

Early Indication Tool

Rationale, Methods and Results

David M. Alexandro, Ph.D.

Charles W. Martie, Ph.D.



TABLE OF CONTENTS

INTRODUCTION.....	1
RATIONALE	2
Overview	2
Special Populations Must Be Supported	3
Interventions and Support Must Follow Early Warning System Predictions.....	4
Connecticut’s Data and Academic Milestones	5
Machine Learning Helps to Understand Patterns in Data and Solve Problems	7
METHODS	9
EIT Models Use Supervised and Unsupervised Learning Techniques	9
Predictors of Academic Milestones Include the ABCs and more	9
Using Balanced Training Datasets Addressed Class Imbalance	12
Using Classification Accuracy Measures and Threshold Optimization Helped Improve Model Performance.....	13
Statistical software and hardware specifications	16
Dataset	17
Analysis	18
RESULTS	20
Balanced Accuracy	20
Variable Importance Rankings	21
DISCUSSION	26
Implications of Results	26
Conclusion.....	29
REFERENCES.....	30
APPENDIX.....	36

INTRODUCTION

In response to the high school dropout crisis, which comes with great economic and social costs, early warning systems (EWSs) have been developed to systematically predict and improve student outcomes. The CSDE created its EWS—the Early Indication Tool (EIT)—as a kindergarten through 12th grade (K-12) system that identifies students who may need additional support to reach academic milestones and facilitates timelier, targeted interventions. The EIT is a critical support component in [Connecticut's ESSA Plan](#) (U.S. Department of Education, 2017). Ultimately, CSDE wants more students to meet academic milestones and graduate from high school.

For the EIT, CSDE developed a unique model for each grade from first grade through Grade 12. The EIT assigns each student to a targeted support level (High, Medium, or Low) based on the individual's likelihood of meeting the academic milestone corresponding to her or his grade. The aim of the EIT is to better identify those students in need of targeted support and inform on-the-ground practitioners who may intervene long before students may be dropping out.

Connecticut's early grade models use factors such as attendance, assessments, disciplinary incidents, free-or-reduced price meal eligibility status, and student mobility to group students using modeling approaches including latent profile analysis (LPA)¹ and random forests.² This combination of modeling approaches takes advantage of sophisticated machine learning algorithms to determine which variables are most essential to predictions, considers how students are clustering together on these variables, and avoids overfitting of models.³ As students advance to middle school and high school, Connecticut's predictive models incorporate course-level variables including course enrollments and course performance. In addition, the EIT models use English learner (EL) status and detailed special education data as predictors, including primary disability, hours of special education services received, and percent of time with non-disabled peers. Finally, school- and district-level predictors are included to capture factors beyond the student level.

¹ Latent Profile Analysis or LPA is a person-centered statistical method of clustering that allows us to identify hidden groups of students in the data based on similar characteristics on the various observed data elements.

² Random Forest is a machine-learning algorithm that can be used for classification tasks. This approach creates a prediction model based on an ensemble of hundreds of decision trees (hence “forest”) that are uncorrelated to each other due to random subsetting of training records and fields (hence “random”). By using not just one but hundreds of decision trees, the model achieves greater accuracy and stability in classifying and predicting.

³ Overfitting occurs when the estimated model performs well with the original data, but poorly when applied to other datasets.

RATIONALE

Overview

For decades, educational researchers have studied high school dropout in efforts to improve student outcomes, especially for students from low-income families, students of color, English learners (ELs), and students with disabilities (SWD) (e.g., DePaoli, Balfanz, Atwell, & Bridgeland, 2018; Ekstrom, Goertz, Pollack, & Rock, 1986; Rumberger, 2011). Although dropout rates have improved over that span, the national high school graduation rate is still below 85 percent, and researchers and policymakers have called the college and career readiness (CCR) of America's high school students into question (DePaoli et al., 2018). While this issue has been at the center of educational research and reform efforts (e.g., Belfield, 2007; Belfield & Levin, 2007; Ferguson, 2007; Rebell, 2007; Rumberger, 2011), the relative lack of public awareness of staggering dropout rates has prompted some experts to deem this problem a "silent epidemic" (Bridgeland, DiIulio, & Morison, 2006, p. 1). The EIT uses education data to systematically make predictions regarding the likelihood of students meeting academic milestones and seeks to help target interventions in order to increase the number of students that are meeting these milestones and graduating from high school college-and-career ready.

High school dropout is a complex issue. Rumberger (2011) describes dropout as a process and problem with four dimensions: nature, consequences, causes, and solutions. With more than 7,000 American students dropping out of high school each day (Rumberger, 2011), the dropout crisis comes with great economic and social costs (Belfield & Levin, 2007). Several studies have explored predictive factors from elementary school through early high school to identify students who are at-risk of dropping out (e.g., Allensworth & Easton, 2005, 2007; DePaoli, Balfanz, & Bridgeland, 2016; Ekstrom et al., 1986). Researchers have found that the ABCs (i.e., attendance, behavior, and course performance/credit accrual) are most predictive of high school dropout (e.g., Mac Iver & Messel, 2013). However, the ABCs are not the sole predictors of missing academic milestones.

Rumberger and his colleagues identified a host of factors that are predictive of dropout, including student (demographics, achievement, attitudes, behaviors); family (parental education, family socioeconomic status [SES], family structure, parental employment, family size, parenting practices, parenting expectations, sibling dropout); school (school composition, school size, resources, academic climate, disciplinary climate, teaching quality); and community (unemployment rates) variables (e.g., Rumberger, 2011; Rumberger & Larson, 1998; Rumberger & Lim, 2008; Rumberger & Palardy, 2005;

Rumberger & Thomas, 2000). Furthermore, researchers have cited the role that specific triggers play in the dropout process, including housing, money, criminal or legal issues, accidents or health problems, suspensions, pregnancy, and personal relationships (Dupéré et al., 2018). Armed with these findings, educators and policymakers have looked to target student interventions and support via EWSs.

Schools and districts have implemented many interventions to raise the high school graduation rate. Levin and Belfield (2007) have concluded that improvements can result from fine-tuning factors from kindergarten through 12th grade, including academic expectations, school and class sizes, personalization, counseling, parental engagement, instructional time, and personnel. Additionally, researchers have found that there is a correlation between school climate and graduation rates (e.g., Boyd, 2016; Freeman et al., 2015). Navigating the vast array of intervention options is challenging, since a one-size-fits-all solution to the dropout problem does not exist.

Predictive modeling is a core component of EWS development. Educational researchers have used student data to develop predictive models to identify students at risk of a host of troublesome outcomes, including dropping out (e.g., Allensworth, 2013). CSDE has developed models to estimate predicted probabilities of meeting academic milestones; the students with the lowest probabilities are deemed most in need of targeted support.

Special Populations Must Be Supported

Although some states and districts have shown incredible progress, there are still low-performing schools and disparities in national graduation rates for students of color (76.4% for Black students and 79.3% for Hispanic students, compared with 88.3% for white students in 2016) and special populations, including students with disabilities (65.5%), students from low-income families (77.6%), and students with limited English proficiency (66.9%) (DePaoli et al., 2018).

In Chicago, researchers examined the graduation rates for students with disabilities and English learners (Gwynne, Lesnick, Hart, & Allensworth, 2009; Gwynne, Pareja, Ehrlich, & Allensworth, 2012). The authors not only found that there was a graduation rate disparity between the major categories (i.e., SWD and non-SWD, ELs and non-ELs), but graduation rates also varied greatly across SWD categories. In fact, four-year graduation rates were below 50 percent for students two or more years below grade level in grade 9, students with learning disabilities, students with mild cognitive disabilities, and students with emotional disturbances.

Malin, Bragg, and Hackmann (2017) expressed concern that if graduating from high school and college and career readiness are “not recognized as important for all students, the nation risks perpetuating inequities among student groups that may have a lasting detrimental impact on society” (p. 813). Wilkins and Bost (2016) acknowledged that implementing early warning systems and other interventions has increased graduation rates of students with disabilities, but cautioned educational leaders to review data regularly, and revise and review school policies accordingly. Balfanz and Legters (2004) asserted the importance of targeting a relatively small number of failing high schools:

High schools with weak promoting power are the engines driving the low national graduation rate for minority students...These high schools must be specifically targeted for reform...Transforming the nation’s dropout factories into high schools that prepare all their students for post-secondary schooling or training and successful adulthood should thus be an urgent national priority. (p. 23)

The Every Student Succeeds Act (ESSA) has marshalled in a “new environment of accountability” in which federal funding to states requires evidence of improved outcomes for all students (Hanover Research, 2018, p. 6).

Interventions and Support Must Follow Early Warning System Predictions

Early warning systems employ models that depend on available data to predict everything from bioterrorism (e.g., Berkowitz, 2013) to landslides (e.g., Battistini et al., 2017). The key to any EWS is the intervention and/or support that follows the prediction. Balfanz (2009, 2011, 2014, 2016) is a leader in the development and dissemination of EWS research in education. He asserts, “Early warning and intervention systems provide the necessary means to unify, focus, and target efforts to improve attendance, behavior, and course performance. Their fundamental purpose is to get the right intervention to the right student at the right time” (Balfanz, 2009, p. 10). He and his colleagues have written extensively about their findings and have highlighted the importance of students being engaged and being at school (e.g., Balfanz & Byrnes, 2012; Balfanz, Byrnes, & Fox, 2014).

While she does not dispute that interventions must be on-time and on-target, Scala (2015) cautions educators and policymakers against making causal claims: “Early warning indicators are used only for

prediction—they *do not cause* students to drop out. Rather, they should be *treated as symptoms* of the dropout process that is in progress” (p. 8, emphasis in original). Since these symptoms exhibit themselves at different times, researchers have made efforts to study indicators and outcomes from pre-kindergarten to the end of the student life cycle.

Many researchers have conducted studies using large datasets at the city and state levels to develop EWSs to improve student outcomes. As of 2013, more than 30 state departments of education had early warning systems (Data Quality Campaign, 2013). The Every Student Succeeds Act (2015) expanded state responsibility over schools, and this legislation is driving all states to develop EWSs and other accountability systems to support local education agencies (Civic Impulse, 2017).

Connecticut’s Data and Academic Milestones

Recent changes in graduation requirements (Connecticut General Assembly, 2017), as well as Connecticut’s adoption of Next Generation Science Standards (NGSS), Smarter Balanced Assessment Consortium (SBAC) mathematics and English language arts (ELA) assessments, and the Next Generation Accountability System (CSDE, 2016, 2017, 2018), have created unique opportunities to develop prediction models that incorporate new and relevant data. The CSDE data warehouse contains the requisite data to train prediction models that integrate course-, school- and district-level data and standardized assessments with other student-level variables to predict a host of outcomes, including college and career readiness (CCR). Each model includes a large pool of predictors in order to determine student probabilities of meeting academic milestones.

Table 1 provides an overview of the academic milestones modeled by the EIT for students in different grades. As the table shows, there are four outcomes of interest, each covering students across a three-grade band. For students in grades 1 to 3, the EIT is considering students’ probabilities of reaching reading proficiency by the end of third grade, as measured by meeting or exceeding expectations on the ELA SBAC assessment administered at the end of third grade. For students in grades 4 to 6, the EIT is considering students’ probabilities of meeting or exceeding expectations on the mathematics and ELA SBAC assessments administered at the end of sixth grade. For students in grades 7 to 9, the EIT is considering students’ probabilities of being on-track at the end of ninth grade. Finally, for students in grades 10 to 12, the EIT is considering students’ probabilities of being college and career ready.

Table 1

Academic milestones modeled by the EIT

Grades	Academic Milestone
1-3	Proficient in reading by the end of Third Grade <ul style="list-style-type: none"> Meeting or exceeding expectations on the 3rd Grade ELA SBAC
4-6	Prepared for Middle School <ul style="list-style-type: none"> Meeting or exceeding expectations on the 6th Grade ELA and mathematics SBACs
7-9	Prepared for High School <ul style="list-style-type: none"> On-track to High School Graduation in Grade 9
10-12	College and Career Ready <ul style="list-style-type: none"> Meeting assessment and course-passing benchmarks

The University of Chicago Consortium on School Research (CCSR) made considerable efforts to study the transition into high school and its relationship with high school success (e.g., Allensworth & Easton, 2005, 2007). The CCSR concluded that a 9th grade on-track indicator combining information on credits and grades earned during freshman year is a stronger predictor of high school graduation than standardized tests. Following the CCSR's lead, CSDE adopted the On-track in 9th grade indicator as a central component of Connecticut's Next Generation Accountability System (CSDE, 2016, 2017, 2018). Table 2 summarizes the on-track criteria that is used for the EIT.

Table 2

Criteria for On Track to High School Graduation in Grade 9 indicator for EIT

Number of semester F's in core courses (1 semester course = 0.5 credit)	Number of credits accumulated during Grade 9 (1 full-year course = 1 credit)	
	Less than 5.0	5.0 or more
2 or more	Off-track	Off-track
0 to 1	Off-track	On-track

Note. Students who fail one full year (i.e., two semesters) of a core course and/or earn less than five total credits during 9th grade are deemed off-track. English, mathematics, science, and social studies are core courses for the purposes of the on-track indicator.

The CCR outcome for students in grades 10 to 12 includes assessment and course-passing components. For purposes of the EIT, College and Career Ready means a student who, by the end of high school, has:

- achieved proficiency on the SAT in both Evidence-Based Reading and Writing (EBRW; 480 or higher) and Mathematics (530 or higher); and
- achieved one or more of the following:
 - passed two courses combined in Advanced Placement (AP), International Baccalaureate (IB), or dual enrollment; or
 - passed two courses in one of 17 Career and Technical Education clusters (CTE); or
 - passed two workplace experience courses.

Machine Learning Helps to Understand Patterns in Data and Solve Problems

Given the importance of prediction models for efforts to identify at-risk students via EWSs and inform practitioners who may intervene, it is necessary to understand the approaches to statistical modeling that undergird these models. Forecasting the academic milestones for Connecticut's students includes the prediction of a binary outcome (e.g., on-track or off-track) from quantitative and categorical independent variables. When creating a model of this type, a logistic regression model is often used, particularly in the social sciences and education (Cizek & Fitzgerald, 1999). Fortunately, new modeling approaches are available that improve on the predictive accuracy of logistic regression models.

Machine Learning. Machine learning involves the use of data mining techniques and computer algorithms to understand patterns in data to solve problems. Conway and White (2012) place machine learning “at the intersection of traditional mathematics and statistics with software engineering and computer science” (p. 1), and Ng describes it as “the science of getting computers to act without being explicitly programmed” (2013, para. 1).

Data mining is distinct from classical statistical methods and covers “a variety of exploratory data analysis techniques that were developed in statistics and computer sciences for analyzing large amounts of data” (Strobl, 2013, p. 678). There are supervised and unsupervised approaches to machine learning. Supervised learning occurs when outcomes are used in the preprocessing of data, such as techniques to classify a set of observations into groups that are directly observed. In unsupervised learning, the outcomes are not used in the preprocessing, as in clustering techniques designed to sort a set of observations into latent or unobserved groups (Kuhn & Johnson, 2013).

Supervised learning techniques train prediction models using observed outcomes. Models created using supervised learning modeling techniques such as classification and regression tree (CART, or decision tree) and random forests benefit from the flexibility of not being constrained by assumptions about the functional form and distribution of the data, which is in stark contrast to parametric models like logistic regression (Strobl, 2013). However, since the relationship between the predictors and outcome is not explicitly reported, data mining is often called a “black box” approach (Breiman, 2001b; Kuhn & Johnson, 2014; Veltri, 2017). Still, their automated data processing and ability to handle and select large numbers of variables at a time make CART and random forests ideal candidates for solving classification problems.

Classification Trees. Decision trees are built by finding variables and cut-points that can be used in combination for yes-no questions to best predict classifications. The optimal classification tree follows the principle of impurity reduction, by which “each split in the tree-building process results in daughter nodes that are more ‘pure’ than the parent node in the sense that groups of subjects with a majority for either response class are isolated” (Strobl, 2013, p. 684).

Random Forests. Random forests is called an ensemble method, since it aggregates the predictions of several decision trees using a bootstrap approach (Breiman, 2001a; Strobl, 2013). This addresses a major disadvantage of CART models: the structure (including splitting variables and cut-points) and predictions of single decision trees are highly variable. With random forests, the forest makes a prediction by tallying votes across all decision trees contained therein. These models capture complex interactions between predictors. Moreover, by drawing samples with replacement—random samples of both data and predictor variables—and aggregating the results, decision boundaries are smoother than those established with a single tree, and the random variation that went into creating the forest of decision trees results in a diverse grouping of splits and predictor variables.

The random selection of splitting variables in random forests creates unique opportunities for all variables. In some datasets, certain variables are clearly preferable for impurity reduction when constructing decision trees. However, since random forests involve a random sampling of records and variables, the strongest splitting variables are excluded from some decision trees in the random forest. “If the stronger competitor cannot be selected then a new variable has a chance to be included in the model and may reveal interaction effects with other variables that otherwise would have been missed” (Strobl, 2013, p. 693). Although random forests do not produce coefficients like regression models, variable importance measures allow for the ranking of which predictors were most crucial in optimizing the model (Breiman, 2001a).

Unsupervised learning techniques seek to determine whether groups exist by identifying how individual records “hang together” (i.e., cluster on the variables of interest). Latent profile analysis (LPA) is a direct application of finite mixture modeling that allows for person-centered analysis: Instead of correlations among variables being of most interest, the relationship among individuals with respect to the variables of interest is the central concern (Pastor et al., 2007). The approach seeks to identify whether underlying (i.e., latent or unobserved) groups of students exist, determine what distinguishes the

groups from one another (i.e., characterize the group profiles), and assign each individual to a group based on her or his observed data. One of the major benefits of mixture model approaches like LPA is that they allow for potential uncertainty with the classification of each case (Morgan et al., 2016).

METHODS

EIT Models Use Supervised and Unsupervised Learning Techniques

The methods used to create EIT models involved data preparation and data handling in addition to model training, testing, and comparison before the working models were established.

Model development. For the EIT, CSDE created models using supervised and unsupervised learning techniques. Supervised learning techniques were used to develop random forests models for the EIT, and unsupervised learning techniques were used to develop latent profile analysis (LPA) models. For the younger grades, a hybrid model was developed that integrated the probabilities from LPA and random forests models to classify students by targeted support level. For middle school and high school students, random forests models were used to classify students by targeted support level. All models were developed in R. Random forests models were developed using the *randomForest* package (Breiman, Cutler, Liaw & Wiener, 2018); LPA models were developed using the *mclust* package (Fraley, Raftery, Murphy, & Scrucca, 2012). The improvement in performance of random forests models was negligible above 500 trees.

Predictors of Academic Milestones Include the ABCs and more

Beyond the ABCs (i.e., attendance, behavior, and course performance/credit accrual), the EIT models include a range of student-, school-, and district-level variables as predictors of academic milestones. Tables 3 through 6 provide an overview of these fields. For all models, grade-level-specific predictors are limited to those from previous grades. So, for a student entering 3rd grade, data through the end of 2nd grade is used to predict whether that student will make sufficient progress by the end of 3rd grade to meet ELA proficiency on the Grade 3 SBAC. When available, the previous two years of data were generally used for attendance, behavior, mobility, assessment, and course performance fields when constructing a model. Otherwise, the most recent year's information was used for all predictors. Table 9 (see Appendix) provides data definitions and additional details for the full list of fields.

Table 3
Predictors Used in EIT Models for Grades 1 to 3

Domain	Elements	Grade levels for which data is available		
		K	1	2
Student demographics	Free and reduced lunch (FRL) eligibility; EL; SWD; age in grade	X	X	X
Attendance	Percentage of school days attended	X	X	X
Behavior	In-school and out-of-school suspensions	X	X	X
Mobility	Schools and districts attended; number of school and district moves outside of the natural progression ^a	X	X	X
Special Education	Primary disability (if applicable); percentage of time with non-disabled peers (TWNDP); hours of special education services	X	X	X
Retention	Grades repeated	X	X	X
Kindergarten Entrance Inventory (KEI)	Literacy, Numeracy, Language, Personal, Creative, and Physical scores on the KEI	X		
Performance Index	Performance index values for school and district	X	X	X
School and district demographics	Enrollment; percent minority; percent high needs; percent poverty; chronic absence rate	X	X	X
Cohort	Cohort aggregates of above	X	X	X

^a A natural progression school move is one in which the student changes schools because of a district's school structure (e.g., a middle school enrollment when the elementary school does not provide the subsequent grade)

Table 4
Predictors Used in EIT Models for Grades 4 to 6

Domain	Elements	Grade levels for which data is available		
		3	4	5
Student demographics	Free and reduced lunch (FRL) eligibility; EL; SWD; age in grade	X	X	X
Attendance	Percentage of school days attended	X	X	X
Behavior	In-school and out-of-school suspensions	X	X	X
Mobility	Schools and districts attended; number of school and district moves outside of the natural progression	X	X	X
Special Education	Primary disability (if applicable); percentage of time with non-disabled peers (TWNDP); hours of special education services	X	X	X
Retention	Grades repeated	X	X	X
SBACs	SBAC mathematics and English language arts (ELA) scale scores	X	X	X
Performance Index	Performance index values for school and district	X	X	X
School and district demographics	Enrollment; percent minority; percent high needs; percent poverty; chronic absence rate	X	X	X
Cohort	Cohort aggregates of above	X	X	X

Table 5
Predictors Used in EIT Models for Grades 7 to 9

Domain	Elements	Grade levels for which data is available		
		6	7	8
Student demographics	Free and reduced lunch (FRL) eligibility; EL; SWD; age in grade	X	X	X
Attendance	Percentage of school days attended	X	X	X
Behavior	In-school and out-of-school suspensions	X	X	X
Course performance	Course enrollments, including subject area, rigor, and available credits; credits earned and failed		X	X
Mobility	Schools and districts attended; number of school and district moves outside of the natural progression	X	X	X
Special Education	Primary disability (if applicable); percentage of time with non-disabled peers (TWNDP); hours of special education services	X	X	X
Retention	Grades repeated	X	X	X
SBACs	SBAC mathematics and ELA scale scores	X	X	X
Performance Index	Performance index values for school and district	X	X	X
School and district demographics	Enrollment; percent minority; percent high needs; percent poverty; chronic absence rate	X	X	X
Cohort	Cohort aggregates of above	X	X	X

Table 6
Predictors Used in EIT Models for Grades 10 to 12

Domain	Elements	Grade levels for which data is available			
		8	9	10	11
Student demographics	Free and reduced lunch (FRL) eligibility; EL; SWD; age in grade	X	X	X	X
Attendance	Percentage of school days attended	X	X	X	X
Behavior	In-school and out-of-school suspensions	X	X	X	X
Course performance	Course enrollments, including subject area, rigor, and available credits; credits earned and failed	X	X	X	X
Mobility	Schools and districts attended; number of school and district moves outside of the natural progression	X	X	X	X
Special Education	Primary disability (if applicable); percentage of time with non-disabled peers (TWNDP); hours of special education services	X	X	X	X
Retention	Grades repeated	X	X	X	X
SBACs	SBAC mathematics and ELA scale scores	X			
PSATs	PSAT mathematics and EBRW scale scores			X	
Performance Index	Performance index values for school and district	X	X	X	X
School and district demographics	Enrollment; percent minority; percent high needs; percent poverty; chronic absence rate	X	X	X	X
Cohort	Cohort aggregates of above	X	X	X	X

Missingness of data was used as a predictor in EIT models. The missingness of assessment scale scores was treated as a predictor: Missing scores were imputed, a flag was retained to indicate whether each scale score was actual or imputed, and all of these variables were included as covariates when training the models. This approach increased the number of student records on which the models were trained; more important, it increased the number of students for whom the prediction models could be applied. Since students with disabilities, students of color, and English learners are disproportionately represented among those with missing scores, imputing assessment scale score values was a critically important technique to ensure the maximum possible number of records were retained for these important student groups.

Using Balanced Training Datasets Addressed Class Imbalance

Classification problems that involve predicting high school graduation or being on-track to milestones such as graduation involve class imbalance, since there is a large discrepancy between the size of the majority (e.g., graduate, on-track) and minority (e.g., non-graduate, off-track) classes. The minority class is commonly called the positive class, since “the interest usually leans towards correct classification of the ‘rare’ class” (Chen, Liaw, & Breiman, 2004, p. 1). Among EIT academic milestones, the On-track in 9th Grade indicator demonstrates the largest imbalance between classes: In Connecticut, more than 85 percent of students meet the criteria for the On-track in 9th Grade indicator at the end of their freshman year, which means less than 15 percent of students are in the minority (or positive) class for this outcome.

To address class imbalance, EIT models were developed using balanced training samples. Balanced training sets were created using oversampling (also known as upsampling), a technique which retains all records from the majority class (e.g., on-track) and creates a bootstrap sample of cases with replacement from the minority (e.g., off-track) class to balance the number of on-track and off-track records. This method generally improves classification accuracy for positive (i.e., rare or off-track) cases at the expense of decreased classification accuracy for negative cases (Haixiang et al., 2017). A goal in developing the EIT models was to increase the odds of correctly identifying students who will be off-track.

Figure 1 shows how the training and validation samples were created. In each two-colored cylinder, the top portion (in green) represents on-track records, and the bottom section (in red) represents off-track records. The imbalanced training and validation sets maintain the same class imbalance as the sample.

The balanced training dataset contains an equal number of records from the positive and negative classes.

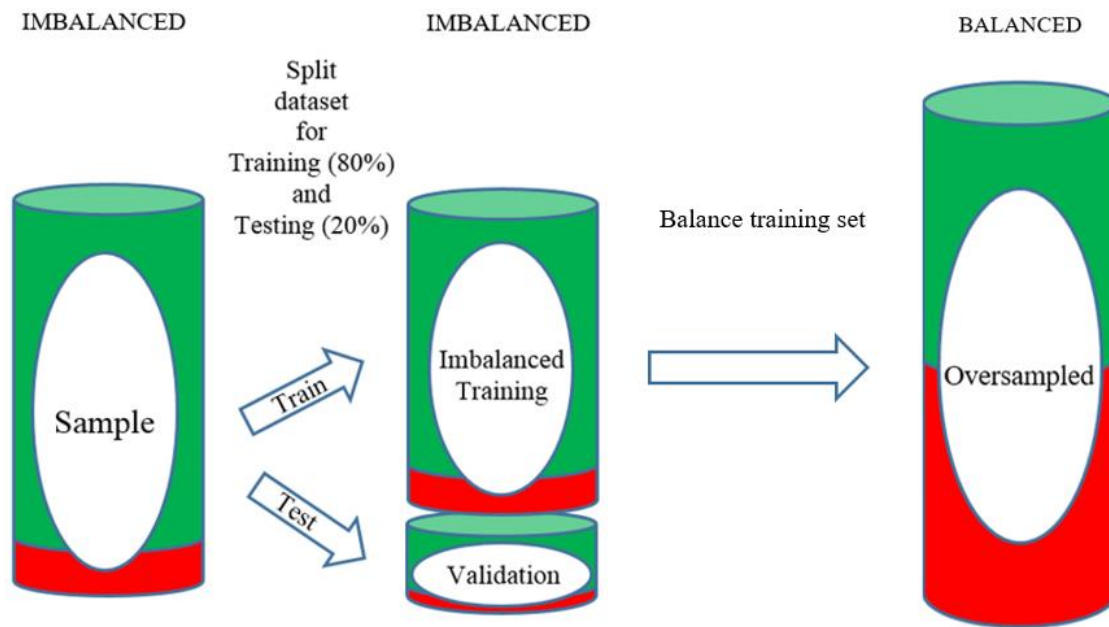


Figure 1. Flowchart to explain how training and validation samples were created

Using Classification Accuracy Measures and Threshold Optimization Helped Improve Model Performance

Classification accuracy measures. The validation dataset was used to test the models, and true-positives (TP), false-positives (FP), false-negatives (FN), and true-negatives (TN) for predicted and true conditions were determined for all models. In those four designations, the true/false indicator identifies whether the predicted classification was correct/incorrect, and the positive/negative indicator denotes the predicted class as off-track /on-track (i.e., not meeting/meeting the academic milestone). In addition, following the recommendation of Bowers, Spratt, and Taff (2013), the precision, sensitivity, specificity, and false-positive rate (FPR; also known as the false-positive proportion [FPP] or $1 - \text{Specificity}$) were also considered for all models. Lastly, AUC (i.e., area under the receiver operating characteristic [ROC] curve), accuracy and balanced accuracy were considered for all models. These measures are explained and the related equations are presented in the *Accuracy Equations* section below.

The contingency table (also known as confusion matrix) shown in Figure 2 summarizes how true and predicted conditions were compared to determine TP, FN, FP, and TN values for all models.

		True condition		
		Condition positive (Off-track)	Condition negative (On-track)	
Predicted condition	Predicted condition positive (Off-track)	<i>a</i> True-positive (TP) Correct	<i>b</i> False-positive (FP) Type I Error	<i>a + b</i> (TP + FP)
	Predicted condition negative (On-track)	<i>c</i> False-negative (FN) Type II Error	<i>d</i> True-negative (TN) Correct	<i>c + d</i> (FN + TN)
		<i>a + c</i> (TP + FN)	<i>b + d</i> (FP + TN)	<i>a + b + c + d</i> (<i>N</i>)

Figure 2. Contingency table (Adapted from Bowers et al., 2013, p. 83)

Accuracy Equations. The equations for calculating classification accuracy measures for each model are an essential component in evaluating and comparing models. All components of the equations can be found in the contingency table in Figure 2. The *confusionMatrix* function in the caret package (Kuhn, 2018) was used to calculate a cross-tabulation of observed and predicted classes and all related statistics in R.

Since accuracy, sensitivity, specificity, and balanced accuracy are metrics that are commonly used to select between classification models, these classification accuracy measures are explained below. Accuracy (also known as the overall accuracy rate) represents the proportion of correct predictions among all cases in the validation sample.

$$\text{Accuracy} = (\text{TP} + \text{TN}) / N \quad (1)$$

Sensitivity (also known as recall or true-positive rate [TPR]) measures the proportion of correct predictions among all observed positive cases in the validation sample.

$$\text{Sensitivity} = \text{TP} / (\text{TP} + \text{FN}) \quad (2)$$

Specificity (also known as true-negative rate [TNR]) measures the proportion of correct predictions among all observed negative cases in the validation sample.

$$\text{Specificity} = \text{TN} / (\text{TN} + \text{FP}) \quad (3)$$

Balanced accuracy is an average of the sensitivity and specificity, and it measures the average accuracy in classifying minority and majority class observations.

$$\text{Balanced Accuracy} = (\text{Sensitivity} + \text{Specificity}) / 2 \quad (4)$$

The balanced accuracy measure is particularly helpful when evaluating models involving rare events or imbalanced data, since overall accuracy rates are weighted and often high due to the classifier favoring the majority class when the validation data is imbalanced. In these cases, although the balanced accuracy will be lower than the overall accuracy rate, the balanced measure helps researchers identify which model does the best job of classifying both minority and majority class observations. While all accuracy measures were calculated using an imbalanced validation dataset, employing oversampling to balance the training datasets helped improve balanced accuracy results.

EIT models maximized balanced accuracy instead of overall classification accuracy. Unfortunately, privileging the negative class by trying to maximize overall classification accuracy results in poor classification accuracy for cases in the crucial rare class. For example, consider a dataset in which 95 of 100 records are graduates (i.e., on-track). Simply classifying all records as graduates yields a 95 percent overall classification accuracy and 100 percent specificity (i.e., the proportion of on-track students who were correctly classified), but it classifies all five non-graduates as graduates—a 0% sensitivity that misclassifies all rare cases in the positive class—and results in a balanced accuracy of 50%. Optimized classification models would look to improve on this simple classification rule by improving accuracy measures. Table 7 provides an overview of the values associated with this example to allow for a clear comparison of accuracy measures.

Table 7
Comparison of accuracy measures

Comparison of accuracy measures					
	Measure	Base Model	Model A	Model B	
True positives: The number of correct off-track predictions	TP	0	2	4	False negatives: The number of incorrect on-track predictions
False positives: The number of incorrect off-track predictions	FN	5	3	1	
	FP	0	1	4	True negatives: The number of correct on-track predictions
Accuracy: The proportion correct among all predictions	TN	95	94	91	
	Accuracy	0.95	0.96	0.95	Balanced Accuracy: (Sensitivity + Specificity)/ 2
Sensitivity: TP / (TP + FN)	Balanced Accuracy	0.50	0.69	0.88	
	Sensitivity (TPR or Recall)	0.00	0.40	0.80	Specificity: TN / (TN + FP)
Precision: TP / (TP + FP)	Specificity	1.00	0.99	0.96	
	Precision (PPV)	-	0.67	0.50	

Note. The Prevalence was 0.05 – 5% of students in the sample were non-graduates (i.e., off-track)

For instance, consider Model A: It classifies 97 records as graduates (including 94 correctly classified) and three records as non-graduates (with two correctly classified), resulting in an increase in overall classification accuracy (96%), sensitivity (40%, with two of five actual non-graduates correctly classified in the prediction model), and balanced accuracy (69%). A second model (Model B) classifies

92 records as graduates (91 of which are correctly classified) and eight records as non-graduates (four correct). Model B results in a decrease in overall classification accuracy (95%) from Model A, but an increase in sensitivity (80%, since four of five rare cases are correctly classified) and balanced accuracy (88%). If resources were limited such that an intervention could only be administered to five students, Model B may be preferred, since it improves on the balanced accuracy and doubles the sensitivity of Model A with minimal loss in specificity and overall classification accuracy. All EIT models were developed with the crucial rare class in mind, and the final EIT models sought to maximize balanced accuracy instead of overall classification accuracy.

Thresholds. Aside from data preprocessing, another approach in dealing with data imbalance is to adjust the threshold (also known as cut-off or cut-point) criteria. Since random forests models result in a probabilistic prediction for the classification for each student record, it is straightforward to assess the predictive accuracy of the models. While it is common to use a 0.5 threshold to assign binary classes based on probabilistic predictions (i.e., cases with a probability of 0.5 or higher of belonging to a particular class are assigned to that class; all others are assigned to the alternate class), this threshold “does not necessarily preserve the observed prevalence [i.e., the overall proportion of cases in which a particular outcome is observed] or result in the highest prediction accuracy, especially for [rare events] data sets with very high or very low observed prevalence” (Freeman & Moisen, 2008, p. 48). Researchers have compared the performance of threshold criteria to improve binary classification models and have shown that adjusting the threshold criteria is a valuable method for model selection and optimization (e.g., Freeman & Moisen, 2008).

For each EIT model, the classification accuracies of different threshold criteria were compared to select the best model and optimize model performance. Optimal thresholds were determined using Youden’s index (also known as Youden’s J statistic), which is the difference between the TPR (i.e., sensitivity) and FPR (i.e., 1 – Specificity) (Youden, 1950). For each model, the cut-point that maximized Youden’s index was selected as the optimal threshold.

$$\text{Youden's index} = \text{Sensitivity} - (1 - \text{Specificity}) = \text{Sensitivity} + \text{Specificity} - 1 \quad (5)$$

Statistical software and hardware specifications

The statistical software used to develop these models included SAS Enterprise Guide Version 7.15 (SAS Institute Inc., 2018), SAS Data Integration Studio Version 4.903 (SAS Institute Inc., 2018), R Version

3.5.0 (R Core Team, 2018), and RStudio Version 1.1.453 (RStudio Team, 2016). The SAS programs were used to acquire the data, and R programs were executed for data cleaning, preparation, analysis, and modeling. The SAS software ran on a computer with the following specifications: Windows 7 Enterprise; 64-bit Operating System; 8 GB of RAM; and Intel Core i5-4570T CPU @ 2.90GHz processor. The R and RStudio software ran on a computer with the following specifications: Windows Server 2012 R2 Datacenter; 64-bit Operating System, x64-based processor; 128 GB of RAM; and Intel Xeon Gold 6128 CPU @ 3.40GHz processor (2 processors).

Dataset

For each model, the training data consisted of data collected from the population of Connecticut public school students who were in the model's grade in the previous academic year.

Samples. While the full dataset contained the students who were in the model's grade in the previous year, the sample for each grade included records for which attendance was not missing. In addition, models for grades 8 to 12 required that the previous year's record included attendance and course-level data. The dataset was in wide format, structured so that each row corresponded to one unique student, and each column corresponded to a single variable. All continuous predictor variables were standardized using student-level standard deviations prior to model development, and samples were created from the cohort described above.

To handle missing data, listwise deletion was used to extract complete records; a complete record is any student record for which all corresponding non-assessment fields for a particular sample have a value. Complete assessment data was not required; the *mice* (Multivariate Imputation via Chained Equations) package in R (van Buuren & Groothuis-Oudshoorn, 2011) was used to impute missing Smarter Balanced Assessment Consortium (SBAC) mathematics and English language arts (ELA) score values, and a flag was retained to indicate whether each SBAC scale score was actual or imputed. Loosening the restrictions allowed for models to be trained using a larger sample and for the impact of including different variable combinations as predictors to be evaluated. All of the fields in this study correspond with information that CSDE stores in its secure data warehouse and mandates public school districts to report, including demographics, attendance, discipline, mobility, and achievement data.

Training and validation datasets. In order to obtain accurate forecasts, all models were developed using holdout sample validation, a process in which part of the sample is designated to model training,

and the remaining part of the sample is dedicated exclusively to model testing (also known as validation) (Cooil, Winer, & Rados, 1987; Mosier, 1951). The large sample size allowed for data splitting to obtain independent training and validation datasets. Stratified random sampling was used to partition the data and preserve the overall class distribution. Since more than 85 percent of students were in the *On-Track in 9th Grade* class, the rare class of off-track students was identified as the positive class. The training sample contained 80 percent of the records and was used to derive the models. The remaining 20 percent of records comprised the validation dataset that was used to evaluate the classification accuracy of the models.

Training datasets. Since the training sample was created by stratified random sampling to preserve the overall class distribution of the data in the overall sample, the training sample was imbalanced (i.e., 87.4 percent of records were on-track for the grade 9 model). This imbalanced training sample was used to create a training sample that was balanced via oversampling. The oversampled training set was created by merging all on-track records from the majority class with records resampled from the minority class with replacement; off-track records were resampled until the on-track and off-track classes were perfectly balanced.

Validation datasets. Balanced datasets were not used for testing. It is important to test models using an imbalanced dataset, since any future application of the models will make predictions using imbalanced data. For each sample, the validation dataset contained a random sample of 20 percent of the records from the imbalanced original sample. Validation samples were used to evaluate the classification accuracy of the models.

Analysis

A series of summary tables and data visualizations were used to analyze the results of various EIT models. In addition to summary tables presenting the classification accuracy measures, data visualizations were generated to help with model comparisons and to understand the relative importance of predictor variables. These visualizations included receiver operating characteristic (ROC) curves—including composite ROC curves with AUC values for all models, and individual ROC curves with optimal thresholds—and variable importance plots.

Generally, the model with the largest area under the ROC curve and highest balanced accuracy was selected as doing the best job. Since a premium was placed on correctly classifying off-track records

(i.e., records from the positive or minority class), precision (i.e., the proportion correct among all off-track predictions) was also among the most important accuracy measures when selecting between models.

RESULTS

Model results will be reported in two sections: (1) balanced accuracy; and (2) variable importance rankings. In all cases, models trained using the oversampled training sample performed best. Moreover, balanced accuracy was maximized at thresholds optimized using Youden's index.

Balanced Accuracy

Table 8 presents the balanced accuracy results for all EIT models. Balanced accuracy was lowest for grades 1 to 3. This result is not surprising, since these early-grade students are just starting their education journey. Moreover, those models do not include any standardized test scores as predictors (Kindergarten Entrance Inventory [KEI] scores are not as reliable as SBAC scores), there is nearly a one-to-one ratio of off-track to on-track records, and there is a limited number of predictor variables available.

Table 8
Percent On-Track in Test Dataset and Balanced Accuracy in Test Results

Grade	Outcome	Percent of records that were On-Track	Balanced Accuracy
1	Grade 3 SB ELA Proficient	54.8	0.727
2	Grade 3 SB ELA Proficient	54.5	0.726
3	Grade 3 SB ELA Proficient	53.9	0.736
4	Grade 6 SB ELA and Math Proficient	40.8	0.837
5	Grade 6 SB ELA and Math Proficient	40.4	0.863
6	Grade 6 SB ELA and Math Proficient	40.1	0.875
7	9th Grade On-Track	87.5	0.735
8	9th Grade On-Track	87.3	0.766
9	9th Grade On-Track	87.0	0.806
10	Grade 11 CCR (Ind 5 passing AND SAT Benchmark)	35.0	0.846
11	Grade 11 CCR (Ind 5 passing AND SAT Benchmark)	34.4	0.848
12	Grade 12 CCR (Ind 5 passing AND SAT Benchmark)	43.2	0.874

Note. Balanced accuracy is calculated as the average of the proportion corrects of the on-track and off-track classes. For example, a model that classified 81.8% of on-track students correctly and 85.6% of off-track students correctly in the test data has a balanced accuracy of 0.837 or 83.7%. The balanced accuracy calculation is not a weighted average.

Variable Importance Rankings

Figures 3 to 14 show the variable importance results for the random forests models. For grades 1 to 3, free-reduced lunch type (as determined by family income level), attendance, and special education status are the most important predictors of the grade 3 academic milestone. The academic components of the KEI—literacy, numeracy, and language—and whether the student was enrolled in an Alliance District (one of Connecticut’s lowest-performing districts) last year are also key predictors of meeting the grade 3 academic milestone.

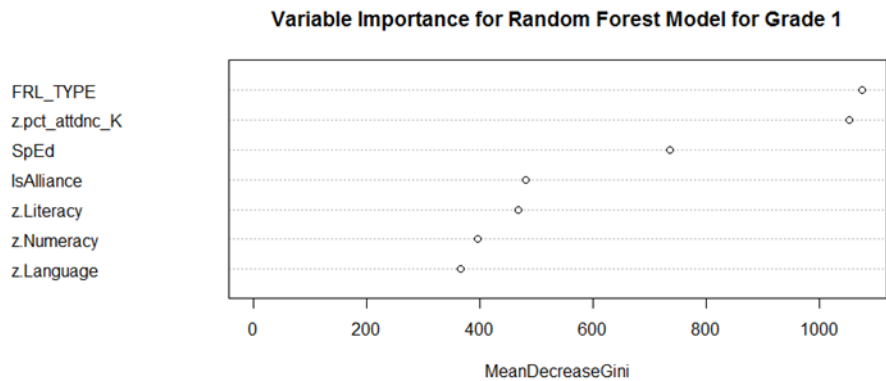


Figure 3. Variable importance for Grade 1 random forests model trained with oversampled training sample

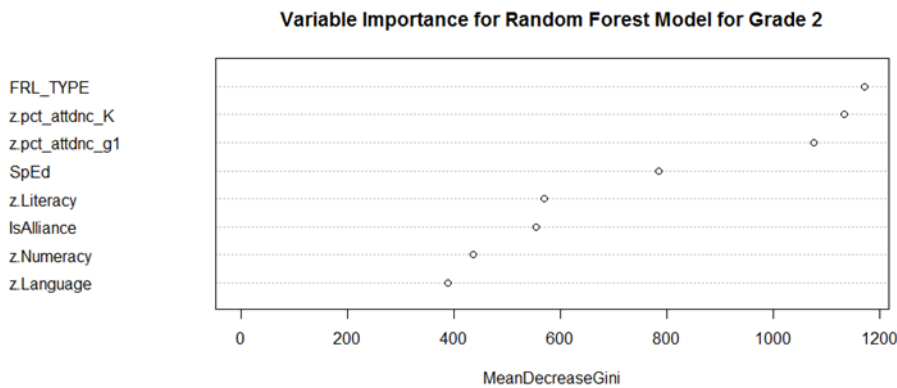


Figure 4. Variable importance for Grade 2 random forests model trained with oversampled training sample

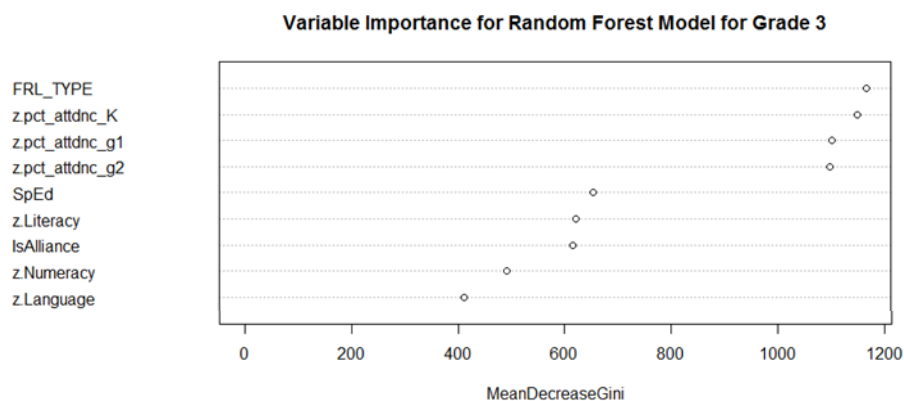


Figure 5. Variable importance for Grade 3 random forests model trained with oversampled training sample

Starting in grade 4, standardized test scores (SBAC ELA and math taken in grades 3 to 8, then PSATs taken in grade 10) have the highest variable importance in predicting whether students will reach their academic milestones. As in earlier models, free-reduced lunch type, attendance, and special education status also have high importance as predictors.

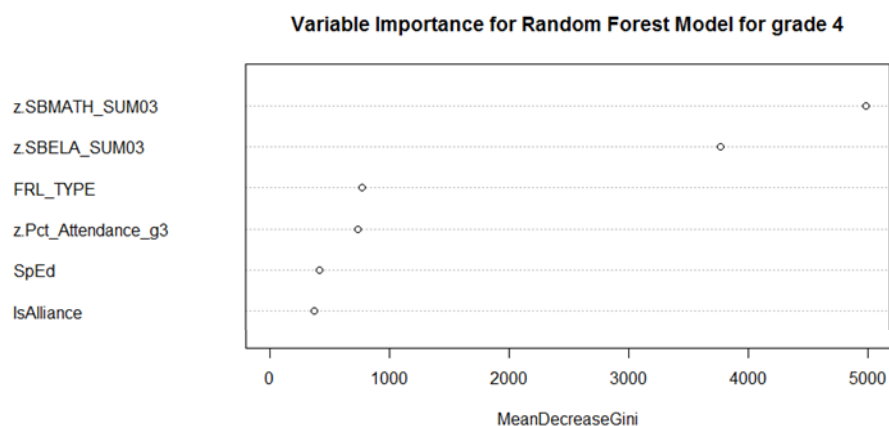


Figure 6. Variable importance for Grade 4 random forests model trained with oversampled training sample

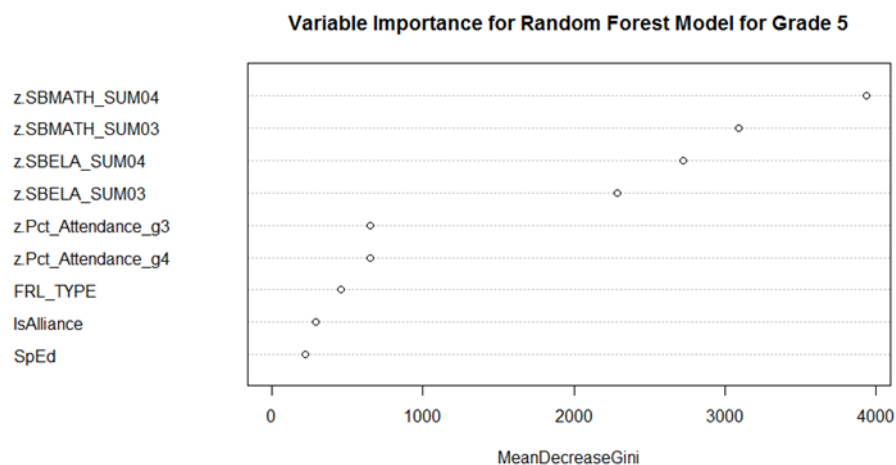


Figure 7. Variable importance for Grade 5 random forests model trained with oversampled training sample

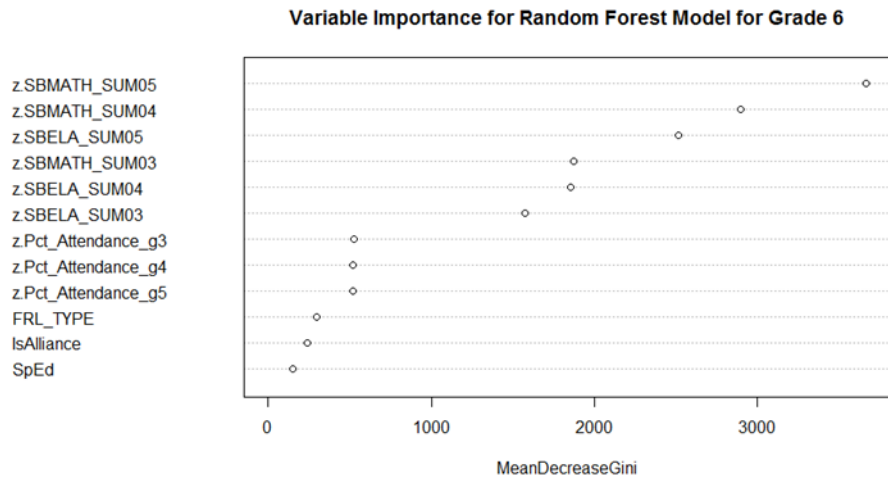


Figure 8. Variable importance for Grade 6 random forests model trained with oversampled training sample

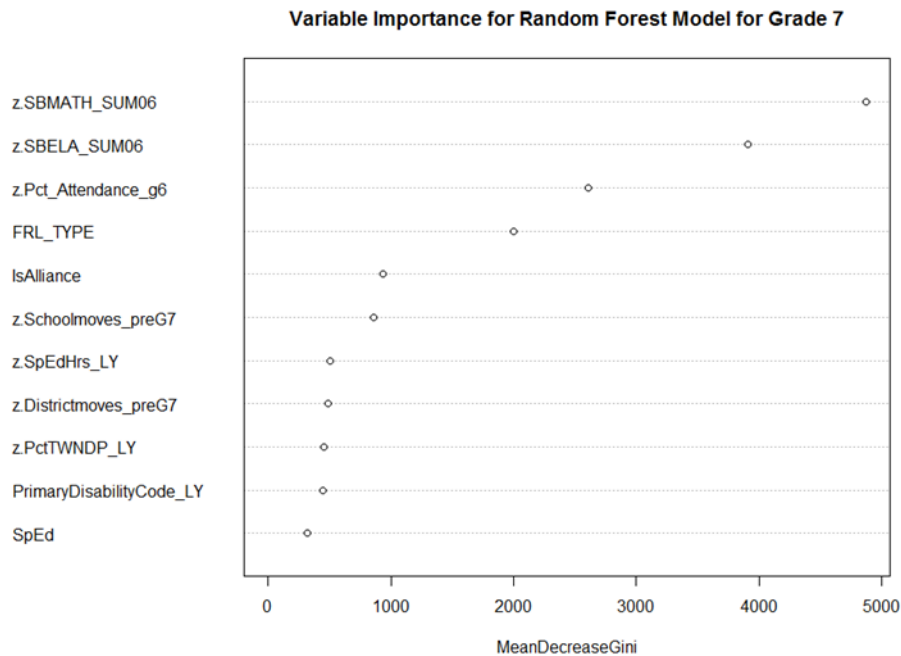


Figure 9. Variable importance for Grade 7 random forests model trained with oversampled training sample

Starting with the grade 8 model, counts for credits earned and failed are important predictors for meeting academic milestones.

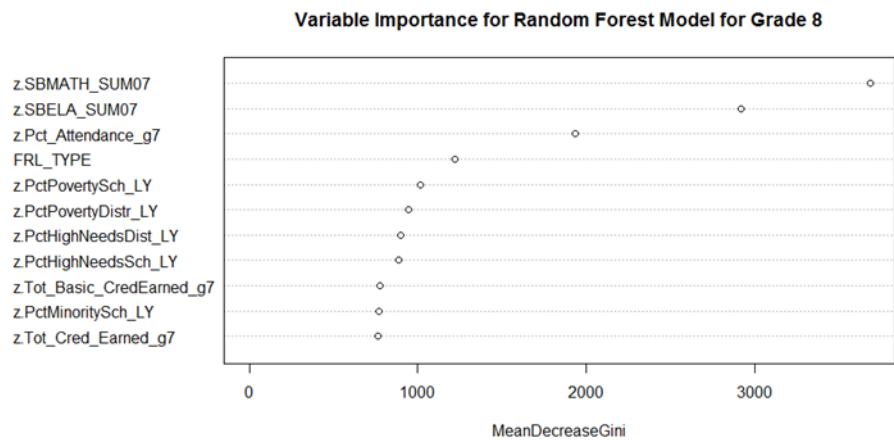


Figure 10. Variable importance for Grade 8 random forests model trained with oversampled training sample

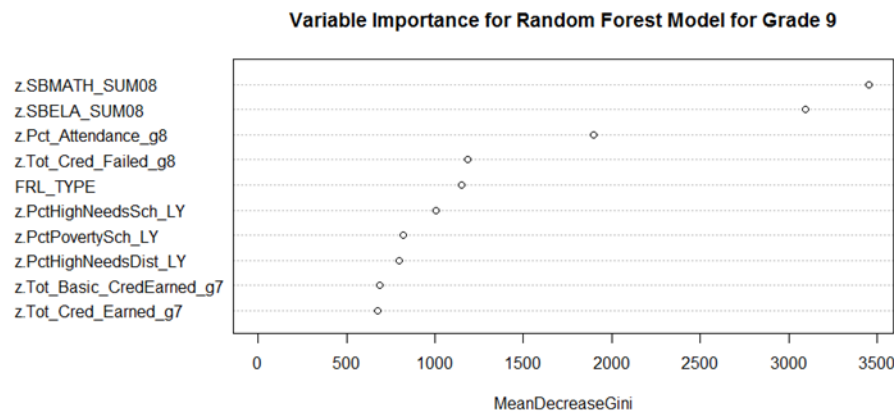


Figure 11. Variable importance for Grade 9 random forests model trained with oversampled training sample

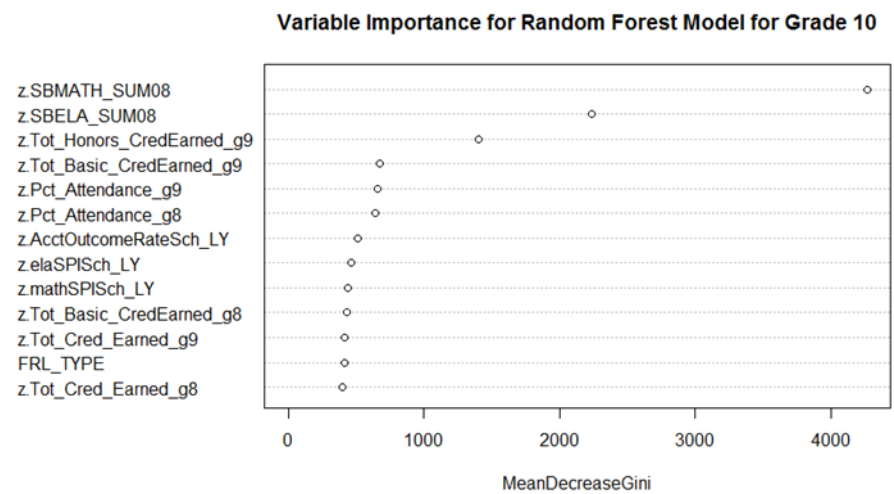


Figure 12. Variable importance for Grade 10 random forests model trained with oversampled training sample

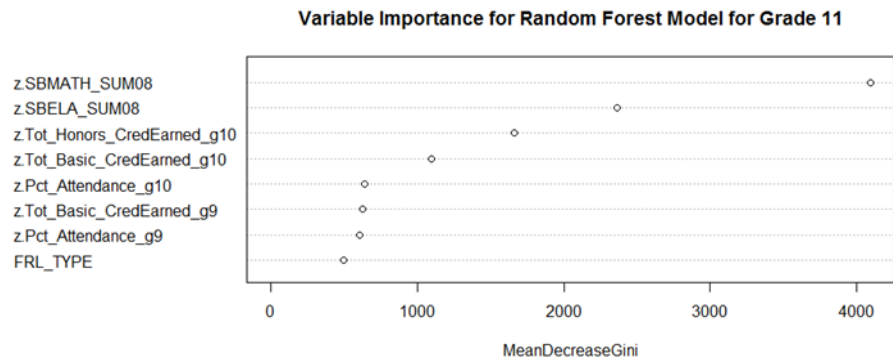


Figure 13. Variable importance for Grade 11 random forests model trained with oversampled training sample

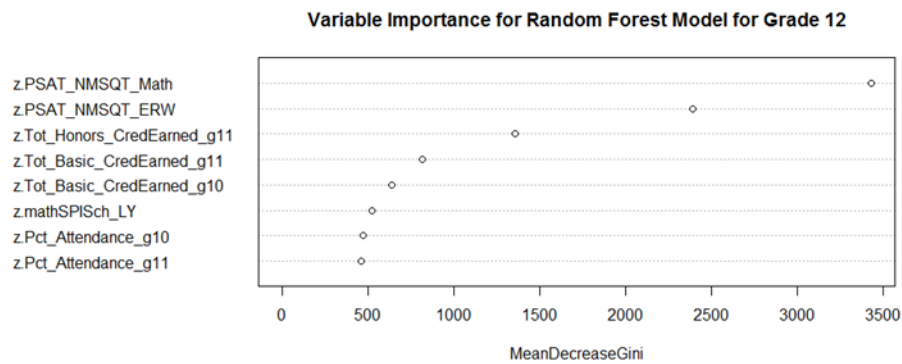


Figure 14. Variable importance for Grade 12 random forests model trained with oversampled training sample

DISCUSSION

The EIT aims to identify those students in need of targeted support and inform on-the-ground practitioners who may intervene long before students may be dropping out. In light of the accuracy comparisons—particularly the area under the ROC curve (AUC) and balanced accuracy for the best-fitting models—the random forests model developed using a training set that was balanced via oversampling does the best job of identifying which students are at risk for not being academic milestones. The random forests model had the largest AUC and, at optimal thresholds determined by Youden’s index, the highest balanced accuracy. Although the CART model can use surrogate variables when missing values are encountered (Breiman et al., 1984), it is very dependent on training data and prone to overfitting. The EIT models use powerful imputation techniques to address missing data concerns, and the random forests model was superior to CART and all other models in terms of performance. Furthermore, the random forests model decorrelates decision trees, handles large numbers of variables and complex interactions, and is not constrained by assumptions. In light of these advantages and its superior performance, the random forests methodology does the best job of identifying which students are at risk for not being on-track to graduate.

The ranked lists of variable importance showed that Smarter Balanced Assessment Consortium (SBAC) mathematics and English language arts (ELA) scale scores had the highest variable importance in predicting the EIT’s academic milestones for grades 4 through 11. Attendance, course performance, and free-and-reduced lunch type occupied the next tier of variables in terms of importance. These findings are consistent with other studies that examined the on-track indicator and graduation outcomes (e.g., Allensworth, 2013; Knowles, 2015; Mac Iver & Messel, 2013).

Implications of Results

The comparison of test results during EIT model development demonstrated that preprocessing data—oversampling and imputing missing standardized test scores, in particular—and optimizing thresholds using Youden’s index improves classification accuracy when identifying which students need support. In addition, there is value in including special education-related predictors to improve classification accuracy for students with disabilities. Course-, school- and district-level variables all added to the classification accuracy of high school prediction models. Lastly, SBAC and PSAT scores are highly predictive of all subsequent academic milestones.

As of August 31, 2018, Connecticut was one of 13 SBAC member states, along with California, Delaware, Hawaii, Idaho, Michigan, Montana, Nevada, North Carolina, Oregon, South Dakota, Vermont, and Washington (SBAC, n.d.). Educational leaders from SBAC member states could look to conduct similar studies and analyze whether their results converge with the current study’s findings with respect to classification accuracy and variable importance. More generally, they can use this study’s findings to explore ways to improve EWSs and increase rates of graduation and college and career readiness in their respective states.

Since the random forests models calculated probabilities of being on-track and performed well at identifying at-risk students, CSDE uses these probabilities to classify students into high, medium, and low support levels and present this information through secure dashboards similar to the Massachusetts Early Warning Indicator System (EWIS) (Massachusetts Department of Elementary and Secondary Education & AIR, 2013a, 2013b, 2014).

Classification into high, medium, and low support levels can also help to illustrate the relationship between variables and classification levels. More important, it can highlight the importance of going beyond single-variable, single-threshold early warning systems (e.g., systems that focus on only one indicator, such as all students with attendance below 90%), which overlook complex interactions among predictors. For example, Figures 15 and 16 show two different “Support levels by predictor” views. Figure 15 shows a mosaic plot of support levels by chronic absentee status (i.e., those students who missed more than 10% of school days).

Support Levels by Chronic Absentee Status

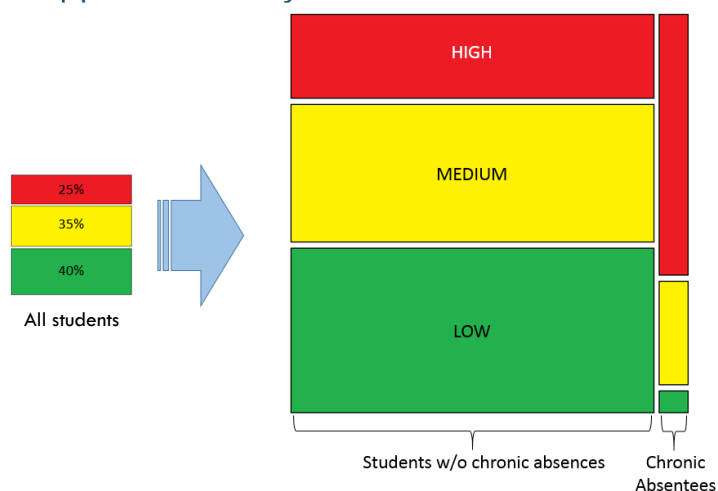


Figure 15. Mosaic plot of support levels by chronic absentee status

The smaller square to the left of the arrow shows the overall distribution of support levels for all students entering 8th grade. The support levels were assigned based on each student's probability of being on-track by the end of 9th grade: The 25 percent of students with the lowest probabilities were assigned to the High Support group, the next 35 percent of students were classified as Medium Support, and the 40 percent of students with the highest probabilities were assigned to the Low Support group. To the right of the arrow, the mosaic plots show that students who are chronically absent are more likely to be in the High or Medium Support groups. However, there are chronic absentees with high probabilities of being on-track (i.e., Low Support) and there are students without chronic absences who have a low probability of being on-track (i.e., High Support). Clearly, one variable is not sufficient to predict a student's probability of being on-track.

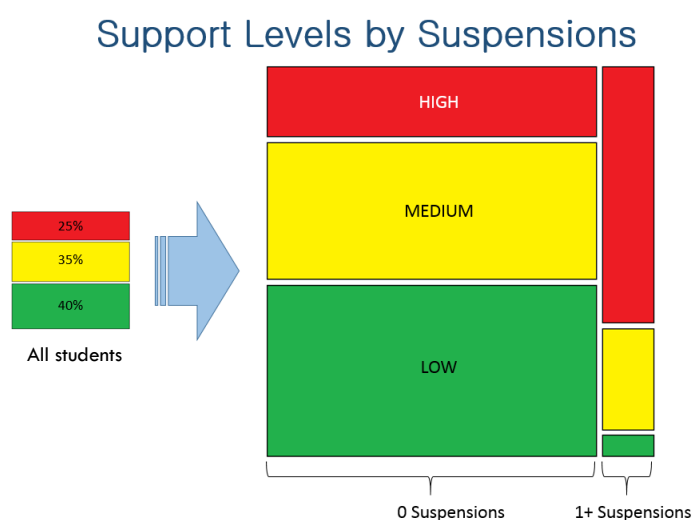


Figure 16. Mosaic plot of support levels by suspensions

Figure 16 shows a mosaic plot of support levels by suspensions. Again, the smaller square to the left of the arrow shows the overall distribution of support levels for all students entering 8th grade. When we look to the right of the arrow, the first thing to notice is that the *1+ Suspensions* section is wider than the *Chronic Absentees* section in Figure 15, which shows that there were more suspended students than chronic absentees in Grade 7. Again, the probabilities from the random forests model did not result in all students who met a particular single-variable threshold (in this case, having one or more suspensions) being placed in the same support level. By considering a student's entire profile—including attendance, behavior, course performance, credit accrual, mobility, detailed special education data, standardized assessment scores, English learner status and family income status—policymakers will have a more complete picture from which to make more informed decisions regarding the timing, type, and target of interventions to implement.

Conclusion

The EIT modeling methods and extensive predictors go beyond the logistic regression methods and binary flag predictors (i.e., 0-1 fields that indicate broad membership in student groups such as students with disabilities) that are pervasive in educational early warning systems. The methods, variable importance and accuracy measures presented herein provide practitioners the opportunity to leverage new knowledge about students who are at-risk and to test interventions at many levels in an attempt to improve graduation outcomes.

REFERENCES

- Allensworth, E. M. (2013). The use of ninth-grade early warning indicators to improve Chicago schools. *Journal of Education for Students Placed at Risk (JESPAR)*, 18(1), 68–83.
- Allensworth, E. M., & Easton, J. Q. (2005). *The on-track indicator as a predictor of high school graduation*. Chicago, IL: Consortium on Chicago School Research, University of Chicago.
- 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*. Chicago, IL: Consortium on Chicago School Research, University of Chicago.
- American Institutes for Research [AIR]. (n.d.). *The National High School Center*. Retrieved from <http://www.air.org/project/national-high-school-center>
- Balfanz, R. (2009). Putting middle grades students on the graduation path. *Policy and practice brief*.
- Balfanz, R. (2011). Back on track to graduate. *Educational Leadership*, 68(7), 54–58.
- Balfanz, R. (2014, June 8). Stop holding us back. *New York Times (1923-Current File)*, SR5.
- Balfanz, R. (2016). Missing school matters. *Phi Delta Kappan*, 98(2), 8–13.
- Balfanz, R., & Byrnes, V. (2012). *The importance of being in school: A report on absenteeism in the nation's public schools*. Baltimore, MD: Johns Hopkins University Center for Social Organization of Schools.
- 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*.
- Balfanz, R., Byrnes, V., & Fox, J. H. (2014). Sent home and put off-track: The antecedents, disproportionalities, and consequences of being suspended in the ninth grade. *Journal of Applied Research on Children*, 5(2), 17–30.
- Battistini, A., Rosi, A., Segoni, S., Lagomarsino, D., Catani, F., & Casagli, N. (2017). Validation of landslide hazard models using a semantic engine on online news. *Applied Geography*, 82, 59–65. <https://doi.org/10.1016/j.apgeog.2017.03.003>
- Belfield, C. R. (2007). The promise of early childhood education interventions. In C. R. Belfield & H. M. Levin (Eds.), *The price we pay: Economic and social consequences for inadequate education*. (pp. 200–224). Washington, DC: The Brookings Institution.
- Belfield, C. R., & Levin, H. M. (2007). The education attainment gap: Who's affected, how much, and why it matters. In C. R. Belfield & H. M. Levin (Eds.), *The price we pay: Economic and social consequences for inadequate education*. (pp. 1–17). Washington, DC: The Brookings Institution.

-
- Berkowitz, M. R. (2013). High-level specification of a proposed information architecture for support of a bioterrorism early-warning system. *Southern Medical Journal*, 106(1), 31–36.
<https://doi.org/10.1097/SMJ.0b013e31827ca83c>
- 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. *High School Journal*, 96(2), 77–100.
- Boyd, B. A. (2016). *Early-warning indicators of high school dropout* (Doctoral dissertation). Retrieved from ProQuest LLC.
- Breiman, L. (2001a). Random forests. *Machine Learning*, 45, 5–32.
- Breiman, L. (2001b). Statistical modeling: Two cultures (with discussion). *Statistical Science*, 16, 199–231.
- Breiman, L., Cutler, A., Liaw, A., & Wiener, M. *Package ‘randomForest’: Breiman and Cutler’s random forests for classification and regression*. R package version 4.6-14. Retrieved from <https://cran.r-project.org/web/packages/randomForest/randomForest.pdf>
- Bridgeland, J. M., DiIulio, J. J., Jr., & Morison, K. B. (2006). The silent epidemic: Perspectives of high school dropouts. *Civic Enterprises*.
- Chen, C., Liaw, A., & Breiman, L. (2004). Using random forest to learn imbalanced data. *University of California, Berkeley*, 110, 1–12.
- Civic Impulse. (2017). S. 1177 — 114th Congress: Every Student Succeeds Act. Retrieved from <https://www.govtrack.us/congress/bills/114/s1177>
- Cizek, G. J., & Fitzgerald, S. M. (1999). Methods, plainly speaking: An introduction to logistic regression. *Measurement and Evaluation in Counseling and Development*, 31(4), 223–245.
- Connecticut General Assembly. (2017). *Sec. 10-221a. High school graduation requirements. Student support and remedial services. Excusal from physical education requirement. Diplomas for certain veterans and certain persons assisting in the war effort during World War II. Student success plans*. Retrieved from https://www.cga.ct.gov/current/pub/chap_170.htm#sec_10-221a
- Connecticut State Department of Education [CSDE]. (2016). *ESEA Flexibility Renewal: Connecticut’s “Next Generation” Accountability System*. Retrieved from http://edsight.ct.gov/relatedreports/next_generation_accountability_system_march_2016.pdf
- Connecticut State Department of Education [CSDE]. (2017). *Draft: 5-Year NGSS Implementation Timeline*. Retrieved from https://portal.ct.gov/-/media/SDE/Science/NGSS_5_Year_Implementation_Plan_for_Transitioning.pdf
- Connecticut State Department of Education [CSDE]. (2018). *Using accountability results to guide improvement*. (3rd ed.). Retrieved from http://edsight.ct.gov/relatedreports/using_accountability_results_to_guide_improvement.pdf
- Conway, D., & White, J. (2012). *Machine learning for hackers*. Sebastopol, CA: O’Reilly Media, Inc.

-
- Cooil, B., Winer, R., & Rados, D. (1987). Cross-validation for prediction. *Journal of Marketing Research*, 24(3), 271–279.
- Data Quality Campaign. (2013). *Data for Action 2013*. Data Quality Campaign.
- DePaoli, J. L., Balfanz, R., & Bridgeland, J. (2016). *Building a Grad Nation: Progress and challenge in raising high school graduation rates. Annual Update 2016*. Civic Enterprises.
- DePaoli, J. L., Balfanz, R., Atwell, M., & Bridgeland, J. (2018). *Building a Grad Nation: Progress and challenge in raising high school graduation rates. Annual Update 2018*. Civic Enterprises.
- Dupéré, V., Dion, E., Leventhal, T., Archambault, I., Crosnoe, R., & Janosz, M. (2018). High School Dropout in Proximal Context: The Triggering Role of Stressful Life Events. *Child Development*, 89(2), 107–122.
- Ekstrom, R. B., Goertz, M. E., Pollack, J., & Rock, D. A. (1986). Who drops out of high school and why? Findings from a national study. In G. Natriello (Ed), *School dropouts: Patterns and policies*. (pp. 52–69). New York, NY: Teachers College Press.
- Every Student Succeeds Act (ESSA) of 2015, Pub. L. No. 114-95 § 114 Stat. 1177 (2015).
- Ferguson, R. F. (2007). Toward excellence with equity: The role of parenting and transformative school reform. In C. R. Belfield & H. M. Levin (Eds.), *The price we pay: Economic and social consequences for inadequate education*. (pp. 225–254). Washington, DC: The Brookings Institution.
- Fraley, C., Raftery, A.E., Murphy, T.B., & Scrucca, L. (2012). *mclust version 4 for R: Normal mixture modeling for model-based clustering, classification, and density estimation*. Retrieved from <http://www.stat.washington.edu/mclust/>
- Freeman, E. A., & Moisen, G. G. (2008). A comparison of the performance of threshold criteria for binary classification in terms of predicted prevalence and kappa. *Ecological Modelling*, 217(1), 48–58.
- Freeman, J., Simonsen, B., McCoach, D., Sugai, G., Lombardi, A., & Horner, R. (2015). An analysis of the relationship between implementation of school-wide positive behavior interventions and supports and high school dropout rates. *The High School Journal*, 98(4), 290–315.
- Gwynne, J., Lesnick, J., Hart, H. M., & Allensworth, E. M. (2009). *What matters for staying on-track and graduating in Chicago Public Schools: A focus on students with disabilities. Research report*. Chicago, IL: Consortium on Chicago School Research, University of Chicago.
- Gwynne, J., Pareja, A. S., Ehrlich, S. B., & Allensworth, E. (2012). *What matters for staying on-track and graduating in Chicago Public Schools: A focus on English language learners. Research report*. Chicago, IL: Consortium on Chicago School Research, University of Chicago.
- Haixiang, G., Yijing, L., Shang, J., Mingyun, G., Yuanyue, H. & Bing, G. (2017). Learning from class-imbalanced data: Review of methods and applications. *Expert Systems With Applications*, 73(C), 220-239.

-
- Hanover Research. (2018). *Trends in K-12 education: 2018. Research report*. Arlington, VA: Hanover Research.
- Hosmer, D. W., Jr., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression*. Hoboken, NJ: John Wiley & Sons.
- 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.
- Kuhn, M. (2018). *caret: Classification and regression training*. R package version 6.0-80. Retrieved from <https://www.rdocumentation.org/packages/caret/versions/6.0-80>
- Kuhn, M., & Johnson, K. (2013) *Applied predictive modeling*. New York: Springer.
- Kuhn, M., & Johnson, K. (2014). Who’s afraid of the big black box?: Statisticians’ vital role in big data and predictive modelling. *Significance*, 11(3), 35–37. <https://doi.org/10.1111/j.1740-9713.2014.00753.x>
- Levin, H. M., & Belfield, C. R. (2007). Educational interventions to raise high school graduation rates. In C. R. Belfield & H. M. Levin (Eds.), *The price we pay: Economic and social consequences for inadequate education*. (pp. 1–17). Washington, DC: The Brookings Institution.
- Mac Iver, M. A., & Messel, M. (2013). The ABCs of keeping on track to graduation: Research findings from Baltimore. *Journal of Education for Students Placed at Risk (JESPAR)*, 18(1), 50–67.
- Malin, J. R., Bragg, D. D., & Hackmann, D. G. (2017). College and career readiness and the Every Student Succeeds Act. *Educational Administration Quarterly*, 53(5), 809–838.
- Massachusetts Department of Elementary and Secondary Education, & AIR. (2013a). *Massachusetts Early Warning Indicator System (EWIS). “Technical descriptions of risk model development”: Early and late elementary age groupings (Grades 1-6)*. Massachusetts Department of Elementary and Secondary Education.
- Massachusetts Department of Elementary and Secondary Education, & AIR. (2013b). *Massachusetts Early Warning Indicator System (EWIS). “Technical descriptions of risk model development”: Middle and high school age groupings (Grades 7-12)*. Massachusetts Department of Elementary and Secondary Education.
- Massachusetts Department of Elementary and Secondary Education, & AIR. (2014). *Early warning implementation guide: “Using the Massachusetts Early Warning Indicator System (EWIS) and local data to identify, diagnose, support, and monitor students in grades 1-12.”* Massachusetts Department of Elementary and Secondary Education.
- Mosier, C. I. (1951). Problems and designs of cross-validation. *Educational and Psychological Measurement*, 11(1), 5–11.
- Ng, A. (2013). *Machine Learning*. Stanford University. Retrieved from <http://online.stanford.edu/course/machine-learning>
- R Core Team. (2018). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>

-
- Rebell, M. A. (2007). The need for comprehensive educational equity. In C. R. Belfield & H. M. Levin (Eds.), *The price we pay: Economic and social consequences for inadequate education*. (pp. 225–254). Washington, DC: The Brookings Institution.
- RStudio Team (2016). *RStudio: Integrated development for R*. RStudio, Inc., Boston, MA. URL <http://www.rstudio.com/>.
- Rumberger, R. W. (2011). *Dropping out: Why students drop out of high school and what can be done about it*. Cambridge, MA: Harvard University Press.
- Rumberger, R. W., & Larson, K. A. (1998). Student mobility and the increased risk of high school dropout. *American Journal of Education*, 107(1), 1–35.
- Rumberger, R., & Lim, S. A. (2008). *Why students drop out of school: A review of 25 years of research*. Santa Barbara, CA: California Dropout Research Project.
- Rumberger, R. W., & Palardy, G. J. (2005). Test scores, dropout rates, and transfer rates as alternative indicators of high school performance. *American Educational Research Journal*, 42(1), 3–42.
- Rumberger, R. W., & Thomas, S. L. (2000). The distribution of dropout and turnover rates among urban and suburban high schools. *Sociology of Education*, 73(1), 39–67.
- SAS Institute Inc. (2018). SAS Enterprise Guide Version 7.15. Cary, NC: SAS Institute Inc.
- SAS Institute Inc. (2018). SAS Data Integration Studio Version 4.903. Cary, NC: SAS Institute Inc.
- Scala, J. (2015). *Early warning systems*. Rural Dropout Prevention Project, American Institutes for Research (AIR). Retrieved from <https://www.nd.gov/dpi/uploads/1331/EWSFinal21115.pptx>
- Smarter Balanced Assessment Consortium [SBAC]. (n.d.). *Members and Governance*. Retrieved August 31, 2018 from <http://www.smarterbalanced.org/about/members/>
- Strobl, C. (2013). Data mining. In T. D. Little (Ed), *The Oxford handbook of quantitative methods: Statistical analysis, Vol. 2*. (pp. 678–700). New York, NY, US: Oxford University Press.
- Swets, J. A. (2014). *Signal detection theory and ROC analysis in psychology and diagnostics: Collected papers*. New York, NY: Psychology Press.
- U.S. Department of Education. (2017). *Connecticut Consolidated State Plan Under the Every Student Succeeds Act*. Retrieved from http://www.sde.ct.gov/sde/lib/sde/pdf/essa/august_4_ct_consolidated_state_essa_plan.pdf
- van Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software*, 45(3), 1–67. URL <https://www.jstatsoft.org/v45/i03/>.
- Veltri, G. A. (2017). Big Data is not only about data: The two cultures of modelling. *Big Data*, 4(1), 1–6.
- Wilkins, J., & Bost, L. W. (2016). Dropout prevention in middle and high schools. *Intervention in School & Clinic*, 51(5), 267–275. <https://doi.org/10.1177/1053451215606697>

-
- Woods, C. M. (2013). Categorical methods. In T. D. Little (Ed), *The Oxford handbook of quantitative methods: Statistical analysis, Vol. 2.* (pp. 52–73). New York, NY: Oxford University Press.
- Youden, W. J. (1950). Index for rating diagnostic tests. *Cancer*, 3(1), 32–35.

APPENDIX

Table 9

Data definitions for fields used to predict academic milestones in EIT models

Field	Description
isOnTime	A student who is on time for grade progression (isOnTime = 1) is one who is < 15 years old on entering 9 th grade (< 14 years old on entering 8 th grade, etc.); otherwise isOnTime = 0.
isHighNeeds	A student with High Needs status (isHighNeeds = 1) is one who is economically-disadvantaged, an English learner (EL), or a student with disabilities (SWD); otherwise isHighNeeds = 0
LunchCode	F for free; R for reduced-price; N for non-subsidized
isSpEd	A student with Special Education status (isSpEd = 1) is one with an identified disability who needs specially designed instruction to meet his/her unique needs and to enable the child to access the general curriculum of the school district; otherwise isSpEd = 0
PrimaryDisabilityCode	Primary disability code; this field contains the code corresponding to the student's primary disability only for those students for whom isSpEd = 1, otherwise it defaults to 99
TWNDP_pct_g<G>^a	Percentage of Time with Non-Disabled Peers (Hours inside regular education classroom ÷ hours in school day) x 100
isEL	A student with English Learner (EL) status (isEL = 1) is one who meets at least one of the following criteria: <ul style="list-style-type: none"> English is not the primary language spoken in the home, regardless of the language spoken by the student; English is not the language most often spoken by the student; English is not the language the student first acquired; AND whose English language proficiency test results met English Learner requirements; otherwise isEL = 0
DistrictEnrollment_g<G>	Number of students enrolled in student's public school district
DPI_ELA_g<G>	ELA District Performance Index (DPI) of student's public school district
DPI_Math_g<G>	Math DPI of student's public school district
DPI_Science_g<G>	Science DPI of student's public school district
DistrictNextGenPct_g<G>	Next Generation Accountability Percentage of Points for student's school district
DistrictPctMinority_g<G>	Percentage of minority students in student's school district

Table 9: Continued

Field	Description
DistrictPctHighNeeds_g<G>	Percentage of high needs students in student's school district
DistrictPctPoverty_g<G>	Percentage of students in poverty in student's school district
DistrictChronAbs_g<G>	Chronic Absenteeism rate for student's school district
SchoolEnrollment_g<G>	Number of students enrolled in student's public school
SPI_ELA_g<G>	ELA School Performance Index (SPI) of student's public school
SPI_Math_g<G>	Math SPI of student's public school
SPI_Science_g<G>	Science SPI of student's public school
SchoolNextGenPct_g<G>	Next Generation Accountability Percentage of Points for student's school
SchoolPctMinority_g<G>	Percentage of minority students in student's school
SchoolPctHighNeeds_g<G>	Percentage of high needs students in student's school
SchoolPctPoverty_g<G>	Percentage of students in poverty in student's school
SchoolChronAbs_g<G>	Chronic Absenteeism rate for student's school
Fac1AttendanceDays_g<G>	Number of school days the student attended
Fac1MembershipDays_g<G>	Total days student's school was in session during Grade 5
Pct_Attendance_g<G>	Percentage of school days attended (Fac1AttendanceDays_g5 ÷ Fac1MembershipDays_g5) * 100
Repeat_g<G>	Flag indicating whether the student repeated grade, with a value of 1 for Yes and 0 for No
ISS_g<G>	Student's in-school suspension incidents in grade <G>
OSS_g<G>	Student's out-of-school suspension incidents in grade <G>
SchoolMoves_g<G>	Number of times the student changed schools during grade <i>NOTE: School moves does not include in-district promotions such as elementary school to middle school.</i>
SBELA_g<G> ^b	Connecticut Smarter Balanced English language arts (ELA) scale score (reported for grades 3 to 8)
SBMath_g<G>	Connecticut Smarter Balanced mathematics scale score (reported for grades 3 to 8)
Tot_Cred_Earned_g<G>	Total credits earned (reported in grades 7 to 11)
Tot_Cred_Failed_g<G>	Total credits failed (reported in grades 7 to 11)
Tot_Core_Earned_g<G>	Total core (i.e., ELA, math, social sciences and history, and science) credits earned (reported in grades 7 to 11)
Tot_Core_Failed_g<G>	Total core credits failed (reported in grades 7 to 11)
isCoreCredFailed1_g<G>	Flag indicating if total core credits failed >= 1; 1 if yes, 0 otherwise (reported in grades 7 to 11)
isOnTrack_g9 ^c	Flag indicating if student is On-Track at the end of Grade 9; 1 if isCoreCredFailed1_g9 = 0 and Tot_Cred_Earned_g9 >= 5; 0 otherwise (reported in grade 9)
Tot_EnrAdv_Cred_Earned_g<G>	Total credits earned in enriched or advanced courses (reported in grades 7 to 11)
Tot_Honors_Earned_g<G>	Total credits earned in honors courses (reported in grades 7 to 11)

Table 9: Continued

Field	Description
Tot_Basic_Cred_Earned_g<G>	Total credits earned in basic or remedial courses (reported in grades 7 to 11)
Tot_Basic_Cred_Failed_g<G>	Total credits failed in basic or remedial courses (reported in grades 7 to 11)
Tot_English_Cred_Earned_g<G>	Total English credits earned (reported in grades 7 to 11)
Tot_English_Cred_Failed_g<G>	Total English credits failed (reported in grades 7 to 11)
Tot_Math_Cred_Earned_g<G>	Total Math credits earned (reported in grades 7 to 11)
Tot_Math_Cred_Failed_g<G>	Total Math credits failed (reported in grades 7 to 11)
Tot_Science_Cred_Earned_g<G>	Total Science credits earned (reported in grades 7 to 11)
Tot_Science_Cred_Failed_g<G>	Total Science credits failed (reported in grades 7 to 11)

^a <G> is used as a placeholder when a variable is reported in more than one grade level, as noted in the description.

^b Starting in 2015, the SBAC was the standard assessment administered in Connecticut grades 3 through 8.

^c A student is on-track in 9th grade if s/he earns at least five full-year credits in the year *and* no more than one failing grade in English, mathematics, science, or social studies (CSDE, 2016).

Inquiries regarding the Early Indication Tool should be directed to the developers of the EIT and authors of this report:

David M. Alexandro, Ph.D. (david.alexandro@ct.gov)

Charles W. Martie, Ph.D. (charles.martie@ct.gov)

Connecticut State Department of Education

Performance Office

The Connecticut State Department of Education is committed to a policy of equal opportunity/affirmative action for all qualified persons. The Connecticut State Department of Education does not discriminate in any employment practice, education program, or educational activity on the basis of **race, color, religious creed, sex, age, national origin, ancestry, marital status, sexual orientation, gender identity or expression, disability (including, but not limited to, intellectual disability, past or present history of mental disorder, physical disability or learning disability), genetic information, or any other basis prohibited by Connecticut state and/or federal nondiscrimination laws. The Connecticut State Department of Education does not unlawfully discriminate in employment and licensing against qualified persons with a prior criminal conviction.** Inquiries regarding the Connecticut State Department of Education's nondiscrimination policies should be directed to:

Levy Gillespie

Equal Employment Opportunity Director/American with Disabilities Act Coordinator

Connecticut State Department of Education

450 Columbus Boulevard, Suite 607

Hartford, CT 06103-1841

860-807-2071

Levy.Gillespie@ct.gov