FACTSheet

REL MID-ATLANTIC
Regional Educational Laboratory at Mathematica

Serving the education community in Delaware, the District of Columbia, Maryland, New Jersey, and Pennsylvania

Using School and Child Welfare Data to Predict Near-Term Academic Risks

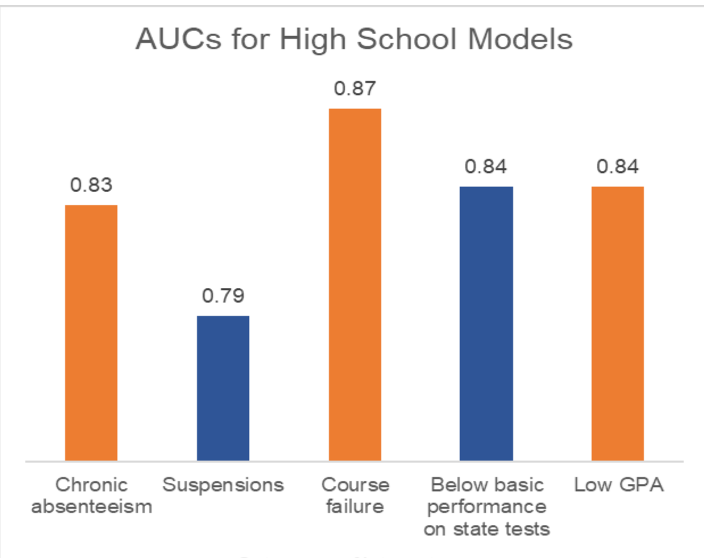
Many districts use early warning systems that identify students who are at risk of academic problems or dropping out of high school so they can better support these students. These systems often track attendance, behavior, and course performance (ABC); indicators that reliably identify students at risk of dropping out in large urban districts.[i] Although districts typically do not incorporate non-academic risk factors into these systems, research has shown that homelessness, teenage pregnancy, child maltreatment, parental substance abuse, and unsafe living conditions are risk factors for negative school outcomes.[ii]

REL Mid-Atlantic partnered with Pittsburgh Public Schools, Propel Schools (a charter school network), and the Allegheny County Department of Human Services to assess how well combinations of in- and out-of-school data predict ABCs in the near-term that are reliable precursors of high school dropout. By flagging students who experience precursors to dropout, educators can target resources to those most at risk and intervene before problems become more serious.[iii]

This fact sheet describes predictors related to near-term academic problems, meaning academic problems in the next quarter or semester. It draws on a *report* that details the approach used to develop a predictive model and assesses how well the model identifies at-risk students.

WHAT WE FOUND

1. **Predictive models based on machine-learning (ML) algorithms successfully identify at-risk students.** The models most accurately predict course failure and state test score performance, though the models for other outcomes also perform well. The model that predicted whether Pittsburgh Public Schools students would have low GPA was the most accurate. Model accuracy is measured using an industry-standard metric called Area Under the Curve (AUC), which can be interpreted as the probability that a randomly selected student with an academic problem is considered at higher risk (by the model) than a randomly selected student without an academic problem.



Generally, numbers above 0.7 indicate a strong model fit in social science.

2. **School data are the strongest predictors of near-term academic problems across all outcomes.** The strongest predictors include prior absences, low prior grades, and low prior performance on state tests.
3. **Predictive performance remains strong when the ML model relies exclusively on school data.** Excluding human services data as predictors in the models did not substantially reduce the ability of the models to make predictions for any of the outcomes. When considering all variables simultaneously, outpatient behavioral health services and Medicaid eligibility are the only two human services predictors that appear among the 10 strongest predictors.
4. **Accuracy varies by sample size and outcome.** The models are more accurate in the larger Pittsburgh Public School than in the smaller Propel charter network. In both districts, they are more accurate at predicting low grade point average, course failure, and

We want to hear from you! Please contact us!

July 2020

below basic performance on state assessments in grades 3 to 8 than they are for chronic absenteeism, suspensions, and below basic performance on end-of-course high school standardized assessments.

5. **Some out-of-school events are individually correlated with near-term academic problems.** Even though the out-of-school variables did not contribute to the predictive power of the ML models, some of the variables are related to the outcomes in descriptive analyses. Only a small percentage of students experience child welfare involvement, but child welfare placement and removal were among the strongest human services predictors of academic outcomes, especially at the high school level. Emergency shelter services and juvenile justice involvement were also strongly related to poor academic outcomes. Pittsburgh students who stayed in emergency shelters seven days or more in the prior period were 27 to 40 percentage points more likely to be chronically absent and 14 to 27 percentage points more likely to perform below the basic level on state tests. Students with an active juvenile justice case were 28 to 36 percentage points more likely to be chronically absent compared to students without active cases.

WHAT DOES THIS MEAN FOR OTHER SCHOOL DISTRICTS?

- Districts could create similar predictive modeling approaches relying solely on in-school data, because previous school outcomes are the strongest predictors of future school outcomes. Out-of-school data, if available, may be useful for providing context—and it is possible that they would improve the prediction of longer-term outcomes such as dropout, which could not be examined in our study.
- Collecting detailed data frequently could improve predictions. Practitioners should maximize the number of time periods available for events (for example, quarterly would be better than yearly).
- For smaller local education agencies, increasing sample size by pooling data with other local education agencies that serve similar populations and have similar data might improve predictive power.
- An ideal early warning system should consider the strength of the prediction in determining how and when to intervene with services. For example, for a more accurate model (like one predicting GPA issues), schools may choose to intervene and provide tutoring services at the first sign. Whereas, for a less accurate model (like one predicting suspensions), schools may set a higher threshold for intervention (perhaps risk flags in two quarters in a row).

Data sources, sample, and methods

Data sources. The study used data from three sources. Pittsburgh Public Schools and Propel schools provided student demographic, enrollment, course, state test, attendance, and behavior data. Allegheny County Department of Human Services provided student data on receipt of social services, justice system involvement, and public benefits.

Sample. The descriptive analysis included 28,719 unique Pittsburgh Public Schools students and 4,614 unique Propel students in kindergarten to 12th grade.

Descriptive and predictive analysis methods. The descriptive analysis measured how, on average, predictors in one period relate to outcomes in the following period, separately by local education agency and grade levels (elementary, middle, and high). The descriptive relationships describe how one variable relates to another without factoring in other variables. Predictive models use in- and out-of-school data from recent periods to identify the probability that each student will experience each outcome in the upcoming period. The predictive models consider how multiple variables relate to the outcome all at once.

ENDNOTES

[i] 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/?q=Preventing+student+disengagement+and+keeping+students+on+the+graduation+path+in+urban+middle-grades+schools%3a+Early+identification+and+effective+interventions.+&id=EJ780922; 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/?q=Do+we+know+who+will+drop+out%3f+A+review+of+the+predictors+of+dropping+out+of+high+school%3a+Precision%2c+sensitivity%2c+and+specificity.+&id=EJ995291

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

[iii] 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/?q=On+track+for+success%3a+The+use+of+early+warning+indicator+and+intervention+systems+to+build+a+grad+nation.&id=ED526421; Edmunds, J. Willse, L., Arshavsky, N., & Dallas, A. (2013). Mandated engagement: The impact of early college high schools. Teachers College Record, 115(1-31). https://eric.ed.gov/?q=Mandated+engagement%3a+The+impact+of+early +college+high+schools.+&id=EJ1020156

This work was funded by the U.S. Department of Education's Institute of Education Sciences (IES) under contract ED-IES-17-C-0006, with REL Mid-Atlantic, administered by Mathematica. The content of this fact sheet 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.