Identifying Students At Risk Using Prior Performance Versus a Machine Learning Algorithm

Regional Educational Laboratory Mid-Atlantic

At Mathematica

REL 2021–126
U.S. DEPARTMENT OF EDUCATION

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

Identifying Students At Risk Using Prior Performance Versus a Machine Learning Algorithm

Lindsay Cattell and Julie Bruch

September 2021

This report provides information for administrators in local education agencies who are considering early warning systems to identify at-risk students. Districts use early warning systems to target resources to the most at-risk students and intervene before students drop out. Schools want to ensure the early warning system accurately identifies the students that need support to make the best use of available resources. The report compares the accuracy of using simple flags based on prior academic problems in school (prior performance early warning system) to an algorithm using a range of in- and out-of-school data to estimate the specific risk of each academic problem for each student in each quarter. Schools can use one or more risk-score cutoffs from the algorithm to create low- and high-risk groups. This study compares a prior performance early warning system to two risk-score cutoff options: a cutoff that identifies the same percentage of students as the prior performance early warning system, and a cutoff that identifies the 10 percent of students most at risk.

The study finds that the prior performance early warning system and the algorithm using the same-percentage risk score cutoffs are similarly accurate. Both approaches successfully identify most of the students who ultimately are chronically absent, have a low grade point average, or fail a course. In contrast, the algorithm with 10-percent cutoffs is good at targeting the students who are most likely to experience an academic problem; this approach has the advantage in predicting suspensions, which are rarer and harder to predict than the other outcomes. Both the prior performance flags and the algorithm are less accurate when predicting outcomes for students who are Black.

The findings suggest clear tradeoffs between the options. The prior performance early warning system is just as accurate as the algorithm for some purposes and is cheaper and easier to set up, but it does not provide fine-grained information that could be used to identify the students who are at greatest risk. The algorithm can distinguish degrees of risk among students, enabling a district to set cutoffs that vary depending on the prevalence of different outcomes, the harms of over-identifying versus under-identifying students at risk, and the resources available to support interventions.

Why this study?

Many school districts use a prior performance early warning system that tracks attendance, behavior, and course performance to identify students at risk of dropping out. For example, school districts might flag students who missed more than 10 percent of school days in the first semester as at risk for chronic absenteeism in the second semester. Research has shown that these indicators can reliably identify students who are at risk of dropping out (Allensworth et al., 2014; Balfanz et al., 2007; Bowers et al., 2013). Districts can use early warning systems to target resources to the most at-risk students and intervene before students drop out (Bruce et al., 2011; Edmunds et al., 2013). Of course, the system needs to correctly identity at-risk students for the intervention to have an impact.

Pittsburgh Public Schools (PPS) requested this study to compare the district's prior performance early warning system to a more sophisticated algorithm that uses a range of in-school and out-of-school data to identify at-risk students. In the 2017/18 school year, PPS rolled out a system that identifies at-risk students based on their prior attendance, behavior, or course performance problems. Many districts have similar early warning systems (U.S. Department of Education, 2016). Support staff can use the flags to identify at-risk students, monitor them, or provide additional supports. This report refers to this approach as the "PPS flags." The system creates four flags relevant to this study:

- **Chronic absenteeism flag:** Every day the system identifies the students who have been absent more than 10 percent of days in that quarter.
- **Course failure flag:** At the end of each quarter, the system identifies the students who failed a course.
- **Low GPA flag:** At the end of each quarter, the system identifies students with a low grade point average, or GPA (less than or equal to 2.2).
- **Suspension flag:** At the end of each quarter, the system identifies the students with any out-of-school suspension.

A 2020 Regional Educational Laboratory Mid-Atlantic study developed an alternative approach to identifying students who are at risk: a sophisticated early warning system for PPS that generates risk scores based on a machine learning algorithm and the district's unique dataset incorporating in-school and out-of-school data (Bruch et al., 2020). Machine learning models use data-driven algorithms designed to extract the most relevant information from a dataset, with a focus on maximizing the predictive performance of the model. The risk scores indicate the likelihood that the student will experience chronic absenteeism, course failure, low GPA, or a suspension in the following quarter. The algorithm generates the risk scores based on in-school data on academics and behavior combined with out-of-school data from the Allegheny County Department of Human Services (DHS), such as child welfare involvement and justice system involvement.

The prior study found that the predictive model risk scores identify at-risk students with a moderate-to-high level of accuracy (Bruch et al., 2020).[1] Across grade levels and predicted outcomes, accuracy ranged from .75 to .92.[2] Data from schools—including prior academic problems and other student characteristics and services—are the strongest predictors across all outcomes. The predictive performance of the model is not reduced much when excluding social services and justice system predictors and relying exclusively on school data.

While the predictive model risk scores and the PPS flag system both are predicting the same outcomes, the two approaches have several key differences:

- **Data:** The PPS flags only use in-school data, whereas the predictive model risk scores also use out-of-school data from DHS.
- **Methods:** The PPS flags simply rely on the binary performance in a prior time period (such as failed a course in the prior quarter to predict failing a course in the next quarter). In comparison, the predictive risk scores are developed from a machine learning model that accounts for many input variables from the previous quarter. The machine learning model automatically determines the relative importance of each input variable.
- **Output:** The PPS flags are binary yes/no predictions of student performance, whereas the predictive risk scores are a likelihood from 0 to 1. The risk scores are converted to binary predictors using a cutoff that sorts students into high- and low-risk categories. This study tests two cutoffs.

PPS would like to know how these two different early warning systems compare regarding who is identified as at risk and how often each prediction method correctly identifies students who ultimately experience academic problems. The findings of this study will inform whether the predictive model risk score early warning system should be adopted. Accurately identifying students at risk for academic problems is a priority for PPS and DHS.

---

[1] The prior study included students from PPS and the Propel charter school network. This study includes only PPS students.

[2] The strength of predictions, or accuracy, is measured using a metric called the area under the curve (AUC); it can have values from 0 to 1, with 1 indicating perfect prediction. An AUC of .7 or higher is considered a strong prediction (Rice & Harris, 2005).

One-third of PPS students missed at least 10 percent of school days in 2018/19 (Pittsburgh Public Schools, 2020). In line with national trends for chronically absent students, more than half of the chronically absent students in PPS in 2011/12 had GPAs below 2.5. In lower grades, about half of chronically absent students were not proficient on their state reading tests (Allegheny County Department of Human Services, 2014). Chronic absenteeism is especially high among students receiving public benefits or mental health services as well as those 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). In school year 2019/20, 9.5 percent of students enrolled at any point in the year were suspended (Pittsburgh Public Schools, 2021).

Box 1. Key terms

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

Over-include. To lean toward identifying more students as being at risk even if that means including some students who will not ultimately have an academic problem.

Prevalence rate. The percent of student-quarters that experience the academic problem.

Risk score. A student-specific score that indicates the predicted probability of each outcome occurring in the upcoming student-quarter. Risk scores range from 0 percent probability to 100 percent probability. During the previous study, the machine learning algorithm produced one risk score for each student in each quarter in the sample.

Same percentage cutoffs. A cutoff that identifies the same proportion of at-risk students as the PPS flags. It varies for each of the outcomes based on the prevalence rate. This is the most direct comparison to the PPS flags and therefore the best way to compare accuracy.

Student-quarter. The level of observation for each outcome. The analyses include one observation for each student for each quarter for which the student had an available outcome.

Ten percent cutoffs. A cutoff that identifies the 10 percent of students most likely to have an academic problem. This cutoff prioritizes the students most at risk.

Under-include. To lean toward identifying fewer students as being at risk even if that means missing some students who will ultimately have an academic problem.

Research questions

The goal of the study is to provide information to PPS about the comparative performance of the predictive model risk scores and the PPS flags in identifying students at risk for academic problems. The study answers the following research questions:

1. **Which approach more accurately predicts near-term student outcomes?** This research question examines which approach is better at predicting the next quarter's outcomes for all students. A more accurate approach will be preferred by PPS because it will help target resources to the students most in need of assistance.
2. **Which approach more accurately predicts near-term student outcomes for student groups of interest (defined by grade span, race/ethnicity, gender, DHS involvement, economic disadvantage, special education status, and English learner status)?** This research question examines which approach is better at predicting outcomes for students with certain demographic characteristics.

Four out of the five types of outcomes examined by Bruch et al. (2020) are included in this study (table 1). State test scores are not included in this study because PPS does not currently produce flags for low test scores. To ensure comparability, the outcomes and prior performance for PPS flags were calculated using the same definition for each student-quarter (see table A1 in appendix A).

For both the risk score analysis and the PPS flags, the predictors come from the quarter prior to the observed outcome. For example, data from the first nine-week quarter of the year predict chronic absenteeism in the second nine-week quarter of the year. This report 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.

Table 1. Definition of outcomes for risk score, outcomes for PPS flags, and prior performance PPS flags

Construct	Definition for outcomes and PPS flags	Grade levels	Level of observation
Chronic absenteeism	Absent from school (excused or unexcused) for more than 10 percent of days	K–12	Student-quarter
Any suspension	One or more out-of-school suspensions during a quarter	K–12	Student-quarter
Course failure	Receipt of a failing grade for a core course, graded on a standard A–F or A–E scale	K–12	Student-quarter
Low grade point average	Quarter-specific grade point average below 2.2	9–12	Student-quarter

PPS is Pittsburgh Public Schools.
Note: See appendix A for more details about the calculation of these variables.
Source: Authors' tabulation based on data from Allegheny County Department of Human Services for 2014/15–2016/17.

Box 2. Data sources, sample, and methods

Data sources. The study used student data from two sources: Pittsburgh Public Schools (PPS) and the Allegheny County Department of Human Services (DHS). PPS provided a range of student academic data (see table A1 in appendix A). The Allegheny County DHS provided student data on use of social services, justice system involvement, and public benefits (see table A2).

Sample. The analysis included 23,848 unique PPS students in kindergarten through grade 12 in school year 2016/17. For each analysis, the previous study team first identified all available observations of that outcome for 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 per term. The sample was limited this way to make the results more relevant to most students. This means the results may not be relevant for students enrolled for very short periods. See table A3 in appendix A for sample sizes and number of observations for each analysis.

Methodology. The study team retrospectively created the PPS flags for the school year 2016/2017, because PPS had not yet created flags for that school year. For the predictive risk scores (the full description is in appendix A), which are continuous on a scale of 0 to 1, the team determined cutoffs to sort students into high- and low-risk categories for each outcome. This study examines two cutoffs. First, the "same percentage cutoffs" which identify the same proportion of at-risk students as the PPS flags. To create the cutoff, the study team first created the PPS flags and found the percent of students flagged for each academic outcome. Then, using the continuous risk scores, the team identified a cutoff that would classify the same percentage of students as the PPS flags as "high-risk." For example, the PPS flag identified 27 percent of students at risk of being chronically absent. The risk score cutoff for chronic absenteeism was set at 0.43, because 27 percent of students had a risk score above 0.43. The same percentage cutoffs are the most direct comparison to the PPS flags and therefore the best way to compare accuracy. Second, the study team created the "ten percent cutoffs," which identify the 10 percent of students most likely to have an academic problem (which prioritizes the students predicted to be at the highest risk and might correspond better to the percent of students a school district could actually provide additional support to with limited resources).[3] For each student, outcome, and quarter of 2016/17, the researchers calculated binary indicators of risk: one based on the PPS flags alone and one for each of the two cutoffs of the continuous predictive risk scores.

[3] The study team also included an analysis of a third cutoff option that is statistically calculated to maximize the difference between the sensitivity and the specificity (see appendixes A and B). The resulting cutoffs cast a wide net and predicted many of the students in the sample would experience an academic problem. In practice, identifying such a large percentage of students is not helpful for a school

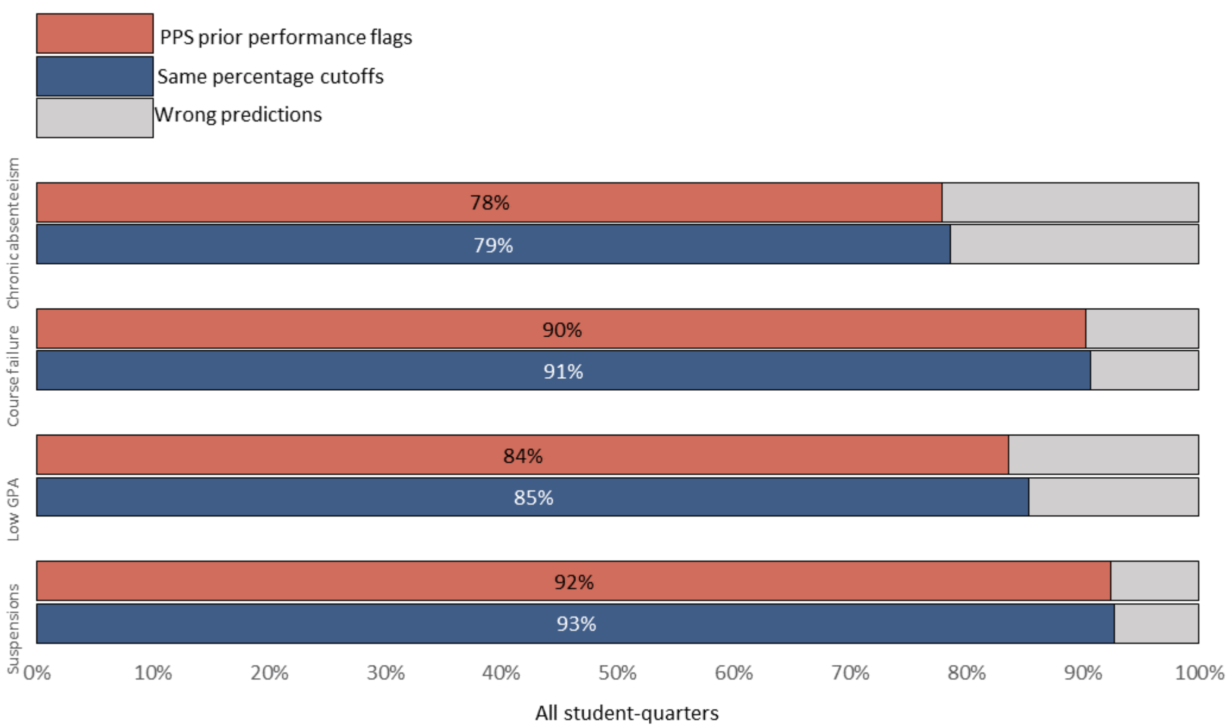
To answer research question 1, this study performs four comparisons of the predictions to the true actual outcomes: the percent of student-quarters with correct or incorrect predictions (figure 1), the percent of student-quarters that were predicted to have an academic problem among the student-quarters that ultimately had an academic problem (figure 2), the percent of student-quarters predicted not to have an academic problem among the students-quarters that ultimately did not have an academic problem (figure 3), and the percent of student-quarters that actually had an academic problem among the student-quarters predicted to have an academic problem (figure 4). To answer research question 2, the study compared the percent of wrong predictions among White students and Black students for each outcome. This report focuses on differences across prediction approaches of more than 5 percentage points.[4]

Findings

The Pittsburgh Public Schools flags and risk scores using the same percentage cutoffs are similarly accurate

Accuracy means how often the predictions are correct (either correctly predicting there will be a problem or correctly predicting there will not be a problem). The risk scores using the same percentage cutoffs and the PPS flags both identify the same percent of students; therefore, a comparison of the accuracy between these two is the fairest approach. The PPS flags and the same percentage cutoffs are similarly accurate for all outcomes (figure 1). The approaches likely perform relatively similarly because they are identifying an overlapping group of students. For all of the outcomes, more than 60 percent of the students flagged by each system are flagged by the other system.

Figure 1. Percent of predictions that correctly predict either there will or will not be a problem, 2016/17



GPA is grade point average. PPS is Pittsburgh Public Schools.
Note: Low GPA only includes high school students. Differences are statistically significant.
Source: Authors' analysis of data from Pittsburgh Public Schools and Allegheny County Department of Human Services for the 2016/17 school year.

district without the resources to serve so many students. Given the funding realities, the findings for optimal cutoffs are not included in the main report but are presented in the appendix for completeness.

[4] There is no standard in the literature for differences that are big or meaningful. Based on the differences observed in the data and subject matter expertise, the authors chose this threshold.

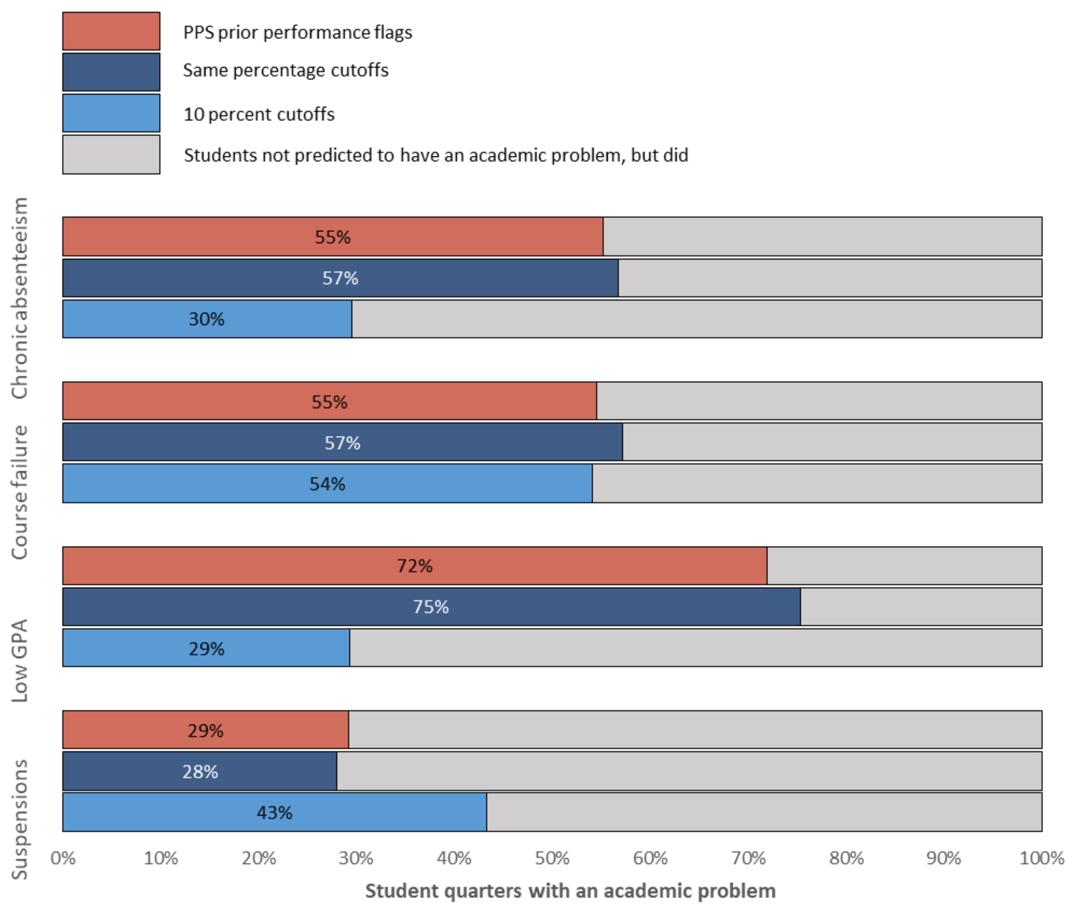
The Pittsburgh Public Schools flags and risk scores with the same percentage cutoffs identify more students who ultimately have an academic problem than does the risk score with the 10 percent cutoffs—except for suspensions, where the risk score with 10 percent cutoffs does better

Another way to think about which system is better is to ask: Of the students who will ultimately have an academic problem, what percent were identified by each prediction? In other words, are at-risk students falling through the cracks? This percent can be calculated for the PPS flags, the same percentage risk score cutoffs (that identify the same percentage of students as the PPS flags), and the 10 percent cutoffs (that identify the top 10 percent of students with the highest risk scores).

The PPS flags and the same percentage cutoffs identify more students who ultimately experience academic problems than the 10 percent cutoffs for chronic absenteeism and low GPA (figure 2). The PPS flags identify 55 percent of students who ultimately are chronically absent, compared to the 57 percent for the same percentage cutoffs and 30 percent for the 10 percent cutoffs. The difference is greater between the 10 percent cutoffs and the other two approaches for low GPA. For suspensions, the 10 percent cutoffs perform better than the other two by 14 to 15 percentage points. All three perform similarly for course failure.

The 10 percent cutoffs might perform best for suspensions because only around 5 percent of students are actually suspended each year. Similarly, the 10 percent cutoffs likely perform relatively poorly for chronic absenteeism and low GPA because there are relatively more students who experience these academic problems (27 percent and 33 percent, respectively).

Figure 2. Among student-quarters *with* an academic problem, percent correctly predicted to have an academic problem, 2016/17



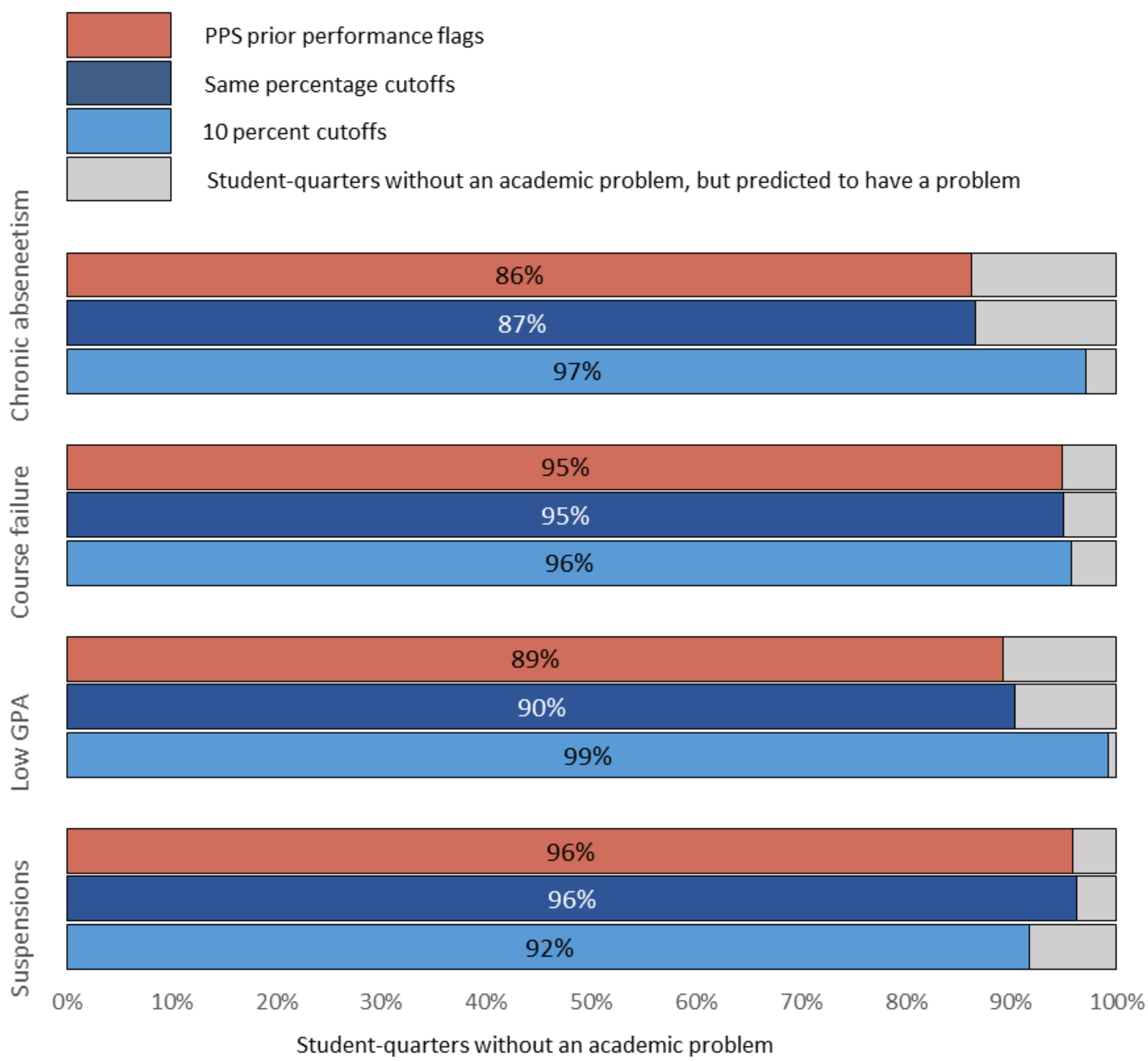
GPA is grade point average. PPS is Pittsburgh Public Schools.
Note: Low GPA only includes high school students.
Source: Authors' analysis of data from Pittsburgh Public Schools and Allegheny County Department of Human Services for the 2016/17 school year.

The risk scores with 10 percent cutoffs correctly identify more students that will not be chronically absent or have a low grade point average

A third way to think about which system is best is to ask: Among those students who do not ultimately have an academic problem, how many were correctly predicted to not be at risk? In other words, are lower-risk students receiving extra resources and supports that they do not need? (See figure 3.)

The 10 percent cutoffs perform best for chronic absenteeism and GPA, likely because this approach identifies the fewest percent of student-quarters. The two risk score cutoffs and the PPS flags perform similarly for course failure. The same percentage cutoffs and the PPS flags perform similarly for each outcome. For suspensions, the same percentage cutoffs and the PPS flags correctly identify more students that will not have an academic problem.

Figure 3. Among student-quarters *without* an academic problem, percent predicted to not have an academic problem, 2016/17



GPA is grade point average. PPS is Pittsburgh Public Schools.
Note: Low GPA only includes high school students.
Source: Authors' analysis of data from Pittsburgh Public Schools and Allegheny County Department of Human Services for the 2016/17 school year.

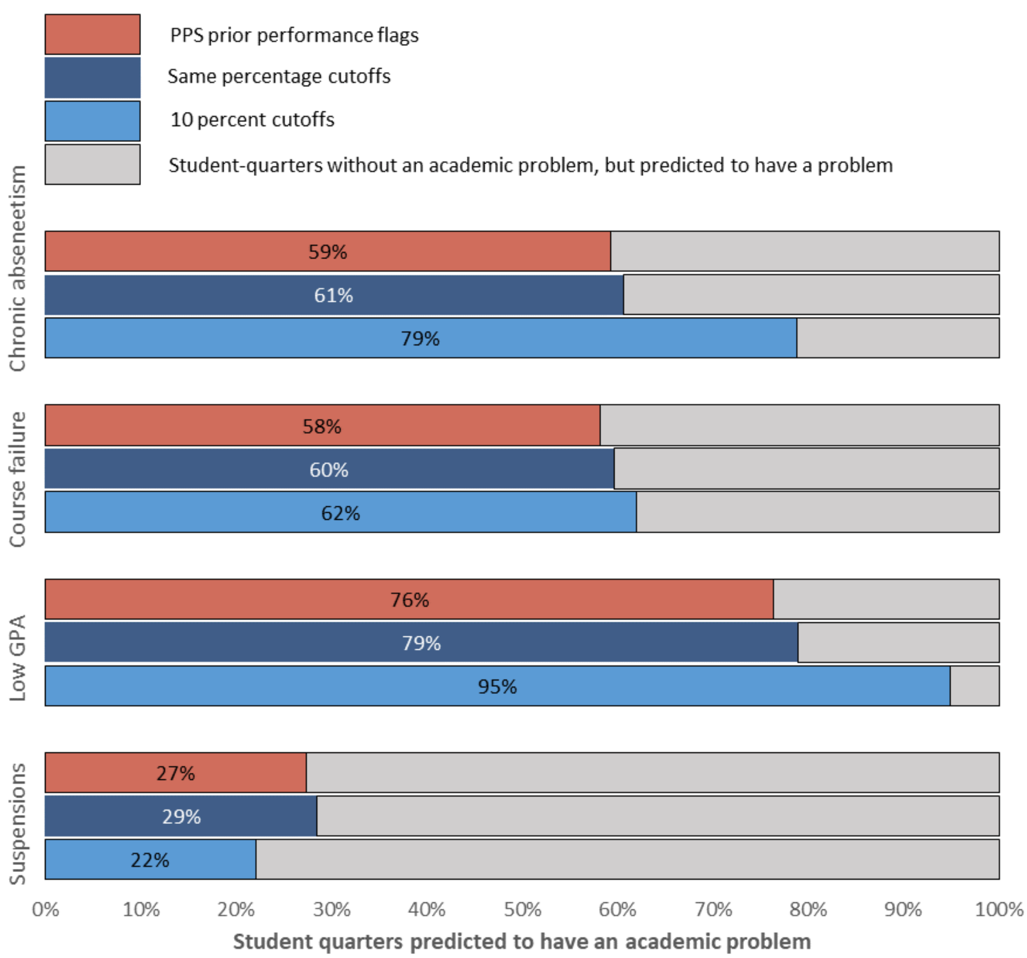
The risk scores with 10 percent cutoffs successfully targets the students most likely to be chronically absent, fail a course, or have a low grade point average

Finally, districts might also wonder: Of the students identified and then provided with extra resources and supports, how many actually need those resources and supports? In other words, are the resources expended going towards students who actually need them? This is a question about the efficient use of resources and also about minimizing potentially stigmatizing interventions for those who don't need them.

Comparing across the two risk score cutoffs and the PPS flags helps answer these questions. The 10 percent cutoffs identify the students at the highest risk, so this approach usually identifies more student-quarters in which an academic problem is observed (figure 4). For chronic absenteeism, course failure, and low GPA, most of the student-quarters identified by the 10 percent cutoffs ultimately have an academic problem (79 percent, 62 percent, and 95 percent, respectively). For both chronic absenteeism and low GPA, the PPS flags and the same percentage cutoffs perform more than 15 percentage points worse than the 10 percent cutoffs. However, all three approaches perform similarly for course failure—among the student-quarters predicted to have academic problem, between 58 and 62 percent of the student-quarters actually experience an academic problem.

Only 5 percent of students are suspended each quarter, and all the approaches substantially over-identify students who will be suspended. Only 27 percent of those identified by the PPS flags, 29 percent identified by the same percentage cutoffs, and 22 percent identified by the 10 percent cutoffs were actually suspended.

Figure 4. Among student-quarters *predicted* to have an academic problem, percent that actually had an academic problem 2016/17



GPA is grade point average. PPS is Pittsburgh Public Schools.
Note: Low GPA only includes high school students.
Source: Authors' analysis of data from Pittsburgh Public Schools and Allegheny County Department of Human Services for the 2016/17 school year.

Both the risk scores and the Pittsburgh Public Schools flags are less accurate when predicting outcomes for students who are Black than for students who are White[5]

In addition to considering the overall accuracy, districts might also want to know how accurate each method is for different groups of students. An ideal early warning system would be equally accurate among different groups of students. Accuracy among Black students and White students may be of particular interest because these are the two largest racial or ethnic groups in PPS. The percentages of student-quarters with a wrong prediction, broken down by Black and White students, are shown in figure 5. For all outcomes, all of the prediction approaches have a higher percentage of wrong predictions for students who are Black compared to students who are White. In other words, Black students are both more likely to be over-identified (predicted to have a negative outcome that does not occur) and more likely to be under-identified (predicted not to have a negative outcome that does occur) (see figures B3 and B4 in appendix B). This does not mean the PPS flags or the predictive algorithms are biased against Black students; it only means that outcomes for Black students are harder to predict from existing data. Below there is a discussion of potential reasons for this finding.

Comparing across outcomes gives some insight into which outcomes are the most difficult to predict for Black students. The Black–White gap in the percent of wrong predictions is smallest for chronic absenteeism and highest for low GPA. Course failure and suspensions fall in the middle. The Black–White gap is more than 5 percentage points for all predictions for course failure, low GPA, and suspensions.

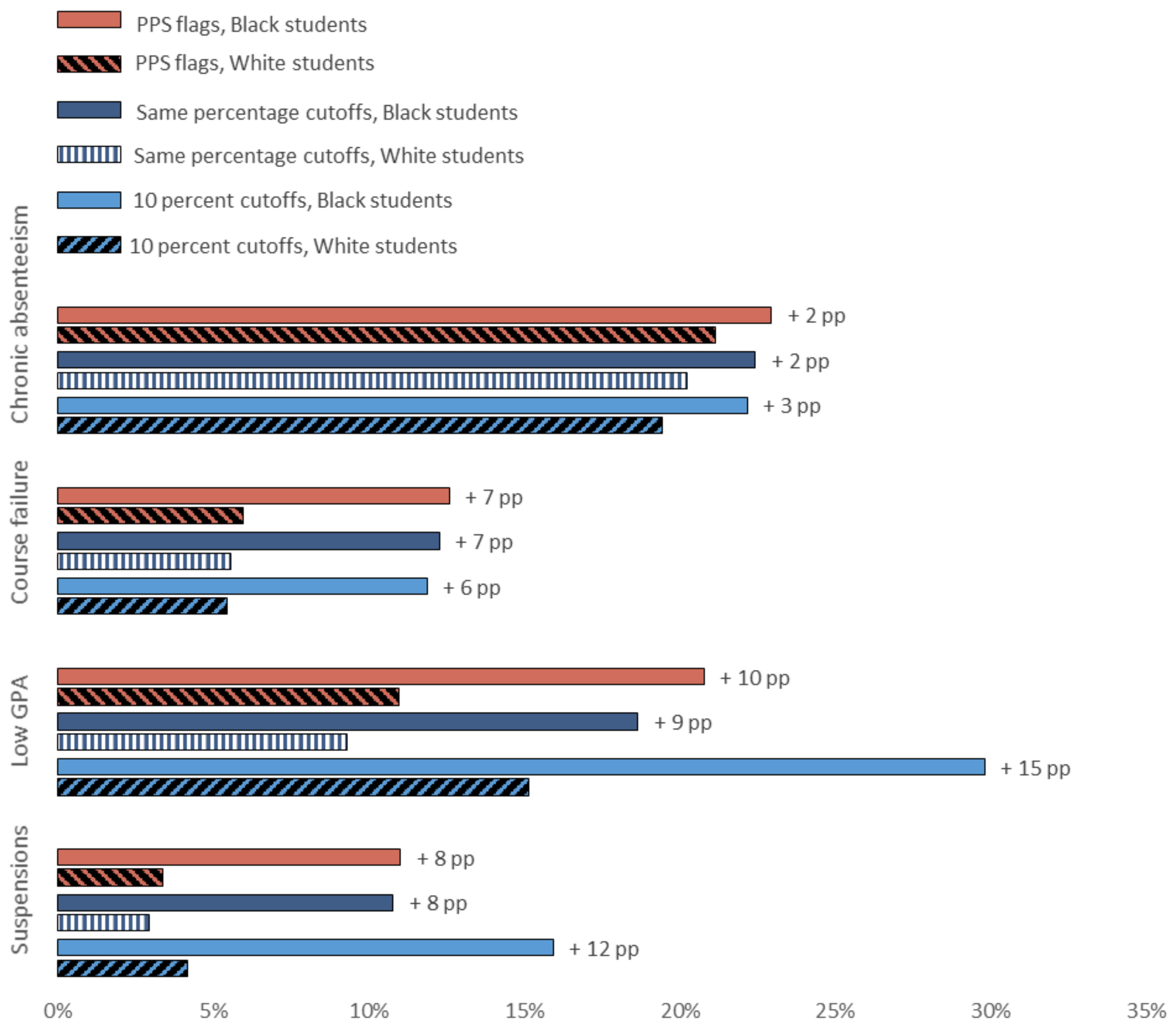
Comparing within outcomes and across prediction approaches reveals some differences between the approaches. For chronic absenteeism, all of the options incorrectly predict more Black students than White students (a difference of 2 to 3 percentage points). For course failure, all of the options incorrectly predict more Black students (a difference of 6 to 7 percentage points). The PPS flags and the same percentage cutoffs have similar race differences for low GPA and suspensions (8 to 10 percentage points more incorrect predictions for Black students than White students), but the 10 percent cutoffs have larger race differences. For low GPA, there is a 15-percentage point difference in inaccurate predictions; for suspensions, there is a 12-percentage point difference.

Predictions for students who are Black may be less accurate because the model might exclude potentially relevant predictors for these students. The study did not have access to a host of data on other issues and events that affect children—including health issues, and issues and events affecting parents—and therefore these data are not included in the model. If the model is missing data on factors that are particularly important for Black students, then the predictions from the current model will be less predictive for Black students than their peers. Indeed, a regression analysis revealed that, for course failure and low GPA, all three prediction approaches accounted for more of the variation in outcomes among White students than Black students (see table B2 in appendix B). However, this pattern was reversed for chronic absenteeism, where all three prediction approaches accounted for more of the variation in chronic absenteeism among Black students than White students. For suspensions, all the prediction approaches had only limited predictive power.

---

[5] The appendix includes a detailed breakdown of the accuracy of each approach for predicting academic problems. Information is presented overall for each outcome and broken into the following student groups: elementary school students, middle school students, high school students, students who are Black, students who are White, male students, female students, students eligible for free or reduced-price lunches, students not eligible for free or reduced-price lunches, students with DHS involvement, and students without DHS involvement (see tables B3–B6 in appendix B). A breakdown of the wrong predictions by race is available in figures B3 and B4.

Figure 5. Percent of student-quarters with a wrong prediction, by race, 2016/17



GPA is grade point average. pp is percentage point. PPS is Pittsburgh Public Schools.

Note: Low GPA only includes high school students. Differences between students of different races are statistically significant for each prediction method and each outcome.

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

Implications

In summary, the findings suggest two key takeaways for districts:

- **Because the accuracy of the PPS flags and the algorithm using the same percentage cutoffs is similar, decisions about which to use should be driven by district circumstances and preferences beyond accuracy.** Both approaches identify the same number of students as at risk (by design), and accuracy is similar across outcomes. The PPS flags only use prior performance, and the same percentage cutoffs use a model with many other in- and out-of-school predictors. This result suggests that the additional out-of-school predictors do not substantially improve the prediction in most cases.

- **When selecting the risk score cutoff thresholds, districts must make a choice between over-including students and under-including students.** Unlike simple binary flags, the algorithm distinguishes degrees of risk among students and a cutoff must be used to divide students into high- and low-risk groups. This can allow PPS to set cutoffs that vary depending on the prevalence of different outcomes, the harms of over-identifying versus under-identifying students at risk, and the resources available to support interventions. For example, if PPS is especially interested in identifying more students who are likely to be suspended, risk scores from the algorithm (unlike flags based on prior suspensions) would enable it to do so. However, there is a tradeoff: widening the net to capture more of the students that will experience an academic problem means also capturing more students that will not experience an academic problem. The opposite is also true: a cutoff that is designed to narrowly target students most likely to experience an academic problem will miss some students that experience an academic problem.

A simple prior performance prediction system might make sense for some districts. There are significant costs to set up and maintain a predictive risk score early warning system, so school districts with limited resources might prefer to use a system like the PPS flags.[6] Furthermore, districts need to consider if they have the high-quality data required to create a predictive risk score early warning system (appendix A details the data). Without the right data, districts would need to rely on a simple prediction based on the prior term. If districts choose to use out-of-school data, there may be challenges working with local agencies to access and store data. This study suggests that a simple prior performance system based solely on in-school data would be just as accurate as the cutoffs examined from the predictive risk score early warning system for many outcomes.

For districts that can implement a predictive risk score early warning system, the system presents several important advantages. A system like the PPS flags that produces binary indicators does not allow the district to distinguish the degree of risk among different students or control the number of students identified based on the relative harms associated with over-identifying versus under-identifying students at risk. In contrast, continuous risk scores rank all students by their risk. Districts can use this information in multiple ways—for example, to implement different interventions for students at different risk levels or to target resources to the highest-risk students.

If districts choose to use a predictive risk score early warning system, there are multiple factors to consider when choosing the cutoffs. An example of the factors to consider is shown in figure 6. It includes a few sample situations and the resulting cutoff conclusions. This flow chart is a simplified example; in reality, the decision will be more complex. Districts will also have to consider other factors (such as staff qualifications and capacity to implement the intervention) and gather information (such as cost and effectiveness of planned interventions that would be assigned to students who meet a certain risk score). Districts will have to consider these questions separately for each outcome because the answers could differ by outcome and the corresponding choice of intervention.

First, a district should consider whether it should under-include students (that is, should it err on the side of missing some students who are at risk rather than include some students who are not). If the intervention could be viewed as stigmatizing, intrusive, or punitive, it might be preferable to narrowly identify students who are most likely to have academic problems. It might also be preferable to under-include students if the planned intervention might take students away from other academic opportunities. To under-include students, the district would need to select a cutoff lower than the prevalence rate.

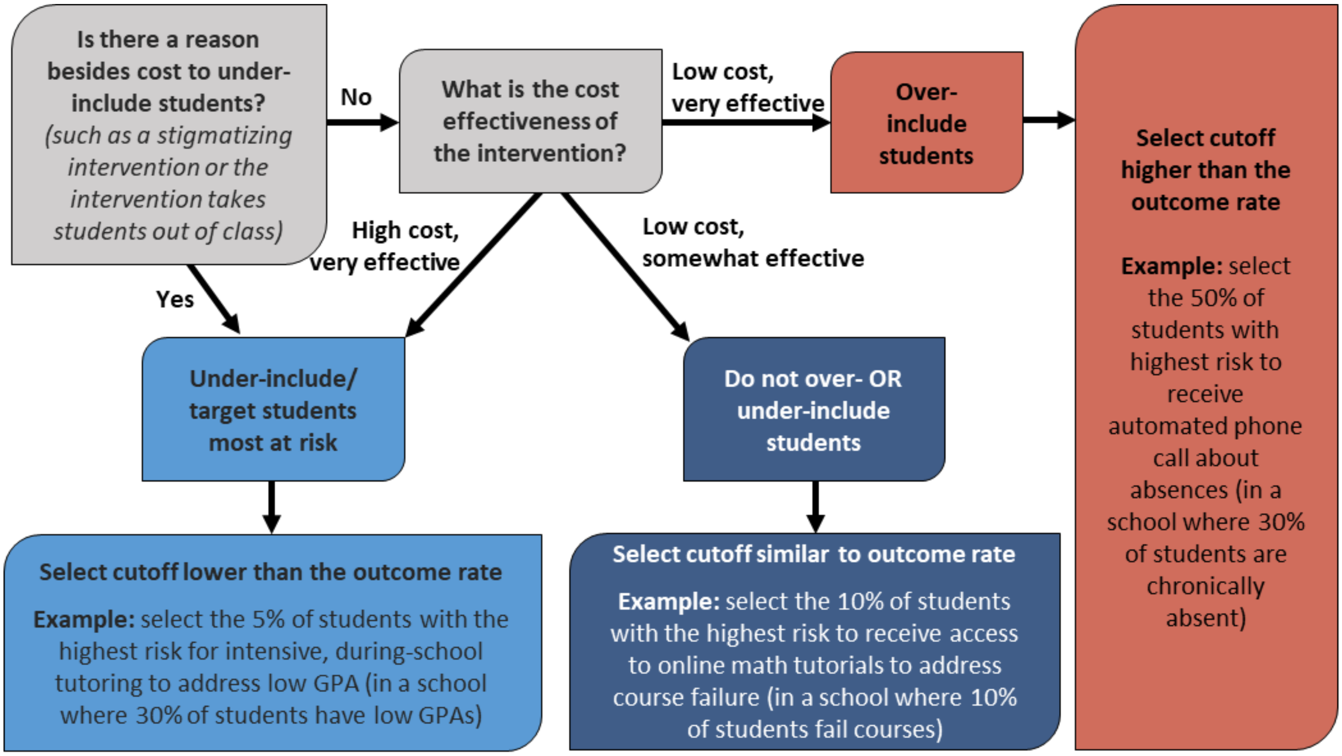
The next consideration is the cost effectiveness of the planned intervention.[7] If the intervention is high cost, districts might want to under-include students by selecting a cutoff lower than the prevalence rate. If the

[6] It would likely take significant staff and computer resources to gain access to out-of-school data, prepare the data, and create the prediction model at the beginning. Then, on an ongoing basis, it would take staff and computer resources to continually re-run the model and generate new predictions for students. Districts may have staff internally with this expertise or may need to partner with an external vendor or local agency.

[7] If districts need help finding effective interventions, the What Works Clearinghouse has developed practice guides and reviewed the quality of research specific to a wide range of interventions (see https://ies.ed.gov/ncee/wwc/).

intervention is low cost, the district can consider effectiveness of the intervention and err on the side of over-including students for more effective low-cost interventions.

Figure 6. Guide for districts in deciding risk score cutoffs



GPA is grade point average.
Source: Authors' creation.

Limitations

This analysis is based on data from the 2016/17 school year for PPS. Policies on absences, grading, GPA calculation, and suspensions vary by district, so these findings may differ for districts with significantly different policies than PPS. Further, policies change over time, so these results may not even hold for PPS if their policies have changed significantly. A dynamic system that creates the risk scores and updates the analyses on a regular basis would address this limitation for PPS because, as the policies changed, so would the data (though there would be a lag).

The analysis sample was limited to students who met an enrollment threshold. This accounted for students who transferred schools (sometimes within the district) just a few days into the quarter. It also made the outcomes more meaningful. For example, a student enrolled five days for the entire quarter and who missed three days would be considered chronically absent, but the exclusion criteria removed these types of student-quarters from the analytic sample. However, it is likely that the enrollment threshold excluded students with a higher than average risk for academic problems. These students are not included in the risk score creation or in this analysis; this means the results may not be relevant for students enrolled for very short periods.

The model used to create the risk scores included data only from the quarter immediately prior. If districts have access to data from other prior quarters (such as the quarters from the prior school year), districts could consider adding in other additional variables from previous quarters. These variables may improve prediction accuracy.

There are also many cutoffs that could be used to cut the risk scores—this analysis considered only two options. Another cutoff may better balance PPS's priorities for identifying enough students correctly while, at the same time, targeting resources to the students who need support. Districts could also consider using the risk scores to create tiers of risk and provide a different level of services to students at different tiers.

These results may also not be completely applicable to districts still dealing with the COVID-19 pandemic and its aftermath. The predictions were developed with data prior to the pandemic and therefore may not account for different factors impacting student outcomes. For example, student anxiety may still be high from the pandemic and may result in more behavior issues than the data used in this study would suggest.

References

Allegheny County Department of Human Services. (2014). *Data brief: Addressing school absenteeism.* https://www.alleghenycountyanalytics.us/wp-content/uploads/2016/06/Data-Brief-Addressing-School-Absenteeism.pdf

Allegheny County Department of Human Services. (2015). *Data brief: The impact of child welfare system involvement on school stability.* https://www.alleghenycountyanalytics.us/wp-content/uploads/2016/06/DATA-BRIEF-The-Impact-of-Child-Welfare-System-Involvement-on-School-Stability.pdf

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.* 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

Bruce, M., Bridgeland, J. M., Hornig Fox, J., & Balfanz, R. (2011). *On track for success: The use of early warning indicator and intervention systems to build a grad nation.* 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

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. https://files.eric.ed.gov/fulltext/ED606230.pdf

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/?q=Mandated+engagement%3a+The+impact+of+early+college+high+schools.+&id=EJ1020156

Pittsburgh Public Schools. (2020). *Attendance summary: District-wide rates.* https://www.pghschools.org/Page/5075

Pittsburgh Public Schools. (2021). *Out of school suspensions: District-wide rates.* https://www.pghschools.org/Page/5070

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.

U.S. Department of Education. (2016). *Issue Brief: Early warning systems.* Office of Planning, Evaluation and Policy Department; Policy and Program Studies Service. https://www2.ed.gov/rschstat/eval/high-school/early-warning-systems-brief.pdf

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

REL 2021–126

September 2021

This report was prepared for the Institute of Education Sciences (IES) under Contract ED-IES-17-C-0006 by 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.

Cattell, L., & Bruch, J. (2021). *Identifying students at risk using prior performance versus a machine learning algorithm* (REL 2021–126). U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Mid-Atlantic. http://ies.ed.gov/ncee/edlabs

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

Identifying Students At Risk Using Prior Performance Versus a Machine Learning Algorithm

Appendix A. Methods

Under the previous study (Bruch et al., 2020), the study team collected and linked five academic years of student-level administrative data from Pittsburgh Public Schools (PPS) and the Allegheny County Department of Human Services (DHS). The sample included the full population of students enrolled in 2015/16 or 2016/17, and each entity provided any data available on those students for 2012/13 through 2016/17.

This study extends that work by retroactively creating PPS flags for the school year 2016/17. The current study team also considered alternative cutoffs for the continuous risk scores beyond the optimal cutoffs based on the Youden statistic used in the prior study. This analysis focused on comparing the PPS flags to each of the cutoffs for the risk scores, examining which students are identified by each system and comparing accuracy rates.

Data acquisition

PPS generated lists of unique state identification numbers (PASecureID) associated with all students enrolled in each local education agency during 2015/16 or 2016/17, which defined the sample. PPS provided the lists to Allegheny County DHS, which compiled files with historical data on each student in the sample. PPS and DHS then provided the previous study team with the data associated with each student from school years 2012/13 through 2016/17, identified by PASecureID. DHS data included the entire five-year period (2012/13–2016/17) for each student, regardless of dates of enrollment in PPS schools. The data did not include student names, birthdays, addresses, or social security numbers, but the previous study team took steps to protect the data given that they included PASecureIDs.

Data elements

PPS provided data for each student on the academic problems examined in the study—absences, suspensions, course performance, and state test performance—as well as demographic characteristics and indicators of eligibility for school services (table A1). The agencies provided some data elements on an annual timescale and some on more granular levels—semesters, quarters, or event dates—in separate files. Allegheny County DHS provided data on its services, justice system involvement, and receipt of public benefits for students in the sample over the five-year study period (table A2).

Table A1. Data elements provided by Pittsburgh Public Schools, 2012/13–2016/17

Types of data elements	Timescale of data
Demographics (race/ethnicity, gender)	Annual
Economic disadvantage[a]	Annual
Grade level	Semester
English learner status	Event dates
Special education status[b]	Event dates
Gifted status	Annual
Type of disability	Annual
Type of absence	Event dates
Behavior incidents and reasons for suspension	Event dates
Course grades and cumulative grade point average	Quarter
State exam score[c]	Annual
School enrollment and withdrawal events and reasons	Event dates

a. Based on eligibility for the national school lunch program.
b. Whether student has an Individualized Education Program.
c. The Pennsylvania System of School Assessment for elementary and middle school students and Keystone exams for high school students.
Source: Authors' compilation.

Table A2. Data elements provided by Allegheny County Department of Human Services, 2012/13–2016/17

Type of data element	Timescale of data
Social services	
Child welfare services[a]	
Home removal episodes	Event dates
Type of child welfare placements	Event dates
Nonplacement services	Event dates and monthly
Other social services	
Type of behavioral health service	Event dates
Type of homeless or housing service	Event dates
Head Start service	Monthly
Low-Income Home Energy Assistance Program	Monthly
Justice system	
Juvenile and adult justice system involvement	
Active cases	Event dates and monthly
Adjudication of cases	Event dates
Jail bookings and stays	Event dates
Family court involvement	
Active cases	Event dates and monthly
Type of family court events	Event dates
Public benefits	
HealthChoices[b]	Monthly
Supplemental Nutrition Assistance Program	Monthly
Temporary Assistance for Needy Families	Monthly
Public housing and section 8 housing vouchers	Monthly

Note: While the previous study team received data beginning with 2012/13, the primary predictive model used data starting in 2014/15. The earlier 2012/13 data were used for testing alternative models.
a. Services were further differentiated by service placement starts or stops, ongoing placement, or removal episodes.
b. Pennsylvania's managed care program for individuals eligible for Medicaid.
Source: Authors' tabulation based on data from Allegheny County Department of Human Services for 2014/15–2016/17.

Data preparation

The previous study team assessed the completeness and quality of the data and then used students' unique state identification numbers (PASecureID) to link school and Allegheny County DHS data. The linked data were used to prepare outcome and predictor data files for the descriptive and predictive analyses.

Outcome data. The previous study team examined four types of academic outcomes (table A3). Chronic absenteeism and suspensions are calculated by aggregating to a particular time period from the raw (event) data, based on the description in table A1. The outcome period was school year 2016/17 for which a student was enrolled at least half the time. Each outcome is binary; in other words, taking a value of 1 if the outcome occurred for the student in the given time period and taking a value of 0 otherwise.

Predictor data. The previous study team defined the predictor period (the time period over which predictors are measured) for each observation to be the period of time immediately preceding—and not overlapping with—the outcome period. The previous study team created data files of predictor variables aggregated to appropriate time periods for each outcome. For chronic absenteeism, suspensions, course failure, and low grade point average (GPA), which capture performance over a fixed academic period, the predictor period is the academic period of the same length as the outcome period that immediately precedes it. In most cases, this is the preceding quarter. This approach makes it possible to examine the relationships between predictors and outcomes in adjacent periods of equal lengths.

For examining absences and suspensions as predictors, the analysis excluded observations for students who were not enrolled for at least 50 percent of the school days in the predictor period (in addition to the restriction for the outcome period). This is because these predictors are counts of events or percentages of possible days on which events occurred, and they are highly related to the number of days enrolled.

In many cases, the timing of observation of the predictor is not directly aligned with that of the outcome. For example, suspensions are defined at the quarter level, but Allegheny County DHS predictors are measured monthly or as date-specific events. In these cases, the previous study team aggregated the predictors to the appropriate level for each outcome using sensible rules. Monthly flags and events (such as for receiving DHS services) were recalculated at the term level, indicating whether the service or event occurred during any month that overlapped that particular term. The aggregation approach was defined for each predictor.

Analytic sample

After the data preparation stage, the Bruch et al. (2020) team created a separate analytic sample file for each outcome. The final sample sizes for each analysis are shown in table A3 for both the number of unique students and the total number of observations. The number of observations varied for each outcome.

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 per term. This method accounted for students who transferred schools (sometimes within the district) just a few days into the quarter. It also made the outcomes more meaningful. For example, a student enrolled five days for the entire quarter and who missed three days would be considered chronically absent, but the exclusion criteria removed these types of student-quarters from the analytic sample. However, it is likely that the enrollment threshold dropped students with a higher than average risk for academic problems.

Students for whom grade-level information was missing were not included in any of the results that are separated by grade span, but they are included in other results. Thus, the total number of observations for any outcome will not be equal to the sum of the observations from elementary, middle, and high school. In addition, the total number of unique students is not equal to the sum of the number of unique students in each grade span because some students are associated with multiple grade levels, even within a single academic year.

Table A3. Sample size in Pittsburgh Public Schools, 2016/17 (number of student-quarters)

Type of observation	Pittsburgh Public Schools			
	Elementary school	Middle school	High school	Total
Chronic absenteeism	43,326	19,427	24,688	87,441
Suspensions	43,326	19,427	24,688	87,441
Course failures	34,770	18,960	23,395	77,125
Low grade point average	na	na	23,415	23,415
Number of unique students	13,086	6,580	7,081	23,848

na is not applicable.
Note: See table 1 in the main report for definitions of outcomes. The sample includes all observations during academic terms for which the student was enrolled for at least 50 percent of possible days in Pittsburgh Public Schools during the 2016/17 school year. Grade ranges are K–5 for elementary school, 6–9 for middle school, and 9–12 for high school.
Source: Authors' calculations using data from Pittsburgh Public Schools for the 2016/17 school year.

Composition of sample

Below are data on the characteristics of students in the descriptive analysis sample (table A4) and the frequency and duration of Allegheny County DHS involvement for this sample (table A5).

Table A4. Demographic characteristics and school service eligibility of sample, 2016/17 (percent of student-quarters)

Student characteristic or school service eligibility	Percent of sample (n = 275,422)
Student characteristic	
Gender	
Male	50.36%
Female	49.64%
Race/ethnicity	
American Indian and Native Hawaiian/Pacific Islander	0.29%
Black	52.70%
Hispanic	2.97%
Multiracial	7.46%
White	33.20%
School service eligibility	
Economic disadvantage (eligible for national school lunch program)	61.61%
In special education (has an Individualized Education Program)	17.58%
Eligible for English as a second language services	2.83%

Note: Table includes all student-quarters in the descriptive analysis sample.
Source: Authors' analysis of administrative data from Pittsburgh Public Schools in 2016/17.

Table A5. Frequency and duration of student involvement with Allegheny County Department of Human Services, 2016/17 (percent of student-quarters)

Type of involvement	Percent of sample	Mean duration of involvement (standard deviation)	Median duration of involvement (interquartile range)
Duration measured in days			
Behavioral health services			
Outpatient behavioral health services	8.71	6.36 (8.177)	4 (2–7)
Counseling services	2.31	13.10 (10.598)	11 (5–18)
Inpatient behavioral health services	0.22	16.19 (17.218)	11 (6–17)
Duration measured in months			
Child welfare services			
Child welfare nonplacement services	1.57	2.99 (0.981)	3 (3–4)
Child welfare placement services	0.13	1.00 (0.000)	1 (1–1)
Housing and family support services			
Any homeless service	1.39	1.00 (0.000)	1 (1–1)
Any homeless service started	0.37	1.00 (0.000)	1 (1–1)
Emergency shelter assistance	0.20	1.00 (0.000)	1 (1–1)
Rental assistance and prevention	0.44	1.45 (0.901)	1 (1–1)
Head Start	0.01	2.36 (0.780)	2 (2–2)
Energy assistance[b]	0.00	3.29 (0.488)	3 (3–4)
Justice system involvement			
Active case in family court	0.69	1.06 (0.262)	1 (1–1)
Active case in the juvenile justice system	1.37	1.00 (0.000)	1 (1–1)
Time spent in county jail	0.05	1.60 (0.956)	1 (1–2)
Adult probation	0.04	7.56 (5.072)	6 (4–12)
Public benefits			
HealthChoice[c]	57.23	3.39 (0.641)	3 (3–4)
Supplemental Nutrition Assistance Program	21.37	2.94 (0.981)	3 (2–4)
Temporary Assistance for Needy Families	9.24	3.09 (0.896)	3 (3–4)
Section 8 housing choice voucher program	13.93	3.42 (0.616)	3 (3–4)
Low-income public housing	5.75	3.42 (0.606)	3 (3–4)

Note: Table includes all student-quarters in the descriptive analysis sample.

a. Calculated based on students who have any involvement during the two-year period (excluding zeroes for students with no involvement).

b. Low-Income Home Energy Assistance Program.

c. Pennsylvania's managed care program for individuals eligible for Medicaid.

Source: Authors' analysis of administrative data from Allegheny County Department of Human Services for the 2016/17 school year.

Prevalence of academic problems in the sample

The prevalence of academic problems varied across outcome. For approximately one-third of student-quarters in PPS, students had a GPA below 2.2, a threshold identified by the stakeholders to identify at-risk students (table A6). Other outcomes examined in this study—including chronic absenteeism, suspensions, and core course failure—were less frequent. High school students experienced academic problems more frequently than did elementary school students.

Table A6. Frequency of outcomes in 2016/17 (percent of student-quarters)

Outcome	Pittsburgh Public Schools: Grades K–8	Grades 9–12	All grades (including those with missing grades)
Chronic absenteeism	21%	35%	27%
Suspensions	5%	7%	5%
Course failures[a]	6%	21%	11%
Low grade point average[b]	na	31%	33%

na is not applicable.

Note: Table includes all students in the analysis sample. Percentages of chronic absenteeism, suspensions, and low grade point average represent proportion of student-quarter with each outcome during 2016/17.

a. Proportion of student-quarters with a failing grade in any course.

b. Proportion of student-quarters where grade point average was less than 2.2.

Source: Authors' analysis of data from Pittsburgh Public Schools for the 2016/17 school year.

Predictive modeling methods

The previous study team built a predictive model to calculate a risk score for each student for each outcome, with the goal of achieving the best predictions possible. The previous study team decided against linear or logistic regression models, which rely on strong parametric assumptions on the functional form (such as linearity and additivity of the effects of predictors) in favor of more flexible machine learning techniques, which take full advantage of the rich data sources available. Machine learning models use data-driven algorithms designed to extract the most relevant information from a dataset, with a focus on maximizing the predictive performance of the model. They are particularly useful when there is no strong theory to guide the way predictors interact, which is common when data come from multiple, loosely related sources. Machine learning approaches are also advantageous when events occur over time and when complex, long-term dependencies exist between predictors and outcomes. Each of these features characterizes the study data.

The previous study team ultimately found the random forest (RF) machine learning model performed best as measured using the area under the curve (AUC). An RF is an ensemble predictive model that is made up of many decision trees. Like decision trees (commonly known as classification and regression trees, or CART models), random forests can identify nonlinear relationships and interactions between predictors. Because they can fit many decision trees, each constructed slightly differently because of randomness, they tend to be more robust than standard CART models. The previous study team used a grid search and 10-fold cross validation to optimize the tuning parameters.

The input predictors for the RF were taken from the set of aggregated predictors that were used for the descriptive analysis. In some cases, the previous study team used different forms of these predictors than the one used in the descriptive analysis. For example, there were a number of continuous variables that took the value of 0 for most students, indicating that the students never used a particular service or experienced the event. For the descriptive analysis, the previous study team dichotomized these predictors into 0 and greater than 0 because these approaches assume linear trends between the predictors and outcomes that are unlikely to hold in their raw form. For the RF model, dichotomization is unnecessary because the RF algorithm automatically identifies relevant thresholds. Therefore, these variables were included as continuous variables in the RF.

Model validation. To assess how the model will perform on a future dataset, the team trained the model only on data through 2015/16. After training, the model was used to predict risk scores for all outcomes in 2016/17, allowing for testing model performance on new data. To assess model performance in 2016/17, the predicted probabilities returned by the model can be compared with the actual outcomes for students in the sample.

Creation of analysis file for the current study

Converting the student-quarter-course risk scores to student-quarter risk scores. The prior study calculated a risk score for failing each student-quarter-course. The PPS flags are calculated at the student-quarter level; in other words, the PPS system predicts whether a student will fail any course in a given quarter, rather than a specific course in that quarter. To make a useful comparison, the team created a single risk score for course failure for each student in each quarter. The combined version is the maximum risk score across each course for the student in a given quarter. This is the equivalent of the risk of failing any course in the next quarter.

Creating the PPS flags. The current study team created the PPS flags retroactively for school year 2016/17. PPS explained the definitions of the PPS flags in detail to the current study team. The current study team then defined variables as similarly as possible to the PPS flags while also ensuring comparability to the machine learning *model.* Students missing data needed to create the PPS flags were dropped from the sample. (See table 1 in the main report for the full definitions of the PPS flags.) Additional notes about the PPS flags are below:

- ***Chronic absenteeism:*** *The PPS flags are calculated every day. This means that, in the first few days of the quarter, students absent for just a day can be identified as chronically absent (more than 10 percent of days absent). In practice, PPS does not consider this a real measure until a few weeks into the quarter. For comparison to the risk scores, the outcomes were calculated once at the end of the quarter.*
- ***Course failure:*** *The PPS flags identify students who receive a failing grade in any course, not just core courses. The study team had access to core course data only, so the risk scores were calculated only with core courses. For comparison to the risk scores, the PPS flags were also calculated among core courses only. The course failure risk scores calculated under the original study were at the student-quarter-course level. For comparison to the PPS flag system, the study team aggregated to the student-quarter level by taking the maximum risk score value for all courses for each student in each quarter.*
- ***Low GPA:*** *The PPS flags use a cumulative GPA over lifetime of academic career. PPS and the study team discussed the idea that cumulative GPA was not a good measure to detect near-term academic problems. PPS and the study team agreed that this study should test the use of a term-specific GPA measure. As done with course failure, for comparison to the risk scores, the PPS flags were calculated among core courses only.*

Selecting risk score cutoff points. To turn the risk scores into a binary prediction, the user must choose a cutoff score. Students with risk scores below the cutoff are deemed to be low risk; student with risk scores above the cutoff are at risk of academic problems. This study examined two cutoff options presented in the main report and a third option presented in appendix B:

- Main report
  1. A cutoff that identifies the same proportion of student as at risk as do the PPS flags. This is the most direct comparison to the PPS flags and, therefore, the best test of accuracy across both approaches.
  2. A cutoff that identifies the students with risk scores in the top 10 percent. The study team discussed this cutoff with PPS. Ideally, this cutoff would be the number of students for whom PPS has the resources to provide services. However, due to limited resources, PPS suggested 10 percent as a good starting point to minimize the number of “false positive” students receiving supports (see below). Using the 10 percent cutoffs, they can determine how a cutoff defined by resource limitations would fare compared to the other options for predicting academic problems.
- Appendix B
  1. An optimal cutoff calculated by maximizing the difference between the sensitivity and the specificity (Youden, 1950). (See definitions below.) The optimal cutoffs were re-created for the current study with the current study’s sample.

Creating accuracy assessment variables. The current study team compared the predictions to the actual outcomes to calculate the following variables:

- **True positive:** The percent of student-quarters predicted to have an academic problem that actually had an academic problem.
- **True negative:** The percent of student-quarters not predicted to have an academic problem that actually did not have an academic problem.
- **False positive:** The percent of student-quarters predicted to have an academic problem that actually did not have an academic problem.
- **False negative:** The percent of student-quarters not predicted to have an academic problem that actually had an academic problem.
- **Sensitivity:** The percent of student-quarters that were predicted to have an academic problem among the student-quarters that ultimately had an academic problem. In other words, the true positives divided by the sum of the true positives and the false negatives. (See statistics in figure 2 in the main report.)
- **Specificity:** The percent of student-quarters that were correctly predicted to not have an academic problem, among students without an academic problem, or the true negatives divided by the sum of the true negatives and the false positives. (See statistics in figure 3 in the main report.)
- **False positive rate:** The percent of student-quarters that were incorrectly predicted to have an academic problem, among students without an academic problem, or the false positives divided by the sum of the true negatives and the false positives.
- **Precision:** The percent of student-quarters that actually had an academic problem among the student-quarters predicted to have an academic problem. In other words, the true positive divided by the sum of the true positives and the false positives. (See statistics in figure 4 in the main report.)

A summary of how these variables relate to each other is shown in table A7, which is known as a confusion matrix. These variables were created for all four predictions: risk score optimal cutoffs, *risk score same proportion cutoffs, risk score 10 percent cutoffs, and the PPS flags. These statistics are presented for the entire student groups of interest.*

Table A7. Confusion matrix

		Predicted values		
		Positive	Negative	
Actual values	Positive	True positive (TP)	False negative (FN)	Sensitivity TP/(TP+FN)
	Negative	False positive (FP)	True negative (TN)	Specificity TN/(TN+FP) False positive rate FP/(TN+FP)
		Precision TP/(TP+FP)		

Source: Authors' creation.

Appendix B. Supporting analysis

This appendix includes the results of some additional analyses including a break down of the demographics of each group of students predicted to have an academic problem, the optimal cutoff findings, a deeper analysis of the wrong predictions by race, and the results of the regression analysis.

Demographic breakdown

The current study team calculated rates of predicted problems and actual problems for student groups of interest (table B1). This table includes those predicted to have at least one of the problems or experienced at least one of the problems. The first column shows the characteristics of the student-quarters that experienced academic problems. For example, 54 percent of the student-quarters that experienced an academic problem involve male students. Across the demographics, the PPS flags identify a set of students with similar demographics to the students who actually experienced an academic problem.

The group identified as at risk by the 10 percent cutoffs includes more Black students. Students who are Black make up 73 percent of the student-quarters identified by the 10 percent cutoffs. In comparison, among those identified by the PPS flags, only 64 percent are Black—the same percentage as the percentage of student-quarters that are Black among all students who actually have an academic problem (the pattern is reversed for White students). There is a similar pattern for students who are economically disadvantaged and students who are involved with DHS.

The predictive risk scores also identify more high school students. High school students make up 54 percent of the student-quarters with an academic problem in the sample. Of the students predicted to have an academic problem by the same percentage cutoffs, 59 percent are in high school. Of the students predicted to have an academic problem by the 10 percent cutoffs, 63 percent are in high school.

There are smaller differences (5 percentage points or less) in the composition of the identified at-risk groups based on gender, Individualized Education Program status, and English learner status.

Table B1. Characteristics of student-quarters predicted to have at least one academic problem, all outcomes, 2016/17

Characteristics	Student-quarters that actually have an academic problem	Student-quarters predicted to have an academic problem		
		By the *PPS flags*	By the *same percentage cutoffs*	By the *10 percent cutoffs*
Gender				
Male	54%	55%	55%	59%
Female	46%	45%	45%	41%
Race/Ethnicity				
American Indian and Native Hawaiian/Pacific Islander	0%	0%	0%	0%
Black	64%	64%	67%	73%
Hispanic	2%	2%	2%	2%
Multiracial	7%	7%	7%	7%
White	24%	24%	22%	17%

Characteristics	Student-quarters that actually have an academic problem	Student-quarters predicted to have an academic problem		
		By the *PPS flags*	By the *same percentage cutoffs*	By the *10 percent cutoffs*
Student service eligibility				
Economic disadvantage	75%	75%	78%	81%
In special education	20%	21%	22%	22%
Eligible for English as a second language services	2%	2%	2%	1%
DHS involvement	25%	25%	28%	31%
Grade level				
Elementary	27%	28%	22%	17%
Middle	19%	19%	18%	20%
High	54%	54%	59%	63%
Number of student-quarters	**44,240**	**41,925**	**41,973**	**27,552**

DHS is Allegheny County Department of Human Services. PPS is Pittsburgh Public Schools.
Note: Table includes student-quarters that either actually have (first column) or are predicted to have (next three columns) at least one of the four possible academic problems.
Source: Authors' analysis of data from Pittsburgh Public Schools and Allegheny County Department of Human Services for the 2016/17 school year.

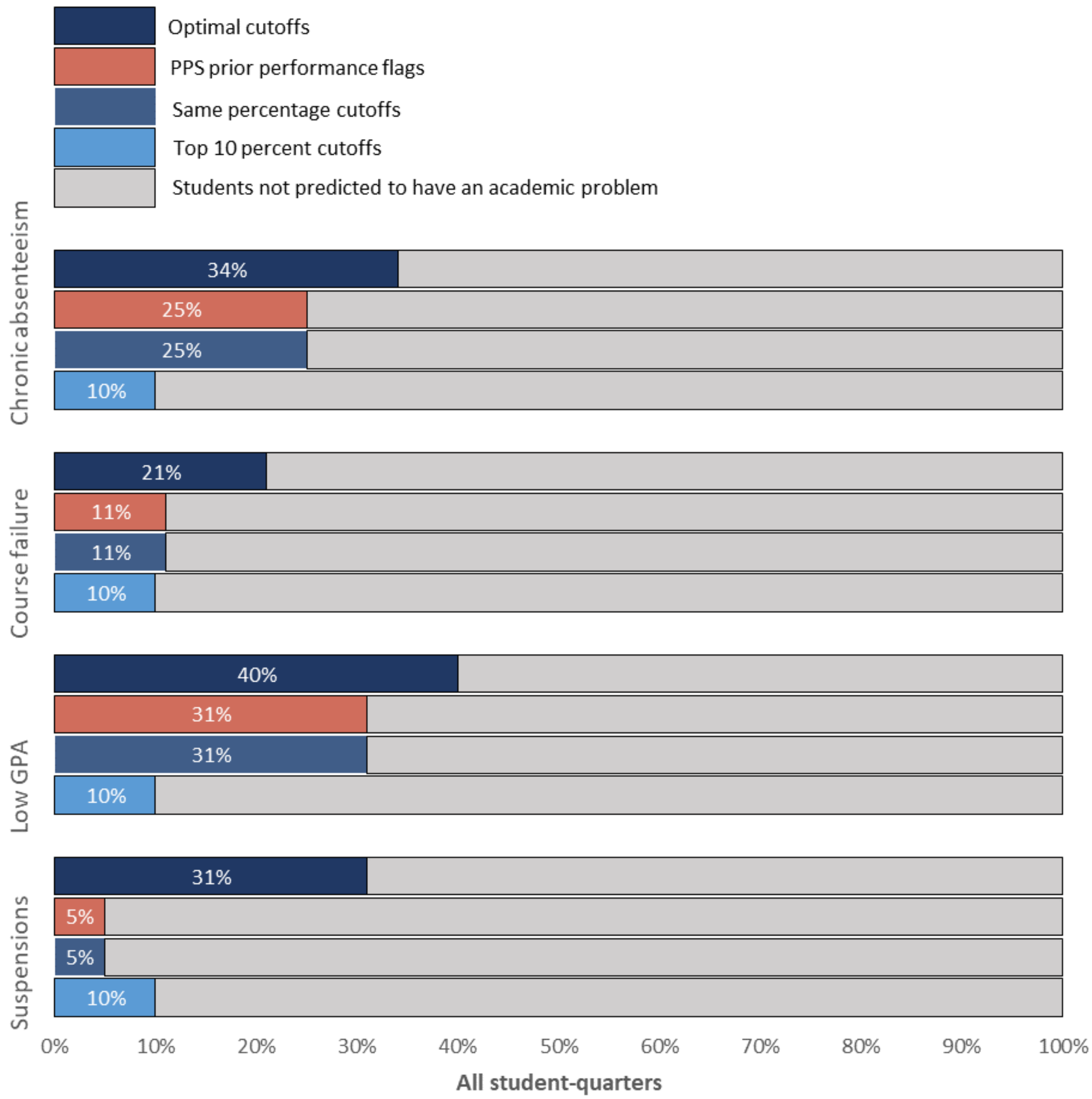
Optimal cutoff findings

The original study focused on a third cutoff option dubbed *the optimal cutoffs*. Those cutoffs were considered optimal because they were statistically calculated to maximize the difference between the sensitivity and the specificity (see the methods in appendix A for more information). The resulting cutoffs cast a wide net and predicted many of the students in the sample would experience an academic problem. In practice, identifying such a large percentage of students is not helpful for a school district without the resources to serve so many students. Given resource constraints, the findings for optimal cutoffs are not included in the main report but are presented here for completeness.

The statistically optimal cutoffs identify more students as at risk

The statistically optimal cutoffs cast a wide net and predict that, depending on the outcome, between 21 percent and 40 percent of all students will have academic problem (figure B1). By design, the 10 percent cutoffs predict 10 percent of students will have an academic problem, and the same percentage cutoffs identify the same proportion of students as the PPS flags. For chronic absenteeism, course failure, and low GPA, the PPS flags and same percentage cutoffs predict fewer students than the optimal cutoffs and more students than the 10 percent cutoffs. Suspensions follow a different pattern: the PPS flags and same percentage cutoffs predict fewer students will have suspensions than the optimal or 10 percent cutoffs.

Figure B1. Percent of student-quarters predicted to have an academic problem, 2016/17

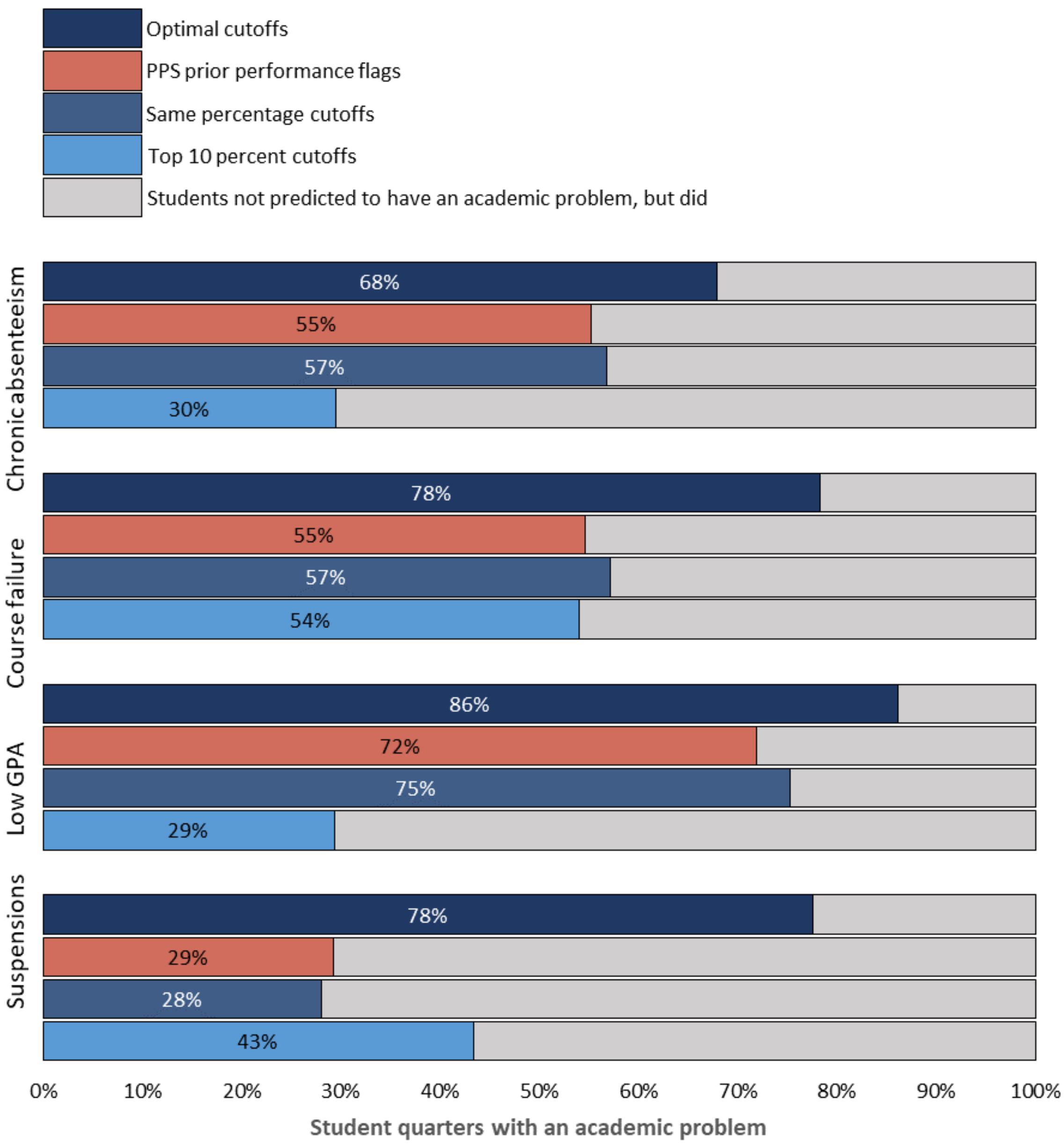


GPA is grade point average. PPS is Pittsburgh Public Schools.
Note: Low GPA only includes high school students.
Source: Authors' analysis of data from Pittsburgh Public Schools and Allegheny County Department of Human Services for the 2016/17 school year.

The statistically optimal cutoffs correctly identify the highest percentage of students that ultimately experience academic problems

The optimal cutoffs identify between 68 and 86 percent of student-quarters in which students actually experience an academic problem (figure B2). The optimal cutoffs are able to identify most of the students who ultimately have an academic problem by casting a wide net and predicting many students will have an academic problem. In comparison, the PPS flags and the same percentage cutoffs miss slightly more of the students who experience academic problems, and the 10 percent cutoffs miss most of the students who experience academic problems .

Figure B2. Among student-quarters *with* an academic problem, percent correctly predicted to have an academic problem, 2016/17



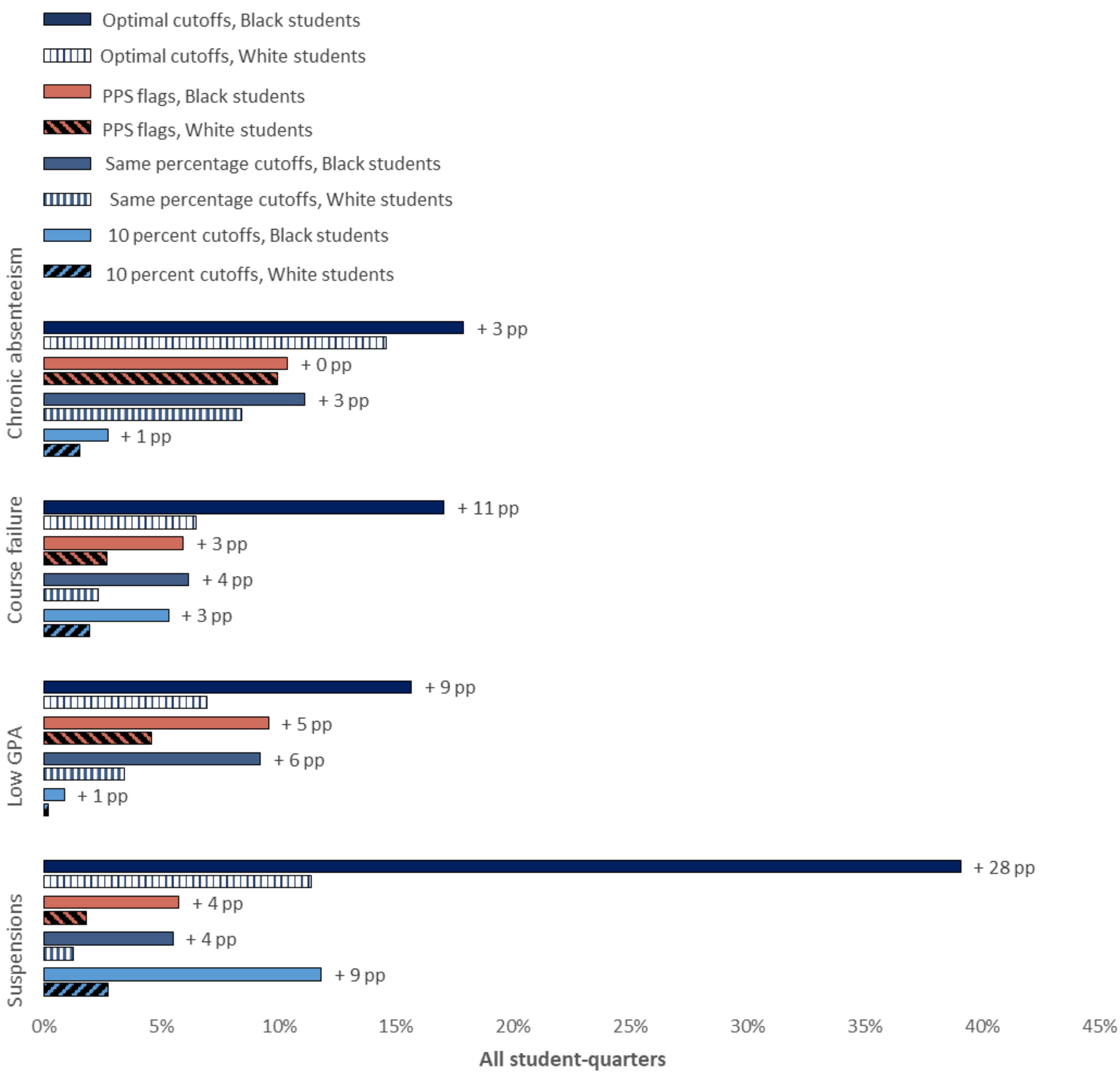
GPA is grade point average. PPS is Pittsburgh Public Schools.
Note: Low GPA only includes high school students.
Source: Authors' analysis of data from Pittsburgh Public Schools and Allegheny County Department of Human Services for the 2016/17 school year.

Breakdown of wrong predictions by race

Predictions can be wrong in two directions—either over-identifying students or under-identifying students. Over-identified students are the students who were predicted to have an academic problem but did not have an academic problem, or the false positives. Under-identified students are the students who were not predicted to have an academic problem but did have an academic problem, or the false negatives. Depending on the outcome being predicted and the intervention planned for those identified, it may be preferred to over-identify or under-identify individuals. Users may especially be concerned about the patterns of over-identifying or under-identifying students of different racial groups.

The study team calculated the differences in the percent of *over-identified* students who are Black compared to students who are White; these numbers are shown next to each set of bars (figure B3). All of the options somewhat over-identify more students who are Black. Compared to the other predictions, the optimal cutoffs over-identify a larger percentage of Black students than White students for course failure (11 percentage points) and low GPA (9 percentage points), but especially for suspensions (28 percentage points).

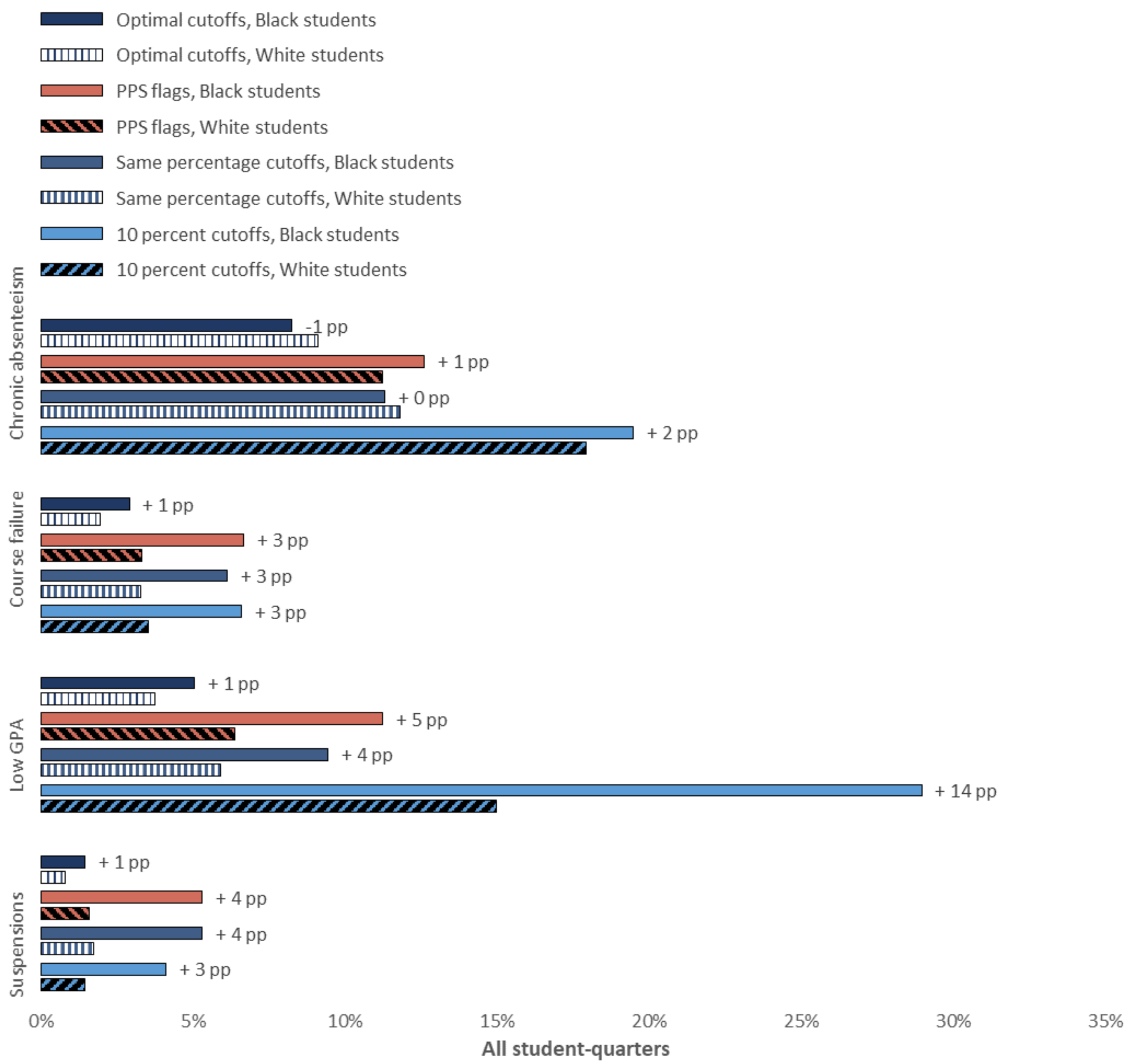
Figure B3. Percent of student-quarters over-identified, by race and outcome, 2016/17



GPA is grade point average. pp is percentage point. PPS is Pittsburgh Public Schools.
Note: Low GPA only includes high school students.
Source: Authors' analysis of data from Pittsburgh Public Schools and Allegheny County Department of Human Services for the 2016/17 school year.

The differences in the percentage of *under-identified* students who are Black compared to students who are White are also shown (figure B4). All of the options slightly under-identify more students who are Black. For low GPA, the 10 percent cutoffs under-identify more students who are Black compared to students who are White (14 percentage point difference). For the other outcomes, the differences in rates of under-identifying students are 5 percentage points or less.

Figure B4. Percent of student-quarters under-identified, by race and outcome, 2016/17



GPA is grade point average. pp is percentage point. PPS is Pittsburgh Public Schools.
Note: Low GPA only includes high school students.
Source: Authors' analysis of data from Pittsburgh Public Schools and Allegheny County Department of Human Services for the 2016/17 school year.

Regression analysis

To examine how much of the variation in outcomes could be captured by the predictions, the study team ran a series of regressions predicting the outcomes using the prediction from the PPS flags and the risk score cutoff options. The results are discussed in the text and listed here (table B2).

Table B2. R-squared values from regression analysis, 2016/17 (percent of student-quarters)

Outcome/prediction method	Black students	White students
Chronic absenteeism		
Optimal cutoffs	0.192	0.151
Same percentage cutoffs	0.213	0.161
10 percent cutoffs	0.173	0.118
PPS flag	0.192	0.153
Course failure		
Optimal cutoffs	0.224	0.283
Same percentage cutoffs	0.264	0.310
10 percent cutoffs	0.262	0.304
PPS flag	0.240	0.287
Low GPA		
Optimal cutoffs	0.365	0.472
Same percentage cutoffs	0.382	0.483
10 percent cutoffs	0.184	0.197
PPS flag	0.328	0.415
Suspension		
Optimal cutoffs	0.042	0.051
Same percentage cutoffs	0.055	0.047
10 percent cutoffs	0.057	0.058
PPS flag	0.052	0.057

GPA is grade point average. PPS is Pittsburgh Public Schools.
Note: Low GPA only includes high school students.
Source: Authors' analysis of data from Pittsburgh Public Schools and Allegheny County Department of Human Services for the 2016/17 school year.

Detailed accuracy data

This section includes the detailed accuracy statistics for each outcome and each prediction method (see the methods appendix for definitions). See table B3 for chronic absenteeism data, table B4 for course failure data, table B5 for low GPA data, and table B6 for suspension data. Each table presents the statistics for all students and the following student groups: elementary school students, middle school students, high school students, male students, female students, students who are Black, students who are White, student eligible for free or reduced-price lunches, students not eligible for free or reduced-price lunches, students with Allegheny County Department of Human Services (DHS) involvement, and students without DHS involvement.[8]

Table B3. Chronic absenteeism

Group	Cutoff	True positive (TP)	True negative (TN)	False positive (FP)	False negative (FN)	Sensitivity (TP/[TP+FN])	Specificity (TN/[TN+FP])	False positive rate	Precision (TP/[TP+FP])
All	Optimal	0.18	0.57	0.16	0.09	0.68	0.78	0.22	0.52
	PPS flags	0.15	0.63	0.1	0.12	0.55	0.86	0.14	0.59
	Same percentage	0.15	0.63	0.1	0.12	0.57	0.87	0.13	0.61
	10 percent	0.08	0.71	0.02	0.19	0.3	0.97	0.03	0.79
Elementary school	Optimal	0.12	0.64	0.14	0.1	0.55	0.82	0.18	0.45
	PPS flags	0.1	0.69	0.1	0.11	0.47	0.87	0.13	0.5
	Same percentage	0.09	0.71	0.08	0.12	0.42	0.9	0.1	0.53
	10 percent	0.03	0.77	0.01	0.18	0.16	0.98	0.02	0.72
Middle school	Optimal	0.17	0.58	0.17	0.08	0.67	0.77	0.23	0.49
	PPS flags	0.13	0.65	0.1	0.12	0.53	0.87	0.13	0.57
	Same percentage	0.13	0.65	0.1	0.11	0.54	0.87	0.13	0.58
	10 percent	0.06	0.73	0.02	0.19	0.22	0.97	0.03	0.74
High school	Optimal	0.31	0.43	0.2	0.07	0.82	0.68	0.32	0.61
	PPS flags	0.24	0.52	0.1	0.13	0.65	0.84	0.16	0.7
	Same percentage	0.27	0.5	0.13	0.1	0.73	0.79	0.21	0.68
	10 percent	0.17	0.59	0.04	0.2	0.47	0.94	0.06	0.83
Male	Optimal	0.18	0.57	0.17	0.09	0.67	0.77	0.23	0.51
	PPS flags	0.14	0.64	0.1	0.12	0.54	0.86	0.14	0.58
	Same percentage	0.15	0.64	0.1	0.11	0.56	0.86	0.14	0.59
	10 percent	0.08	0.72	0.02	0.19	0.29	0.97	0.03	0.77

[8] DHS involvement means the student appeared in the Allegheny County DHS database as having received social services or public benefits or had been involved with the justice system.

Table B3. Chronic absenteeism *(continued)*

Group	Cutoff	True positive (TP)	True negative (TN)	False positive (FP)	False negative (FN)	Sensitivity (TP/[TP+FN])	Specificity (TN/[TN+FP])	False positive rate	Precision (TP/[TP+FP])
Female	Optimal	0.19	0.57	0.16	0.09	0.68	0.78	0.22	0.54
	PPS flags	0.15	0.63	0.1	0.12	0.56	0.86	0.14	0.6
	Same percentage	0.16	0.63	0.1	0.12	0.58	0.87	0.13	0.62
	10 percent	0.08	0.71	0.02	0.19	0.3	0.97	0.03	0.8
Black	Optimal	0.21	0.53	0.18	0.08	0.72	0.75	0.25	0.55
	PPS flags	0.17	0.6	0.1	0.13	0.58	0.85	0.15	0.62
	Same percentage	0.18	0.59	0.11	0.11	0.62	0.84	0.16	0.62
	10 percent	0.1	0.68	0.03	0.19	0.34	0.96	0.04	0.79
White	Optimal	0.14	0.62	0.15	0.09	0.61	0.81	0.19	0.49
	PPS flags	0.12	0.67	0.1	0.11	0.51	0.87	0.13	0.54
	Same percentage	0.11	0.69	0.08	0.12	0.49	0.89	0.11	0.57
	10 percent	0.05	0.75	0.02	0.18	0.22	0.98	0.02	0.77
Eligible for free or reduced-price lunches	Optimal	0.23	0.49	0.19	0.09	0.72	0.72	0.28	0.55
	PPS flags	0.18	0.57	0.11	0.13	0.58	0.84	0.16	0.63
	Same percentage	0.19	0.57	0.12	0.12	0.61	0.83	0.17	0.62
	10 percent	0.1	0.66	0.03	0.21	0.32	0.96	0.04	0.79
Not eligible for free or reduced-price lunches	Optimal	0.1	0.69	0.12	0.08	0.56	0.85	0.15	0.46
	PPS flags	0.09	0.73	0.09	0.1	0.46	0.89	0.11	0.5
	Same percentage	0.08	0.75	0.07	0.1	0.44	0.92	0.08	0.55
	10 percent	0.04	0.8	0.01	0.15	0.22	0.99	0.01	0.78
DHS involvement	Optimal	0.31	0.39	0.23	0.08	0.80	0.63	0.37	0.58
	PPS flags	0.25	0.50	0.12	0.14	0.63	0.81	0.19	0.67
	Same percentage	0.27	0.47	0.14	0.12	0.70	0.77	0.23	0.65
	10 percent	0.14	0.58	0.03	0.24	0.37	0.95	0.05	0.82
No DHS involvement	Optimal	0.16	0.61	0.15	0.09	0.64	0.80	0.20	0.51
	PPS flags	0.13	0.66	0.10	0.11	0.53	0.87	0.13	0.57
	Same percentage	0.13	0.67	0.09	0.11	0.53	0.88	0.12	0.59
	10 percent	0.07	0.74	0.02	0.18	0.27	0.97	0.03	0.77

DHS is Allegheny County Department of Human Services. PPS is Pittsburgh Public Schools.
Source: Authors' analysis of administrative data from Pittsburgh Public Schools in 2016/17.

Table B4. Course failure

Group	Cutoff	True positive (TP)	True negative (TN)	False positive (FP)	False negative (FN)	Sensitivity (TP/[TP+FN])	Specificity (TN/[TN+FP])	False positive rate (1-specificity)	Precision (TP/[TP+FP])
All	Optimal	0.09	0.76	0.12	0.02	0.78	0.86	0.14	0.42
	PPS flags	0.06	0.84	0.04	0.05	0.55	0.95	0.05	0.58
	Same percentage	0.07	0.84	0.04	0.05	0.57	0.95	0.05	0.6
	10 percent	0.06	0.85	0.04	0.05	0.54	0.96	0.04	0.62
Elementary school	Optimal	0.02	0.89	0.07	0.02	0.58	0.92	0.08	0.25
	PPS flags	0.02	0.93	0.03	0.03	0.37	0.97	0.03	0.36
	Same percentage	0.01	0.94	0.02	0.03	0.32	0.98	0.02	0.4
	10 percent	0.01	0.94	0.02	0.03	0.29	0.98	0.02	0.42
Middle school	Optimal	0.08	0.77	0.12	0.03	0.72	0.87	0.13	0.4
	PPS flags	0.05	0.85	0.04	0.06	0.48	0.95	0.05	0.55
	Same percentage	0.05	0.85	0.04	0.06	0.49	0.96	0.04	0.57
	10 percent	0.05	0.86	0.03	0.06	0.46	0.96	0.04	0.6
High school	Optimal	0.2	0.57	0.2	0.03	0.86	0.73	0.27	0.49
	PPS flags	0.14	0.7	0.07	0.09	0.62	0.9	0.1	0.66
	Same percentage	0.15	0.69	0.09	0.08	0.67	0.89	0.11	0.64
	10 percent	0.15	0.7	0.07	0.08	0.64	0.9	0.1	0.66
Male	Optimal	0.11	0.72	0.14	0.03	0.81	0.83	0.17	0.43
	PPS flags	0.08	0.81	0.05	0.06	0.57	0.94	0.06	0.6
	Same percentage	0.08	0.81	0.05	0.05	0.6	0.94	0.06	0.61
	10 percent	0.08	0.82	0.05	0.06	0.57	0.95	0.05	0.63
Female	Optimal	0.07	0.8	0.1	0.02	0.75	0.88	0.12	0.4
	PPS flags	0.05	0.87	0.04	0.05	0.51	0.96	0.04	0.55
	Same percentage	0.05	0.87	0.04	0.04	0.53	0.96	0.04	0.58
	10 percent	0.05	0.88	0.03	0.05	0.5	0.97	0.03	0.6

Table B4. Course failure *(continued)*

Group	Cutoff	True positive (TP)	True negative (TN)	False positive (FP)	False negative (FN)	Sensitivity (TP/[TP+FN])	Specificity (TN/[TN+FP])	False positive rate (1-specificity)	Precision (TP/[TP+FP])
Black	Optimal	0.12	0.68	0.17	0.03	0.8	0.8	0.2	0.41
	PPS flags	0.08	0.79	0.06	0.07	0.55	0.93	0.07	0.58
	Same percentage	0.09	0.79	0.06	0.06	0.59	0.93	0.07	0.58
	10 percent	0.08	0.8	0.05	0.07	0.56	0.94	0.06	0.61
White	Optimal	0.05	0.86	0.06	0.02	0.73	0.93	0.07	0.45
	PPS flags	0.04	0.9	0.03	0.03	0.54	0.97	0.03	0.59
	Same percentage	0.04	0.91	0.02	0.03	0.54	0.98	0.02	0.63
	10 percent	0.04	0.91	0.02	0.04	0.51	0.98	0.02	0.66
Eligible for free or reduced-price lunches	Optimal	0.11	0.71	0.15	0.03	0.8	0.82	0.18	0.42
	PPS flags	0.08	0.8	0.05	0.06	0.56	0.94	0.06	0.59
	Same percentage	0.08	0.8	0.06	0.06	0.59	0.94	0.06	0.6
	10 percent	0.08	0.81	0.05	0.06	0.56	0.94	0.06	0.62
Not eligible for free or reduced-price lunches	Optimal	0.05	0.85	0.08	0.02	0.73	0.92	0.08	0.41
	PPS flags	0.04	0.9	0.03	0.04	0.51	0.97	0.03	0.56
	Same percentage	0.04	0.9	0.03	0.03	0.51	0.97	0.03	0.59
	10 percent	0.03	0.91	0.02	0.04	0.48	0.98	0.02	0.61
DHS involvement	Optimal	0.15	0.60	0.22	0.03	0.84	0.73	0.27	0.41
	PPS flags	0.10	0.75	0.07	0.08	0.57	0.92	0.08	0.60
	Same percentage	0.12	0.74	0.08	0.06	0.64	0.90	0.10	0.58
	10 percent	0.11	0.75	0.07	0.07	0.61	0.91	0.09	0.60
No DHS involvement	Optimal	0.08	0.79	0.11	0.02	0.77	0.88	0.12	0.42
	PPS flags	0.05	0.86	0.04	0.05	0.54	0.96	0.04	0.58
	Same percentage	0.06	0.86	0.04	0.05	0.55	0.96	0.04	0.60
	10 percent	0.05	0.87	0.03	0.05	0.52	0.97	0.03	0.63

DHS is Allegheny County Department of Human Services. PPS is Pittsburgh Public Schools.
Source: Authors' analysis of administrative data from Pittsburgh Public Schools in 2016/17.

Table B5. Low grade point average

Group	Cutoff	True positive (TP)	True negative (TN)	False positive (FP)	False negative (FN)	Sensitivity (TP/[TP+FN])	Specificity (TN/[TN+FP])	False positive rate (1-specificity)	Precision (TP/[TP+FP])
All	Optimal	0.28	0.56	0.12	0.05	0.86	0.83	0.17	0.71
	PPS flags	0.23	0.6	0.07	0.09	0.72	0.89	0.11	0.76
	Same percentage	0.25	0.61	0.07	0.08	0.75	0.9	0.1	0.79
	10 percent	0.1	0.67	0.01	0.23	0.29	0.99	0.01	0.95
Elementary school	Optimal	na	na	na	na	na	na	na	na
	PPS flags	na	na	na	na	na	na	na	na
	Same percentage	na	na	na	na	na	na	na	na
	10 percent	na	na	na	na	na	na	na	na
Middle school	Optimal	na	na	na	na	na	na	na	na
	PPS flags	na	na	na	na	na	na	na	na
	Same percentage	na	na	na	na	na	na	na	na
	10 percent	na	na	na	na	na	na	na	na
High school	Optimal	0.28	0.56	0.12	0.05	0.86	0.83	0.17	0.71
	PPS flags	0.23	0.6	0.07	0.09	0.72	0.89	0.11	0.76
	Same percentage	0.25	0.61	0.07	0.08	0.75	0.9	0.1	0.79
	10 percent	0.1	0.67	0.01	0.23	0.29	0.99	0.01	0.95
Male	Optimal	0.35	0.48	0.13	0.05	0.88	0.79	0.21	0.73
	PPS flags	0.29	0.53	0.08	0.1	0.74	0.87	0.13	0.79
	Same percentage	0.31	0.53	0.07	0.08	0.79	0.88	0.12	0.81
	10 percent	0.13	0.6	0.01	0.26	0.33	0.99	0.01	0.95
Female	Optimal	0.22	0.63	0.11	0.04	0.83	0.86	0.14	0.67
	PPS flags	0.18	0.67	0.07	0.08	0.69	0.91	0.09	0.73
	Same percentage	0.18	0.68	0.06	0.08	0.71	0.92	0.08	0.76
	10 percent	0.06	0.74	0	0.2	0.24	1	0	0.95
Black	Optimal	0.38	0.42	0.16	0.05	0.88	0.73	0.27	0.71
	PPS flags	0.31	0.48	0.1	0.11	0.74	0.83	0.17	0.77
	Same percentage	0.33	0.48	0.09	0.09	0.78	0.84	0.16	0.78
	10 percent	0.14	0.57	0.01	0.29	0.32	0.99	0.01	0.94

Table B5. Low GPA *(continued)*

Group	Cutoff	True positive (TP)	True negative (TN)	False positive (FP)	False negative (FN)	Sensitivity (TP/[TP+FN])	Specificity (TN/[TN+FP])	False positive rate (1-specificity)	Precision (TP/[TP+FP])
White	Optimal	0.16	0.73	0.07	0.04	0.81	0.91	0.09	0.7
	PPS flags	0.13	0.76	0.05	0.06	0.68	0.94	0.06	0.75
	Same percentage	0.14	0.77	0.03	0.06	0.7	0.96	0.04	0.8
	10 percent	0.05	0.8	0	0.15	0.25	1	0	0.97
Eligible for free or reduced-price lunches	Optimal	0.37	0.44	0.14	0.05	0.88	0.75	0.25	0.72
	PPS flags	0.31	0.5	0.09	0.11	0.74	0.85	0.15	0.78
	Same percentage	0.33	0.5	0.09	0.09	0.79	0.85	0.15	0.79
	Top t10 percent	0.14	0.58	0.01	0.28	0.33	0.99	0.01	0.95
Not eligible for free or reduced-price lunches	Optimal	0.17	0.71	0.08	0.04	0.8	0.9	0.1	0.68
	PPS flags	0.14	0.73	0.05	0.07	0.67	0.93	0.07	0.73
	Same percentage	0.14	0.75	0.04	0.07	0.67	0.95	0.05	0.78
	10 percent	0.04	0.78	0	0.17	0.21	1	0	0.96
DHS involvement	Optimal	0.46	0.34	0.16	0.04	0.92	0.68	0.32	0.74
	PPS flags	0.39	0.41	0.09	0.11	0.78	0.82	0.18	0.82
	Same percentage	0.41	0.41	0.09	0.09	0.82	0.82	0.18	0.82
	10 percent	0.17	0.49	0.00	0.34	0.33	0.99	0.01	0.98
No DHS involvement	Optimal	0.25	0.60	0.11	0.05	0.85	0.85	0.15	0.70
	PPS flags	0.21	0.63	0.07	0.09	0.70	0.90	0.10	0.75
	Same percentage	0.22	0.64	0.06	0.08	0.73	0.91	0.09	0.78
	10 percent	0.08	0.70	0.01	0.21	0.28	0.99	0.01	0.94

DHS is Allegheny County Department of Human Services. GPA is grade point average. na is not applicable. PPS is Pittsburgh Public Schools.
Note: Low GPA only includes high school students.
Source: Authors' analysis of administrative data from Pittsburgh Public Schools in 2016/17.

Table B6. Suspensions

Group	Cutoff	True positive (TP)	True negative (TN)	False positive (FP)	False negative (FN)	Sensitivity (TP/[TP+FN])	Specificity (TN/[TN+FP])	False positive rate (1-specificity)	Precision (TP/[TP+FP])
All	Optimal	0.04	0.68	0.27	0.01	0.78	0.72	0.28	0.13
	PPS flags	0.01	0.91	0.04	0.04	0.29	0.96	0.04	0.27
	Same percentage	0.01	0.91	0.04	0.04	0.28	0.96	0.04	0.29
	10 percent	0.02	0.87	0.08	0.03	0.43	0.92	0.08	0.22
Elementary school	Optimal	0.02	0.83	0.14	0.01	0.61	0.85	0.15	0.12
	PPS flags	0.01	0.94	0.02	0.02	0.29	0.97	0.03	0.27
	Same percentage	0.01	0.95	0.01	0.03	0.2	0.99	0.01	0.3
	10 percent	0.01	0.94	0.03	0.02	0.32	0.97	0.03	0.27
Middle school	Optimal	0.06	0.56	0.36	0.01	0.82	0.61	0.39	0.15
	PPS flags	0.02	0.87	0.06	0.05	0.3	0.94	0.06	0.29
	Same percentage	0.02	0.87	0.05	0.05	0.3	0.95	0.05	0.32
	10 percent	0.03	0.83	0.09	0.04	0.43	0.9	0.1	0.26
High school	Optimal	0.06	0.51	0.42	0.01	0.87	0.55	0.45	0.12
	PPS flags	0.02	0.88	0.05	0.05	0.29	0.94	0.06	0.26
	Same percentage	0.02	0.87	0.06	0.04	0.34	0.93	0.07	0.26
	10 percent	0.03	0.78	0.15	0.03	0.54	0.84	0.16	0.18
Male	Optimal	0.05	0.62	0.32	0.01	0.79	0.66	0.34	0.13
	PPS flags	0.02	0.89	0.05	0.04	0.3	0.95	0.05	0.28
	Same percentage	0.02	0.9	0.04	0.04	0.29	0.95	0.05	0.28
	10 percent	0.03	0.84	0.1	0.03	0.45	0.9	0.1	0.22
Female	Optimal	0.03	0.74	0.22	0.01	0.75	0.77	0.23	0.13
	PPS flags	0.01	0.92	0.03	0.03	0.28	0.97	0.03	0.27
	Same percentage	0.01	0.93	0.03	0.03	0.26	0.97	0.03	0.29
	10 percent	0.02	0.9	0.06	0.03	0.41	0.94	0.06	0.23
Black	Optimal	0.06	0.53	0.39	0.01	0.81	0.58	0.42	0.13
	PPS flags	0.02	0.87	0.06	0.05	0.3	0.94	0.06	0.28
	Same percentage	0.02	0.87	0.05	0.05	0.3	0.94	0.06	0.29
	10 percent	0.03	0.81	0.12	0.04	0.45	0.87	0.13	0.22

Table B6. Suspensions *(continued)*

Group	Cutoff	True positive (TP)	True negative (TN)	False positive (FP)	False negative (FN)	Sensitivity (TP/[TP+FN])	Specificity (TN/[TN+FP])	False positive rate (1-specificity)	Precision (TP/[TP+FP])
White	Optimal	0.01	0.86	0.11	0.01	0.64	0.88	0.12	0.11
	PPS flags	0.01	0.96	0.02	0.02	0.27	0.98	0.02	0.24
	Same percentage	0	0.97	0.01	0.02	0.2	0.99	0.01	0.26
	10 percent	0.01	0.95	0.03	0.01	0.33	0.97	0.03	0.21
Eligible for free or reduced-price lunches	Optimal	0.05	0.58	0.35	0.01	0.81	0.62	0.38	0.13
	PPS flags	0.02	0.88	0.05	0.05	0.3	0.95	0.05	0.29
	Same percentage	0.02	0.89	0.05	0.05	0.29	0.95	0.05	0.29
	10 percent	0.03	0.83	0.1	0.04	0.46	0.89	0.11	0.23
Not eligible for free or reduced-price lunches	Optimal	0.02	0.84	0.14	0.01	0.63	0.86	0.14	0.1
	PPS flags	0.01	0.95	0.02	0.02	0.25	0.98	0.02	0.21
	Same percentage	0.01	0.96	0.02	0.02	0.22	0.98	0.02	0.25
	10 percent	0.01	0.94	0.04	0.02	0.33	0.96	0.04	0.18
DHS involvement	Optimal	0.09	0.42	0.47	0.01	0.87	0.48	0.52	0.17
	PPS flags	0.04	0.82	0.07	0.07	0.37	0.92	0.08	0.35
	Same percentage	0.04	0.81	0.09	0.07	0.38	0.90	0.10	0.32
	10 percent	0.06	0.73	0.17	0.05	0.54	0.81	0.19	0.26
No DHS involvement	Optimal	0.03	0.74	0.23	0.01	0.72	0.77	0.23	0.11
	PPS flags	0.01	0.93	0.03	0.03	0.25	0.97	0.03	0.24
	Same percentage	0.01	0.93	0.03	0.03	0.23	0.97	0.03	0.26
	10 percent	0.01	0.90	0.06	0.02	0.38	0.94	0.06	0.20

DHS is Allegheny County Department of Human Services. PPS is Pittsburgh Public Schools.
Source: Authors' analysis of administrative data from Pittsburgh Public Schools in 2016/17.