A Better Cold-Start for Early Prediction of Student At-Risk Status in New School Districts

Chad Coleman
Teachers College, Columbia University
525 W 120th St
New York, NY 10027
+1 347-205-4919
cjc2238@tc.columbia.edu

Ryan S Baker, Ph.D.
University of Pennsylvania
3700 Walnut St
Philadelphia, PA 19104
+1 215-573-2990
rybaker@upenn.edu

Shonte Stephenson, Ph.D.
BrightBytes
717 Market St STE 300
San Francisco, CA 94103
+1 877-433-4036
shonte@brightbytes.net

ABSTRACT
Determining which students are at risk of poorer outcomes -- such as dropping out, failing classes, or decreasing standardized examination scores -- has become an important area of research and practice in both K-12 and higher education. The detectors produced from this type of predictive modeling research are increasingly used in early warning systems to identify which students are at risk and intervene to support better outcomes. In K-12, it has become common practice to re-build and validate these detectors, district-by-district, due to different data semantics and risk factors for students in different districts. As these detectors become more widely used, however, it becomes desirable to also apply detectors in school districts without sufficient historical data to build a district-specific model. Novel approaches that can address the complex data challenges a new district presents are critical for extending the benefits of early warning systems to all students. Using an ensemble-based algorithm, we evaluate a model averaging approach that can generate a useful model for previously-unseen districts. During the ensembling process, our approach builds models for districts that have a significant amount of historical records and integrates them through averaging. We then use these models to generate predictions for districts suffering from high data missingness. Using this approach, we are able to predict student-at-risk status effectively for unseen districts, across a range of grade ranges, and achieve prediction goodness comparable to previously published models predicting at-risk

Keywords
Learning Analytics, High School Graduation, Machine Learning, Ensembling.

1. INTRODUCTION
Graduating from high school is an educational achievement that is strongly linked to gainful well-paying employment, higher personal income, better personal health, reduced risk of incarceration, and lowered reliance on social welfare programs [23, 2]. Fortunately, graduation rates have been rising in the United States, trending towards reaching 90% nationwide by the year 2020 [18]. While this is a positive accomplishment, it leaves millions of students not completing high school, representing a continuing crisis within the American educational system. This crisis is not evenly distributed; in the USA, there are much higher dropout rates for African American, Native American, and Hispanic/Latinx students [35, 19], up to 4 times the rate for White students, as well as for learners from low-income families and with disabilities [37].

There is a clear benefit in completing a high school education, so why do we still see alarmingly high levels of high school dropout? A great deal of research has been conducted trying to answer this one question, with the hope that once identification is achieved educators and administrators can apply a preventative or remedial intervention to curb student dropout [11]. However, many factors appear to lead to student dropout, including lack of social support from parents, poor motivation, low self-esteem, parental educational achievement and value, and economic factors, making it difficult to create a single intervention that works for all students [19, 30]. While demographic factors correlate with eventual dropout, these indicators are not considered actionable. A school district generally does not have the capacity to improve a student’s economic condition, nor is it possible to alter a student's racial identity or gender. As such, the educational research community has focused on more actionable factors such as behavior, attendance, engagement, and social-emotional learning [21]. The most successful interventions have attempted to address issues related to specific indicators while also attempting to improve overall student academic engagement [14]. There are a range of potential interventions and many are costly, driving a need to identify the students that could benefit most from specific forms of support. Identifying these students can be a difficult task [10] which has led to an ongoing effort within the educational research community to determine which students are at risk of not graduating from high school [20] to apply proactive interventions that can help get students back on track [8].

This goal, along with the growing availability of student data, has led to Early Warning Systems, which leverage statistical methods applied to historical student data in order to predict outcomes for new students. Early work on predicting high school graduation tended to use statistical methods in order to infer the relationship between graduation and indicators such as grades and attendance. For example, the seminal Chicago Model developed an "on-task" indicator built from first-year high school student performance indicators and then used this newly defined feature within logistic regression to model student risk [1]. This method proved effective with 80+ percent accuracy in predicting student dropout, leading to high popularity and wide-scale implementation and use [6].

More recently, researchers have begun to leverage machine learning and data mining methods, sometimes termed predictive analytics, to find complex patterns associated with future student outcomes [28, 17]. In K-12 education, this approach was used by Lakkaraju et al. [29] to predict student dropout in two districts, finding that the Random Forest algorithm outperformed several other algorithms. Some of the efforts to use machine learning in predicting student success have been scaled beyond single districts to entire states [27]. However, it still remains a challenge to deploy predictive analytics for use in schools at scale. District data often contain substantial information about its schools and students: demographic data about the student and teacher populations, academic performance information, financial information, disciplinary actions, and attendance records [36]. However, in many school districts, data quality is limited, with key information only available by integrating across multiple data warehouses, incompatible student ID numbers, errors in data entry, and local idiosyncratic interpretations of often ambiguous data fields. Even when data is today excellent, key data from past years is often unavailable due to the absence of a formal data system or having a data system which is difficult to query. Semantics may also change -- for example, the definition of "not graduated" is not stable across years and contexts [35] -- but these changes may not always be clearly understood when reviewing past data.

One solution is to use models that involve simple variables that are feasible for almost any district's data and assume that model will be valid in new contexts, even where that context may be quite different from the context where the model was initially developed [32]. The Chicago model [1] is a common choice for this type of application.

In this paper, we propose and evaluate an alternate solution to providing a model for a new district. Our approach attempts to generate predictions for a specific “Target” school district based on models from other school districts where full datasets are available, using a simple average of the district models, where all existing models are given equal weight. We compare the quality of our averaging approach to the earlier solution of using a simple generic model, specifically, the Chicago model.

2. METHODS

In the following section, we will discuss our method for making at-risk student predictions for school districts which have insufficient data to create a district-based prediction model. In brief, we develop and validate predictive analytics models for each school district with sufficient data. These models predict each student’s probability of graduating (or risk of not graduating). We then conduct a simple ensembling approach, averaging each model’s predictions, to produce a single prediction for each student. We test the quality of this approach by conducting it for held-out districts where data is available.

We validate our new approach by comparing its performance to the widely-used “Chicago model” [1, 6] on the same test data and comparing the performance of our detector to the classic Chicago model, which can be used for entirely new districts with no re-training. The Chicago model utilizes freshman-year GPA, the number of semester course failures, and freshman-year absences to determine the student’s risk of failing to graduate [1]. Since the Chicago model relies on data collected within the first year of high school, we were only able to compare the performance of our approach to the Chicago model for high school students.

2.1 Data

Data for this research originate from the BrightBytes data analytics and visualization platform, Clarity®. The Clarity® platform ingests disparate datasets, transforms them to a standardized format by mapping district-specific variables to a common schema, prepares the data for analysis, and then visualizes the data in a meaningful, easy-to-understand way. The Clarity® platform is used by 1 in 5 schools across 47 states to empower educational leaders to use data for decision making. The value derived by districts from the Clarity® platform comes from using data to drive change within an organization. The anonymized dataset used in this paper (n =3,575,724) represents a large spectrum of K-12 students in terms of free/reduced lunch eligibility, school urbanicity, and school demographic makeup, and is drawn from a range of school districts, educational organizations and agencies.

We have nearly complete data (with only small numbers of variables unavailable) from an educational agency with decision-making power over a large geographical region (Pillar 1) and three large individual school districts (Pillar 2, Pillar 3, Pillar 4). These datasets are referred to as “Pillars” because they serve as bases for our ensemble-based approach. The four Pillars differ in terms of their predominant student demographic groups, with Caucasians representing the largest group of students in Pillar 1 (n=1,681,988), Hispanic/Latinx students representing the largest group of students in Pillar 2 (n= 392,148), and split demographics in Pillar 3 (n=158,991) and Pillar 4 (n=140,132).

We test our models on 30 “Target” districts that were not used to develop the models, due to having fewer years of data, more missing variables, or smaller samples overall. These Target districts span a diverse range of predominant demographics, with one Target district being over ninety percent Caucasian at one extreme, and other districts being almost completely Hispanic/Latinx or African American. District performance is equally as diverse: some Target districts achieve graduation rates over 90% while others have graduation rates as low as 36%. Table 1 below highlights the number of records available in the Pillar districts and Target districts. The Target districts are generally smaller than the Pillar districts, with some having as few as 271 total historical student records, with the percent of data missingness within the Target districts also quite high in some cases (M= 41.65%, SD=7.498%).

Table 1: Average Number of Outcome Records in Target Districts

Grade Band	Graduates	SD	Dropouts	SD
1st – 5th	4,307	9,784	634	1,267
6th – 8th	7,673	16,837	996	2,012
9th – 12th	24,552	39,524	1,920	3,833
All Grades	**12,177**	**95,962**	**1,183**	**7,377**

2.2 Predictor Variables

The potential set of predictor variables was selected in partnership with the American Institutes for Research (AIR) team [22], this paper’s authors, and other researchers and developers at BrightBytes. This collaboration resulted in a theory-based [9] framework of success indicators, along with definitions of those success indicators that are used to map and align district data. Due to the data ingestion and transformation process, the same data features can be used across all districts. Below is a distillation of the broad range of potential variables into a small set of meaningful buckets:

General Coursework: student academic performance such as total credits earned or student grade point averages [25, 38, 5].
Student Assessments: interim or summative assessments related to math, science, reading and social studies performance [26,16].
Student Attendance: recorded as absences or tardies [33].
Student Behavior: disciplinary incidents the student has on file [4, 10].

2.3 Model Fitting

The first step toward building an at-risk prediction model for districts without sufficient data is to build models for districts with sufficient data. For each of these models, we took the data from a single district. We filtered down to only the students who were flagged as 'dropped' and 'graduated'. Even if students took extra time to graduate, they were still counted as graduating. Only these students were used for building the model; all other outcomes (such as transferring to another school district) were removed from the filtered dataset. The resultant datasets were generally highly imbalanced, with substantially more students graduating than dropping out. To account for this imbalance, the training data was manually rebalanced by adding duplicate copies of students who dropped out to the data set. Specifically, duplicates were created such that every grade level (10th, 11th, 12th, etc.) of students in the training datasets had an equal number of students who dropped out as students who remained. The original data distribution was used when testing the models.

Decision trees, support vector machines, XGBoost, logistic regression and random forest were all tested to build the initial model. The best performance across data was obtained with random forest classifiers with n=15 estimators and a max depth of 10. Since the algorithm was tree based, we utilized arbitrary value substitution to replace missing values with a high integer [39]. The goodness of each district's model was evaluated, within-district, using a train-test split method (note that models are also evaluated within entirely new districts; see below). In each case, the training set consisted of a randomly selected 70 percent of the data with label-based stratification used across grades. The test set held out to validate the model consisted of the remaining 30 percent of the data.

Models were evaluated using the Area Under the Curve for the Receiver Operator Characteristic graph. AUC ROC was selected as our primary evaluation statistic due to its interpretability and validity for highly-imbalanced test sets [24]. AUC ROC calculates the tradeoff between true positive and false negative for every possible threshold used for labeling data points as positive and negative; as such, it is well-suited for evaluating how well an algorithm ranks students relative to their risk.

2.4 Pillar Selection

Selection of the four Pillar districts was based on two factors, data quality and model performance. To evaluate data quality, we calculated the proportion of missing values within the total feature set, expressed as a percentage. Districts were not included as Pillar districts if they had high amounts of missing data, over 40% of values missing, as these districts would be less useful for modeling other districts where these features were present. Districts were also not included as Pillar districts if they lacked historical data spanning all grades 1 through 12; districts without historical data for some grades would be less useful for developing models that could be applied to all grades.

Models developed for specific districts as potential pillar model candidates were fit and evaluated using held-out test sets from that district's own data. Districts for which we were able to produce a model with AUC higher than 0.7 on the district's test sample, averaged across all student class years, were designated as Pillar districts/models and used in our predictions for districts for which models could not be generated for all grade levels, or for which models were insufficient in quality. Of the 30 Target districts within our study, 25 do not have enough historical records spanning all 12 grades and 5 had sufficient data but were unable to produce an AUC over 0.7. All 30 districts suffered from at least some degree of feature missingness.

2.5 Applying Models to New Districts

Having developed models for Pillar districts, where data are abundant, data quality is high, and where it is possible to develop a high-quality model, we next applied each Pillar model to each Target district. These Target districts had at least one of the following attributes; 1) Under 20,000 students, 2) Over 40% missing values, 3) Missing historical records for some grades in K-12, 4) AUC ROC when applied to new students within-district.

Our first step to applying the Pillar models was simply to run each of them on the Target district's data (n = 1,202,465) and obtain predictions for each student. This provides us with a set of predictions for each student and for each model. We then took the average of the probability estimates, across districts, to generate the final student prediction. When we applied Pillar models to Target districts, we evaluated these models using all historical records present in the data as none of their records were used within model training. As with models tested on the district for which they were built, we use AUC ROC as our metric of model goodness.

3. RESULTS

3.1 Within-District Performance

We first applied each Pillar district model to new students from the same district, to evaluate within-district performance. As shown in Table 2, the four Pillar districts achieved AUC ROC values ranging between 0.899-0.936 when predicting graduation/dropout, for 9th through 12th grades (the typical high school years in U.S. classrooms). Performance was moderately lower for 6th-8th graders, where longitudinal predictions of up to 6 years are being made, with AUC ROC ranging from 0.849-0.884. Performance was again moderately lower for 1st-5th graders, where longitudinal predictions of up to 11 years are being made, with AUC ROC ranging from 0.758-0.810.

Table 2. AUC of Pillar Model Performance on Pillar Model Test Data (new students) by Grade Band

District	1 – 5	6 - 8	9 - 12	Average
Pillar 1	0.778	0.849	0.899	0.865
Pillar 2	0.758	0.884	0.908	0.858
Pillar 3	0.810	0.884	0.936	0.888

Pillar 4	0.794	0.850	0.903	0.886

Our attempts to build a model for each Target district proved less successful, with 9th – 12th grade model AUC averaging 0.729, 0.1825 lower than for the Pillar Districts, 6th-8th model AUC averaging 0.669, 0.198 AUC lower than the Pillar Districts, and 1st -5th grade obtaining an average AUC of 0.635, 0.15 points lower than the Pillar Districts' within-district performance for these grade levels.

It should be noted that there were three exceptions to this poor AUC trend. Three Target districts produced relatively more successful models, averaging AUC = 0.81 for 9th-12th grade students, AUC = 0.709 for 6th-8th grade students, and AUC = 0.666 for 1st-5th grade students. Despite the initial successful results of these three Target districts, they still lacked the data to produce models for all years, with all three missing records for 1st and 2nd grades. When additional historical records become available in the future, these models will almost certainly make it into the pool of Pillar models in future model iterations.

3.2 Feature Importance

We can understand which features are particularly important to each Pillar model by calculating feature importance. Feature importance was calculated using the mean decrease impurity method, sometimes referred to as the *gini importance* [12]. This metric allows us to calculate how much each feature contributes to the model's eventual predictions of a student's outcomes (in this case, risk of dropout). A range of different types of features were found to be important in the four models. The Pillar 1 model relied heavily on features related to student age, attendance and academic achievement. The Pillar 2 model was similar to the Pillar 1 model in that it relied strongly on academic and attendance related features. However, student summative reading scores were also important to the Pillar 2 model. The Pillar 3 model was less similar to the first two Pillars. For Pillar 3, student assessment data and student behavioral data (i.e., disruption, defiance, etc.) were the primary contributors to the model's predicted outcomes, a difference in feature importance that is likely due to a multitude of reasons. Pillar 4 was most similar to Pillar 1 as it also relied heavily on student academic indicators. One reason could be that there were differences in the data availability of features for each district. For example, no assessment data was available for Pillar 1, whereas Pillar 3 had assessment data available for almost all of their historical student records. Another cause could be the difference in the populations of students in each Pillar district. For example, attendance may play a larger role in graduation in urban districts (e.g. Pillar 2), whereas behavioral incidents could play a larger role in the path to dropping out for students in more rural districts (e.g. Pillar 3).

3.3 Performance on New Districts

We applied the Pillar models to each student's data from the 30 Target districts, and averaged the probability across models for each student. These districts had considerable variation in size, graduation rate, and degree of missingness of data (and which features were missing), with values for these variables that were substantially higher or lower than the values for the Pillar districts. In other words, applying models from the Pillar districts to these thirty Target districts represents substantial extrapolation.

Table 3 shows average performance of each individual Pillar model detectors on the Target district data, as well as using averaged probabilities. Despite the high degree of extrapolation required, performance was generally good, with an average AUC (across districts) of 0.783 (*SD* = 0.100), with three districts achieving AUC above 0.9. AUC results within grade band produced similar outcomes, with 9th-12th obtaining an average AUC of 0.813 (*SD = 0.078*), 6th-8th model AUC averaging 0.736 (*SD = 0.13*), and 1st -5th grade model AUC averaging 0.646 (*SD = 0.141*). However, two districts (Target 12 and Target 28) had poor overall AUC values of 0.539 and 0.469. It is worth noting that these two districts had the highest rate of missing data for features that ranked most important in the Pillar models, with over 80% of students in these Target districts missing data related to coursework, over 90% of the records not containing any assessment scores, and the data for 40% of the students not containing attendance information. Overall, the districts with the highest amounts of missing data in core features were also the districts with the lowest AUC ROC values. None of the individual Pillar models did as well as their average when applied to the Target districts; individual Pillar models achieved an AUC between 0.718 - 0.756 on the Target districts, significantly underperforming compared to averaging the produced model probabilities.

Table 3: Average Performance of Pillar Model and Mean Detectors on Target District Data.

Detector	1 – 5	6 - 8	9 - 12	All Grades
Mean Model	0.646	0.74	0.813	0.783
Pillar 1 Model	0.631	0.673	0.749	0.719
Pillar 2 Model	0.568	0.678	0.764	0.718
Pillar 3 Model	0.631	0.720	0.789	0.750
Pillar 4 Model	0.591	0.685	0.788	0.756

3.4 The Chicago Model On-Task Indicator

Comparing our Mean Model detector to the Chicago model detector was limited by data availability: the Chicago model relies on freshman year high school student data, specifically the number of course credits and courses failed during freshman year (9th grade). Due to the detector's reliance on these two data points, our validation sample was limited to only those districts that contained valid information for these two features. However, many of our Target districts lacked data for the features in the Chicago model, for some students. If at least one feature was available for the Chicago model, the model was used; a student was assigned a .5 probability of graduating if the Chicago model was missing all features and therefore incapable of producing a prediction. The Pillar models performed also poorly for these students with very high data missingness.

Table 4: Average AUC Performance of Mean Model vs. Chicago On-Task Indicator: 9th-12th grades

Detector	Avg AUC	Standard Deviation
Mean Model	0.821	0.084
Chicago Model	0.624	0.121

The Mean Model outperformed the Chicago model in every Target district, except for one district. In that district, the Chicago model (*AUC*=0.77) performed .068 better than the Mean Model (*AUC*=0.702). Overall, the mean Pillar model detector achieved an average AUC of 0.197 higher than the Chicago model when measuring performance of predictions on high school student Target district data.

One might argue that this comparison biases against the Chicago model, by including cases where the specific data needed was unavailable. Although our approach is designed to work in cases of high missingness, we can also compare our Mean Model to the Chicago Model only for cases which have the data the Chicago model needs. This resulted in a significant reduction of the data used to calculate AUC metrics, with only 351,902 out of the original 1,202,465 student records able to populate the On-Track Indicator, 29.3% of our initial sample.

Table 5: AUC Performance of Mean Model vs. Chicago On-Task Indicator on Complete Records: 9th-12th grades

Detector	Avg AUC	Standard Deviation
Mean Model	0.874	0.061
Chicago Model	0.734	0.082

Both the Mean Model and the Chicago Model saw an increase in their average model performance across the new sample, with the Mean Model increasing by 0.053 from 0.821 to 0.874 and the Chicago model increasing by 0.11, from 0.624 to a more respectable AUC of 0.734. However, the Mean Model still achieves an AUC 0.14 higher despite these conditions designed to be more favorable to the Chicago Model.

4. CONCLUSIONS

In this paper we propose an approach to predict student risk of not graduating from high school for districts where the quality, quantity, or availability of data is insufficient to produce a satisfactory model of student risk, using an ensemble of models from other districts where data is available. This method achieves good predictive power for students in districts that were not used to develop the model, without any fitting or modification to the models or their application. Furthermore, it achieves substantially better results than a popular alternate approach to predicting at-risk status in new districts, the Chicago model.

It is worth noting that our approach and study have several limitations that should be investigated in future work. Though our sample of Target districts was large, we have not yet applied this method across the full diversity of students in the U.S. In particular, districts with substantial Native American populations or those located in extremely rural regions, such as northern or western Alaska, are not represented in our study. Similarly, we have not studied whether our models are equally good for all subgroups within the school districts–a limitation that is common in the field.

There are several ways in which we could probably improve model performance. Research has shown that contextual factors can contribute to identifying students at risk of dropping out and that factors associated with dropout can differ between populations [7, 15]. Altering the detector to weight the Pillar model probabilities by leveraging characteristic information such as student and school demographics, urbanicity (urban, rural, suburban), and the proportion of military-connected or otherwise highly-mobile students could help account for similarities between students and districts better. Additionally, future iterations of our method could take an empirical approach to selecting the Pillar model weights using measure of similarity based on model performance [31], rather than limiting the approach to the current simple averaging method where features are weighted equally.

Ultimately, the performance of the Mean Model presents new opportunities in identifying students at risk of dropping out for districts with minimal or no data. Given the potential benefit of interventions for at-risk students, this new approach has the capacity to improve the future of student outcomes within a large number of schools where it is not yet possible to develop predictive models. Students educated by districts where data are insufficient can now be presented with greater opportunities through the use of proactive interventions driven by predictive modeling rather than being limited to receiving reactive interventions that are often applied too late, if ever.

5. REFERENCES

[1] Allensworth, E. M., & Easton, J. Q. (2007). What Matters for Staying On-Track and Graduating in Chicago Public High Schools: A Close Look at Course Grades, Failures, and Attendance in the Freshman Year. Research Report. Consortium on Chicago School Research.

[2] Amos, Jason (2008). *Dropouts, Diplomas, and Dollars: US High Schools and the Nation's Economy*. Washington, DC: Alliance for Excellent Education.

[3] Baker, R. S., & Yacef, K. (2009). The state of educational data mining in 2009: A review and future visions. Journal of Educational Data Mining, 1(1), 3-17.

[4] Balfanz, R., Byrnes, V., & Fox, J. H. (2015). Sent home and put off track: The antecedents, disproportionalities, and consequences of being suspended in the 9th grade. In D. J. Losen (Ed.).Closing the School Discipline Gap (pp. 17-30). New York, NY: Teachers College Press.

[5] Balfanz, R., DePaoli, J. L., Ingram, E. S., Bridgeland, J. M., & Fox, J. H. (2016). Closing the College Gap: A Roadmap to Postsecondary Readiness and Attainment. *Civic Enterprises*.

[6] Balfanz, R., Herzog, L., & Mac Iver, D. J. (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.

[7] Balfanz, R., & Legters, N. (2004). Locating the Dropout Crisis. Which High Schools Produce the Nation's Dropouts? Where Are They Located? Who Attends Them? Report 70. Center for Research on the Education of Students Placed at Risk CRESPAR.

[8] Belfield, C. R., & Levin, H. M. (2007). *The price we pay: Economic and social consequences of inadequate education*. Washington, D.C.: Brookings Institution Press.

[9] Bernhardt, V. L., & Bernhardt, V. (2013). Data analysis for continuous school improvement. Routledge.

[10] Bowers, A. J., Sprott, R., & Taff, S. A. (2012). 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, 77-100.

[11] Bowers, A.J., & Sprott, R. (2012). Examining the multiple trajectories associated with dropping out of high school: A

growth mixture model analysis. Journal of Educational Research, 105(3), 176-195.

[12] Breiman, L., & Cutler, A. (2007). Random forests-classification description. Department of Statistics, Berkeley, 2.

[13] Cha, S. H. (2007). Comprehensive survey on distance/similarity measures between probability density functions. International J. Mathematical Models and Methods in Applied Science, vol. 1, no. 4, (pp. 300–307).

[14] Christenson, Sandra L., and Martha L. Thurlow (2004). School Dropouts: Prevention Considerations, Interventions, and Challenges. Current Directions in Psychological Science 13(1): 36–39.

[15] Christle, C. A., Jolivette, K., & Nelson, C. M. (2007). School characteristics related to high school dropout rates. Remedial and Special education, 28(6), 325-339.

[16] Cumpton, G., Schexnayder, D., & King, C. T. (2012). Factors associated with education and work after high school for the classes of 2008 and 2009: A research report of the Central Texas Student Futures Project.

[17] Dekker, G. W., Pechenizkiy, M., & Vleeshouwers, J. M. (2009). Predicting Students Drop Out: A Case Study. *International Working Group on Educational Data Mining.* Paper presented at the International Conference on Educational Data Mining (EDM) (2nd, Cordoba, Spain, Jul 1-3, 2009)

[18] DePaoli, Jennifer L., Joanna Hornig Fox, Erin S. Ingram, et al. (2015). Building a Grad Nation: Progress and Challenge in Ending the High School Dropout Epidemic. Annual Update 2015. Civic Enterprises.

[19] Driscoll, Anne K. (1999). Risk of High School Dropout among Immigrant and Native Hispanic Youth. International Migration Review 33(4): 857–875.

[20] Ensminger, M. E., & Slusarcick, A. L. (1992). Paths to high school graduation or dropout: A longitudinal study of a first-grade cohort. Sociology of education, 95-113.

[21] Finn, Jeremy D. (1989). Withdrawing from School. Review of Educational Research 59(2): 117–142.

[22] Heppen, J. B., & Therriault, S. B. (2008). Developing Early Warning Systems to Identify Potential High School Dropouts. Issue Brief. National High School Center.

[23] Hoffman, Nancy, Joel Vargas, Andrea Venezia, and Marc S. Miller (2007). Minding the Gap: Why Integrating High School with College Makes Sense and How to Do It. ERIC.

[24] Jeni, L. A., Cohn, J. F., & De La Torre, F. (2013). Facing Imbalanced Data--Recommendations for the Use of Performance Metrics. Proceedings of the International Conference on Affective Computing and Intelligent Interaction, 245-251.

[25] Kemple, J. J., Segeritz, M. D., & Stephenson, N. (2013). Building on-track indicators for high school graduation and college readiness: Evidence from New York City. *Journal of Education for Students Placed at Risk (JESPAR)*, *18*(1), 7-28.

[26] Koon, S., & Petscher, Y. (2016). Stated briefly: Can scores on an interim high school reading assessment accurately predict low performance on college readiness exams?

[27] Knowles, J. E. (2015). Of needles and haystacks: Building an accurate statewide dropout early warning system in Wisconsin. Journal of Educational Data Mining, 7(3), 18-67.

[28] Kotsiantis, S. B., Pierrakeas, C. J., & Pintelas, P. E. (2003). Preventing student dropout in distance learning using machine learning techniques. Proceedings of the International Conference on Knowledge-Based and Intelligent Information and Engineering Systems, 267-274.

[29] Lakkaraju, H., Aguiar, E., Shan, C., Miller, D., Bhanpuri, N., Ghani, R., & Addison, K. L. (2015). A machine learning framework to identify students at risk of adverse academic outcomes. In Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining (pp. 1909-1918). ACM.

[30] Legault, Lisa, Isabelle Green-Demers, and Luc Pelletier (2006). Why Do High School Students Lack Motivation in the Classroom? Toward an Understanding of Academic Amotivation and the Role of Social Support. Journal of Educational Psychology 98(3): 567.

[31] McCune, B., Grace, J. B., & Urban, D. L. (2002). Analysis of ecological communities (Vol. 28). Gleneden Beach, OR: MjM software design.

[32] Neild, R. C., Stoner-Eby, S., & Furstenberg, F. (2008). Connecting entrance and departure: The transition to ninth grade and high school dropout. Education and Urban Society, 40(5), 543-569.

[33] Peugh, J. L., & Enders, C. K. (2004). Missing data in educational research: A review of reporting practices and suggestions for improvement. Review of educational research, 74(4), 525-556.

[34] Rafa, A. (2017). Chronic absenteeism: A key indicator of student success. Policy Analysis. Education Commission of the States. Retrieved from https://files.eric.ed.gov/fulltext/ED574526.pdf

[35] Rumberger, R. W. (1987). High school dropouts: A review of issues and evidence. Review of educational research, 57(2), 101-121.

[36] Schildkamp, K., Lai, M. K., & Earl, L. (Eds.). (2012). Data-based decision making in education: Challenges and opportunities (Vol. 17). Springer Science & Business Media.

[37] Stark, Patrick, and Amber M. Noel (2015). Trends in High School Dropout and Completion Rates in the United States: 1972-2012. Compendium Report. NCES 2015-015. National Center for Education Statistics.

[38] Stuit, D., O'Cummings, M., Norbury, H., Heppen, J., Dhillon, S., Lindsay, J., & Zhu, B. (2016). Identifying early warning indicators in three Ohio school districts. National Center for Education Evaluation and Regional Assistance.

[39] Venables, W. N., & Ripley, B. D. (2002). Tree-based methods. In Modern Applied Statistics with S (pp. 251-269). Springer, New York, NY.