling regression inputs by dividing by two standard deviations. *Statistics in Medicine, 27*(15), 2865–2873.

Johnson, M. T., Lipscomb, S., & Gill, B. (2015). Sensitivity of teacher value-added estimates to student and peer control variables. *Journal of Research on Educational Effectiveness, 8*(1), 60–83. https://eric.ed.gov/?id=EJ1049638.

Kearney, C. (2008). School absenteeism and school refusal behavior in youth: A contemporary review. *Clinical Psychology Review, 28*(3), 451–471.

National Forum on Education Statistics. (2018). *Forum guide to early warning systems*. Washington, DC: U.S. Department of Education. https://eric.ed.gov/?id=ED594388.

Nembrini, S., Koenig, I. R. & Wright, M. N. (2018). The revival of the Gini Importance? *Bioinformatics 34*(21), 3711–3718.

Pittsburgh Public Schools. (2020). *Attendance summary: District-wide rates.* Pittsburgh, PA: Author. Retrieved February 25, 2020, from https://www.pghschools.org/Page/5075.

Rice, M. E., & Harris, G. T. (2005). Comparing effect sizes in follow-up studies: ROC area, Cohen's d, and r. *Law and Human Behavior, 29*(5), 615–620.

Twala, B., Jones, M. C., & Hand, D. (2008). Good methods for coping with missing data in decision trees. *Pattern Recognition Letters, 29*(7), 950–956.

U.S. Department of Education. (2016). *Issue brief: Early warning systems.* Washington, DC: Office of Planning, Evaluation and Policy Development, Policy and Program Studies Service. https://eric.ed.gov/?id=ED571990.

Wooldridge, J. M. (2013). A binary dependent variable: The linear probability model. *Introductory econometrics: A modern approach*. Mason, OH: South-Western.

Youden, W. J. (1950). Index for rating diagnostic tests. *Cancer, 3*(1), 32–35.

REL 2020–027

July 2020

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 Policy Research. 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.

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). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Mid-Atlantic. Retrieved from http://ies.ed.gov/ncee/edlabs.

This report is available on the Regional Educational Laboratory website at http://ies.ed.gov/ncee/edlabs.