education analytics

// March 17, 2020

PRODUCING LONG-TERM FORECASTS OF INDIVIDUAL STUDENT OUTCOMES: AN APPLICATION OF CHAIN-LINKED PREDICTIVE ANALYTICS MODELS WITH SHORT-SPAN DATA AND EDUCATION POLICY REGIME CHANGE

Jeff Dominitz, Soobin Kim, Robert H. Meyer, and Andrew Rice

PRELIMINARY AND INCOMPLETE DRAFT: NOT FOR CIRCULATION. PLEASE DO NOT QUOTE OR CITE

1. INTRODUCTION

Parents, students, teachers, and counselors routinely monitor student outcomes to assess progress toward educational goals. Although many goals may be short-term (e.g., promotion to the next grade, proficiency on annual tests, or improved grades), each of these stakeholders is also concerned with students' long-term goals and outcomes. We focus in this paper on using predictive analytics models to produce forecasts of high school graduation and college persistence outcomes for students as early as 3rd grade. As these forecasts are updated over time with new information on student attendance, standardized test scores, coursework, and other intermediate outcomes, progress toward long-term goals may be tracked and obstacles identified as they arise.

Our approach can be compared to efforts to implement "early warning systems" to identify students at risk of dropping out (e.g., McMahon & Sembiante, 2019). Many early warning systems typically identify students as "at-risk" based on whether they meet certain criteria included in a checklist (e.g., chronic absenteeism, suspension, or insufficient credit accumulation) that were previously identified as predictive of dropping out (e.g., Allensworth & Easton, 2007; Bruce, Bridgeland, Fox, & Balfanz, 2011). This approach has several limitations, including that it is not straightforward to (1) implement in early grades, (2) update risk assessments in response to changes in policies, and (3) compute the accuracy of at-risk classifications, a necessary factor in determining the efficacy of interventions.

In contrast, this paper explores statistical methods using individual students' historical data to forecast their outcomes. Ideally, we would use longitudinal data on a panel of students followed from elementary school through high school and onto college, as in Bowers (2010). However, researchers and practitioners may face two key challenges with this approach: (1) the historical data may not go far enough back in time to track students all the way from third grade through college, and (2) education policy may have systematically shifted in the interim such that extrapolating from the distant past to the future would be ill advised. Indeed, we *hope* that the use of early warning data to target resources to at-risk students will change the relationships between predictors and outcomes, for example, by reducing the predictive power of student characteristics such as poverty.

We develop and apply chain-linked predictive analytics models that account for both types of challenges, which we refer to as short-span data and education policy regime change, respectively. Chain-linking uses short-span models to make predictions that are linked across short spans of time. For instance, the models presented in this paper link predictions from 3rd grade to 6th grade, 6th grade to 8th grade, 8th to 10th, 10th to 12th, 12th to the second year of college. Alternatively, one could link consecutive grades: 3rd to 4th, 4th to 5th, 5th to 6th, and so on. The decision as to which spans to model depends not only on the time span of the data but also beliefs about how recent policy changes impact the credibility of extrapolating from observed statistical relationships to future outcomes.

We estimate the chain-linked models using four years of panel data from the CORE Data Collaborative—a consortium of more than 120 California school districts serving more than two million students, equivalent to more than 1/3 of the students in all of California. With just four years of data, we cannot directly estimate prediction models for more than a three-year forecast horizon. In addition, we were particularly sensitive to systemwide changes in enrollment patterns in 8th grade math; with the adoption of Common Core standards, there was a steep decline in 8th grade algebra enrollment just prior to our sample period (McEachin et al., 2020). This type of policy regime change affects the statistical relationships among observed mathematics coursework in 8th grade and in high school and subsequent graduation and college persistence outcomes. Thus, we sought to estimate these relationships using only more recent data on high school students who went through 8th grade in the "new regime."

The truest test of the predictive value of our chain-linked models will come over the next several years of implementation, as we make out-of-sample forecasts for new cohorts of students and then observe their graduation and college outcomes. In the interim, we have assessed the predictive value of our models by comparing our chain-linked predictions to more conventional types of at-risk indicators developed to designate 8th and 9th graders as meeting the criteria for one of five discrete "on-track" categories. A priori, a quantitative predictive analytic model should outperform any such qualitative indicators in predicting graduation outcomes. However, these qualitative indicators successfully utilized supplemental information to perform surprisingly well. After presenting our empirical results, we discuss how this comparison influenced some of the functional-form choices we made in subsequent iterations of our predictive models.

2. RELEVANT LITERATURE

2.1. Early Warning Systems

Twenty years ago, the Consortium on Chicago School Research developed "an indicator to gauge whether students make sufficient progress in their first year of high school to be on track to graduate within four years" (Allensworth & Easton, 2005, p. 1). This on-track indicator is notable for its simplicity and predictiveness. Students labeled as "on-track" in Chicago Public Schools—those who earned five full-course credits and failed no more than one semester-length course in ninth grade—had a four-year graduation rate of 81% versus 22% for off-track students (Allensworth & Easton, 2005).

The predictiveness is impressive, but Allensworth and Easton (2007) subsequently note that a timelier indicator would provide information prior to the summer after ninth grade.

Moreover, the indicator "does not provide information precise enough to allow specific students to be targeted for specific interventions" (p. 6).

Indeed, researchers have called for EWS that provide actionable information to trigger interventions and supports. For example, Rumberger et al. (2017) produced a What Works Clearinghouse guide to EWS for dropout prevention calling for EWS to signal when schools should "proactively intervene" in response to signs of problems and "provide intensive, individualized support" to off-track students (p. 2). However, McMahon and Sembiante (2019) find that efforts have fallen short of meeting these goals. One particularly relevant concern is the mismatch between students identified as at-risk and potential interventions to be triggered. For example, EWS may classify broad, heterogenous groups of students as at-risk, whereas interventions may be tailored to benefit certain students with specific risk factors (Bowers & Sprott, 2012; Freeman & Simonsen, 2015). McMahon and Sembiante (2019) join others (e.g., Carl et al., 2013) calling for EWS that not only accurately classify students as at-risk but also can effectively allocate school resources to improve student outcomes. Although EWS may focus on identifying students at risk of dropout, other outcomes are of interest to both researchers and practitioners. Carl et al. (2013) report on a system for Milwaukee Public Schools that predicts not only dropout but also ninth-grade promotion, college persistence, and college graduation. Allensworth et al. (2014) supplement the original Chicago Consortium on-track indicator with an eighth-grade indicator of high school readiness. Kieffer et al. (2014) take a similar approach back to fourth grade, using attendance rates to predict ninth-grade on-track status.

This line of research raises questions about how early an EWS can and should be implemented to predict student outcomes and how well this system tracks progress over time

(McMahon & Sembiante, 2019). Bowers (2010) estimates survival models using longitudinal data on students from grades one through 12 and argues that differential risks of dropout can be found beginning in grade seven. His work is notable for the length of the data panel and for producing quantitative estimates of student-specific dropout risk rather than qualitative indicators.

Probabilistic reports of dropout risk produce fine-grained results that are interpersonally and intertemporally comparable and will track progress over time towards a single, distal outcome of interest (Bowers, 2010; Carl et al., 2013). In contrast, traditional EWS methods rely on qualitative (e.g., binary or ordinal) measures typical of at-risk status, or the like. In fact, we he found that practitioners prefer such qualitative indicators, even if one may demonstrate measurable benefits of probabilistic indicators for many purposes. Of course, one may transform any quantitative, probabilistic indicator into a binary or ordinal indicator by choosing threshold probabilities that define the category into which each student falls. We have done just that in the CORE Districts to enhance the actionability of our results.

2.2. Algebra for All

Focusing on dropout risk, discussions in EWS literature on interventions have not typically been concerned with student coursework. However, with our added focus on college readiness, it is natural to consider how coursework not only predicts but also causes observed variation in postsecondary outcomes.

Policymakers have attempted to enhance college readiness by implementing policies to accelerate placement in Algebra I, known as "Algebra for All." These innovations led to research utilizing advanced statistical methods to identify causal effects of variation in course placement.

We seek to incorporate such methods in our proposed research and then apply them in an actionable EWS. For now, we simply seek to produce predictive models that account for these policy changes.

“Algebra for All” encouraged efforts to increase middle school enrollment in algebra, leading to a reported doubling of the eighth-grade enrollment rate nationally from 1990 to 2015, with a peak of 68 percent in California (McEachin et al., 2017). Clotfelter et al. (2015) argue that the push for middle-school algebra “has been inspired in large part by correlational research documenting significant differences in later-life outcomes between those students who enroll in algebra by 8th grade and those who do not” (p. 160). To assess causation, they exploit policy variation over time in a North Carolina school district to estimate local average treatment effects. They find that students who were affected by the enrollment policy—i.e., those who would not have taken Algebra I so early but for the policy change—were either adversely impacted or not impacted with respect to subsequent math success.

Dougherty et al. (2017) report on the results of a targeted policy in another North Carolina school system, whereby students were recommended for the Algebra I track if they were found based on prior test scores to have a 70 percent chance of passing a standardized algebra test. They use a regression discontinuity design (RDD) to estimate causal effects and find positive effects on equity in enrollment patterns but mixed impacts on math achievement, as the initial impact of accelerated coursework decays substantially over time. McEachin et al. (2017) employ a related approach (fuzzy RDD) in California, with empirically identified enrollment policies that vary across 335 schools. They find heterogeneous treatment effects of accelerated math, providing “descriptive evidence about the contexts in which early Algebra is and is not effective” (p. 28).

3. CHAIN-LINKED MODELS

3.1. A General Representation of a Chain-Linked Predictive Analytics System

To produce probabilistic forecasts of high school and postsecondary outcomes for students in elementary, middle, and high school, we developed and applied chain-linked models. Chain-linking uses short-span models and data to make predictions linked together from, for example, 3rd grade to 6th grade, 6th to 8th, 8th to 10th, and 10th to high school graduation.

A more straightforward approach would estimate models using longitudinal student data over the entire time span, as in Bowers (2010). Armed with such data, we could directly estimate graduation probabilities conditional on third-grade attributes (e.g., test scores and attendance rates). However, we faced two key challenges. First, we currently have just four years of data. Second, even if we did have a sufficiently long panel, California education policy has systematically shifted in the interim (such as eighth-grade math enrollment); thus, extrapolating from the distant past to the future would be ill advised. Our chain-linked models account for these problems.

Before discussing details of our predictive analytics models, we begin by describing a simple chain-linked system on which our predictive models are based. Consider three periods or “spans” of schooling s, $s \in \{1,2,3\}$. These spans may correspond to elementary school ($s = 1$), middle school ($s = 2$), and high school ($s = 3$). Student i attends school j in span s. Let y_{sij} denote a schooling outcome observed for student i at the end of span s. Let student i be characterized by a vector of attributes $\boldsymbol{x}_{(s-1)ij}$ observed at the beginning of span s, including variables listed above, such as test scores, demographics, course enrollment, and grades.

Suppose we want to forecast high school graduation outcomes y_{3i} for a sample of elementary school students with observed attributes $\boldsymbol{x}_{0ij}$, $i=1,\ldots,N,$ measured early in elementary school—e.g., 3rd grade attributes in our data. For simplicity, we suppress the j subscripts and consider linear models with common intercepts and slopes across schools, as described by this series of equations:

$$y_{1i} = \alpha_{(1,0)} + \boldsymbol{x}'_{0i}\boldsymbol{\beta}_{(1,0)} + \varepsilon_{1i} \qquad \text{(Ia)}$$

$$y_{2i} = \alpha_{(2,1)} + \boldsymbol{x}'_{1i}\boldsymbol{\beta}_{(2,1)} + \varepsilon_{2i} \qquad \text{(Ib)}$$

$$y_{3i} = \alpha_{(3,2)} + \boldsymbol{x}'_{2i}\boldsymbol{\beta}_{(3,2)} + \varepsilon_{3i} \qquad \text{(Ic)}$$

where α_s denotes the span-s intercept, $\boldsymbol{\beta}_{(s,s-1)}$ denotes the vector of slope parameters relating span-s outcomes to attributes observed at the end of span s-1, and ε_{si} denotes idiosyncratic error.

Suppose the conditional expectation functions are linear:

$$E\left[y_{si}|\boldsymbol{x}_{(s-1)i}\right] = \alpha_{(s,s-1)} + \boldsymbol{x}'_{(s-1)i}\boldsymbol{\beta}_{(s,s-1)},\ s = 1,2,3 \qquad \text{(II).}$$

Suppose further that span-1 outcomes y_{1i} are actually the span-2 right-hand-side attributes x_{1i} and, similarly, the span-2 outcomes y_{2i} are the span-3 attributes x_{2i}. Thus, x_{1i} and x_{2i} are now each one-dimensional. Then, we may write down a chain-linked model that connects span-3 outcomes to the attributes $\boldsymbol{x}_{0i}$ observed at the beginning of span 1:

$$y_{3i} = \alpha_{(3,2)} + \left[\alpha_{(2,1)} + \left(\alpha_{(1,0)} + \boldsymbol{x}'_{0i}\boldsymbol{\beta}_{(1,0)} + \varepsilon_{1i}\right)\beta_{(2,1)} + \varepsilon_{2i}\right]\beta_{(3,2)} + \varepsilon_{3i} \quad \text{(III)}.$$

Under the assumption that errors $\varepsilon_{si}, s = 1, 2, 3$, are mean independent of $\boldsymbol{x}_{s'i}$ for $s' < s$, we have the following conditional expectation function:

$$E[y_{3i}|\boldsymbol{x}_{0i}] = \left[\alpha_{(3,2)} + \alpha_{(2,1)}\beta_{(3,2)} + \alpha_{(1,0)}\left(\beta_{(2,1)}\beta_{(3,2)}\right)\right] + \boldsymbol{x}'_{0i}\boldsymbol{\beta}_{(1,0)}\left(\beta_{(2,1)}\beta_{(3,2)}\right) \text{(IV)}.$$

Thus, we use parameters of the short-span models (Ia-c) to produce chain-linked predictions of outcomes y_{3i} via equation (IV). In practice, our models are more complicated than (Ia-c), with numerous predictor variables and many more spans. This representation, however, captures the essence of our approach.

3.2. A taxonomy of student outcomes in a predictive analytics system

Table 1 presents a taxonomy that shows the types of student-level variables hypothesized to determine student outcomes. The variables are listed in order of the hypothesized causal impact of the variables. For current and future reference, suggested variable names are included for each type of predictor. A key difference between the predictors is that: (a) some are measured for all students, for example, enrollment variables and subject-specific test scores (typically measured at a given grade) and (b) some are measured only for subsets of students, for example, course level grades and end-of-course test scores.

Table 1. Taxonomy of Student Outcomes

Variable Name	Student-Level Variables
X	Student, family, and neighborhood resources and attributes
	Poverty status (free lunch eligibility) Parental education Student peer effects Note: additional variables could include those available in integrated data systems, for example information on health, family status, and homelessness.
S	School enrollment
D	Student diagnostics and program eligibility determinations Note: These diagnostics are measures of student status and are used to determine access to school-based inputs (see below)
	Gifted and talented Special education English language learner (ELL) Related: low performance outcomes
E	Inputs and Interventions
E_1	Course enrollments
	By subject and difficulty level
E_2	Programs and supports
	Tutoring Supplemental resources
F	Student inputs
	Student attendance
	Student Performance Outcomes: Cognitive. Non-cognitive, and Behavioral
$E \cdot G$	Outcomes for subsets of students
	Course grade points (on 4.0 scale) End-of-course test scores Program evaluation scores
Y	Outcomes for all students
	Subject-specific test scores (summative, interim, end-of-unit) Social emotional learning (SEL) Incidents and suspensions Interactions with the juvenile justice system Related: student attendance
A	Educational attainment outcome(s) High school graduation or graduation and completion of courses required for college admission (A-G requirements in California) Enrollment in college (two year or four year) Persistence in college, college completion
W	The complete vector of all student outcomes (including all of the above outcomes) measured at a given grade level. This vector will also include a column of ones to include intercepts in each equation, unless school indicators are included for all schools, in which case the intercept is excluded and is absorbed into the school effects.

Although our focus in this memo is on student inputs and outcomes, we are ultimately interested in evaluating the factors that determine variation in school effectiveness with respect to the numerous student outcomes listed above. The models presented below allow for school effects but do not include higher-level equations that allow for estimation of the impacts of school programs and policies. Table 2 illustrates what a taxonomy of school-level variables could look like.

Table 2. Prototype Taxonomy of School-Level Variables

Variable Name	School-Level Variables
Z	School and educator factors and quality
	Are advanced (e.g., AP) course offered? Are schools effective in hiring, developing, and retaining high quality educators? Is instruction aligned with (high quality) learning standards? Do educators have access to data on student progress during the school year? Are systems or protocols in place to ensure that resources are targeted to at-risk students?
S	Residual school effects not explained by observed variables

3.3. A predictive analytics system

In this section we present a predictive analytics system that includes student outcomes and school enrollments recorded at grades *7, 8, 9*, plus an educational attainment variable measured at some point after grade 9 (call this timepoint *A*). We include the latter variable so that the system includes a single outcome which can be used to rank predicted outcomes along a single dimension.[1] The vectors of student outcomes in grade g are represented by the row vector

[1]The educational attainment variable could be a continuous variable such as years of education. Alternatively, A_i could represent the latent variable underlying a model that measures educational attainment as a binary outcome such as graduation/non-graduation from high school.

W_{gi}. As indicated in Table 1, this vector includes all student outcomes measured in grade g. The educational attainment measure is represented by the continuous variable A_i. The school enrollment variables are represented by the row vector S_{gi}. To simplify the model, we assume that students only change schools between grades 8 and 9, but not between middle school grades and high school grades. Hence, the grade subscripts for the middle school and high school enrollment variables are equal to *8/7* and *A/9*, respectively. Since school enrollments only change between middle school and high school, the system includes only a single school enrollment equation. The complete system is given by:

$$A_i = W_{9i}\theta_A + S_{A/9i}\alpha_A + e_{Ai} \tag{1}$$

$$W_{9i} = W_{8i}\theta_9 + S_{A/9i}\alpha_9 + e_{9i} \tag{2}$$

$$S_{A/9i} = W_{8i}\pi_9 + S_{8/7i}\tau_9 + u_{A/9i} \tag{3}$$

$$W_{8i} = W_{7i}\theta_8 + S_{8/7i}\alpha_8 + e_{8i} \tag{4}$$

$$W_{7i} = W_{6i}\theta_7 + S_{8/7i}\alpha_7 + e_{7i} \tag{5}$$

where $\theta_A, \theta_9, \theta_8, \theta_7, \pi_9$ are conformable coefficient vectors or matrices for the student outcome variables, $\alpha_A, \alpha_9, \alpha_8, \alpha_7, \tau_9$ are conformable coefficient vectors or matrices for the school enrollment variables, and $e_{Ai}, e_{9i}, e_{8i}, e_{7i}, u_{A/9i}$ are conformable student-level residuals.[2]

[2] Note that some of the variables included the vector of predictors may be constant over time; for example, race/ethnicity and gender. These variables thus need not be predicted from one year to the next. The model notation can be extended to allow for variables that do not change over time. Alternatively, our preferred approach, the current model can be interpreted as including equations that

An important feature of the above models is that they include student outcome predictors measured at the end of each grade, but school enrollment indicators recorded in the subsequent grade. This model specification thus shares the same form as a value-added model, although the focus of a value-added model is to measure school effects (here, represented by α_g), whereas in a predictive analytics model we are interested in the predictive power obtained primarily from the student-level predictors. Nonetheless, predictions of educational attainment could include a part due to student-level predictors and a part due to predictions of future school effects.

The above models adopt the assumptions of a recursive model and thus include student outcome and school enrollment variables as predictors only from a single grade. If the assumptions of a recursive model are valid and variables are not measured with error, we can obtain valid predictions of future variables by chain-linking multiple models given the assumption that the predictive relationships between variables stays the same.[3] We refer to this assumption as the "steady state" or "no regime change" assumption.

In order to compute future predictions of student outcomes, for example, predictions of education attainment A_i given information as of 7th grade, it is necessary to estimate the parameters of each predictive analytics equation – the parameter calibration stage – and then chain link the models to allow predictions of outcomes for future cohorts (indexed by c) over periods longer than those of each of the models. Below, we illustrate the method of chain linking by linking equations (1) and (2). This chain-linked equation yields predictions of education

perfectly predict the variables that do not change, specifying coefficients equal to one for the variables that do not change and coefficients equal to zero for the other predictors.

[3] If variables are measured with error, then it is necessary to estimate all equations using a method, such as errors in variables (EIV) (Fuller, 1987) that controls for measurement error.

attainment given student outcomes as of 8th grade. The chain-linked equation is obtained by substituting equation (2) into equation (1), which yields:[4]

$$A_{ic} = [W_{8ic}\theta_9\theta_A + S_{A/9ic}(\alpha_{Ac} + \alpha_{9c}\theta_A)] + [e_{Aic} + e_{9ic}\theta_A] \tag{6}$$

Predicted attainment $\tilde{A}_{ic}$ and the error in predicting/forecasting attainment $\tilde{\varepsilon}_{Aic}$ are given by the two terms in brackets in the above equation:

$$\tilde{A}_{ic} = [W_{8ic}\theta_9\theta_A + S_{A/9ic}(\alpha_{Ac} + \alpha_{9c}\theta_A)] \tag{7}$$

$$\tilde{\varepsilon}_{Aic} = [e_{Aic} + e_{9ic}\theta_A] \tag{8}$$

The key feature of this pair of equations is that the coefficient vector θ_A from the attainment equation multiplies all components from the equations for grade *9*. This coefficient vector acts as scoring weights in making predictions of education attainment A_i given prior grade variables. In addition, the coefficient matrix on W_{8i} is the product of θ_9 (a $P_8 x P_9$ matrix) and θ_A (a $P_9 x1$ vector), yielding a ($P_8 x1$) vector of scoring weights for 8th grade student outcomes (where P_g = the dimension of the vector and matrix). The variance of prediction/forecast errors $\tilde{\varepsilon}_{Aic}$ is constructed using values of the variances of each error component. Note that some of the error variances are also multiplied by the scoring weights θ_A .

---

[4] Note that this equation can readily be extended to allow future school effects for cohort *c* (α_{Ac}, α_{9c}) to be forecasted, an extension that allows for the possibility that school effects may not be perfectly stable.

4. EMPIRICAL RESULTS *(VERY PRELIMINARY AND INCOMPLETE)*

We have estimated our chain-linked models using four years of panel data from the CORE Data Collaborative. We begin this section by reporting on some initial results for about 95,000 students who were enrolled in 9th grade during the 2018-19 school year and for whom we have sufficient data to calculate both our chain-linked predictions of high school graduation outcomes and a more conventional, qualitative indicator of at-risk status. After presenting some descriptive comparisons of these quantitative and qualitative indicators, we assess the predictiveness of these types of indicators using data on graduation outcome for an earlier cohort on 9th grade students.

4.1. 2018-19 Cohort of 9th Graders

Currently, our chain-linked predictions for 9th graders begin with 9th grade data on program participation (gifted and talented, special education, English Language Learner (ELL)), attendance rate, course enrollment by difficulty level (ELA, math, science, social science), course grades (separately for ELA, math, science, and social science, as well as GPA in other courses), and standardized scores on math and ELA SBAC tests administered in 8th grade. We focus here on the chain-linked models that are used to predict which of three high school on-time graduation outcomes will be realized: no on-time graduation, graduation not meeting "A-G requirements", or graduation meeting A-G requirements, where the meeting of A-G requirements based on successful completion of "A-G coursework" determines eligibility for admission as a freshman to universities in the University of California (UC) and the California State University (CSU) systems.[5]

Here we compare predictions based on these 9th-grade models to qualitative at-risk indicator. This indicator takes on five values—1, 2, 3, 4, or 5—where a student's value is determined by a checklist approach using information on attendance rates, overall GPA, grades in A-G courses, and total credits earned. The lower the value, the more "at-risk."

To facilitate comparisons, as well as to enhance the actionability of our predictive analytic models, we created a similar qualitative indicator based on our probabilistic predictions but only taking on five values—I, II, III, IV, or V—where lower values are associated with lower probabilities of graduation. Table 3 reports the share in each indicator category. We have the same 94,959 students in each sample.

---

[5] See, for example, https://www.ousd.org/domain/1724, which states: The University of California (UC) and California State University (CSU) require entering freshmen to complete certain courses in high school, commonly known as the "A-G requirements" because of the letter each subject area is assigned:

A. History/Social Science - 2+ years including 1 of World, 1 of US and/or Government
B. English - 4 years of college preparatory English
C. Mathematics - 3+ years of including topics covered in basic and advanced algebra and two-and three-dimensional geometry
D. Laboratory Science - 2+ years of laboratory science in at least 2 of the 3 disciplines of biology, chemistry and physics
E. Language Other than English - 2+ years of the same language (other than English)
F. Visual/Performing Arts - 1+ years of dance, music, or visual arts.
G. College Preparatory Elective - 1+ years (additional year of any requirement above or other elective)

To receive "A-G" credit, a high school course must be certified through the UC course approval process.

Table 3. Predictive Indicator and Checklist Indicator Category Shares

9th Grade Predictive Analytic Indicator		9th Grade Checklist Indicator	
Category	Share	Category	Share
I	6.1%	1	32.3%
II	40.4%	2	24.5%
III	15.9%	3	14.1%
IV	15.6%	4	18.3%
V	22.0%	5	10.7%

Table 3 shows that the checklist indicator assigns almost one-third of students to the lowest category and another one-quarter to the next lowest category. It is our understanding that practitioners may err on the side of labelling more students as at-risk in order to avoid lulling too many students into a false sense of security. In contrast, the probability-based indicator assigns just over five percent of the sample to the lowest category, which includes those students for whom the high school graduation probability is estimated to be no more than a 50 percent chance. Further, whereas the checklist indicator assigns just over ten percent of the sample to the highest category, more than twice that share of the sample is found in the highest category of the probability-based indicator, which includes students estimated to have at least a 75 percent chance of graduating high school while meeting the A-G requirements.

Table 4A reports some descriptive statistics for the 9^{th} graders in our sample, conditional on this qualitative indicator. Table 4B presents the same statistics for the checklist-based indicator.

Table 4A. Descriptive Statistics Conditional on Predictive Analytic Indicator

9th Grade Predictive Analytic Indicator Category	N	Program Participation		8th Grade SBAC Scores			
				Math (normalized)		ELA (normalized)	
		Percent ELL	Percent GATE	N	Mean	N	Mean
I	5,791	49.6%	3.5%	4,722	-1.18	4,622	-1.2
II	38,385	32.7%	5.0%	36,752	-0.72	36,368	-0.71
III	15,065	19.4%	10.4%	14,792	-0.15	14,720	-0.12
IV	14,833	12.1%	19.2%	14,658	0.25	14,605	0.27
V	20,905	4.1%	49.9%	20,775	0.92	20,750	0.88

Table 4B. Descriptive Statistics Conditional on Checklist Indicator

9th Grade Checklist Indicator Category	N	Program Participation		8th Grade SBAC Scores			
				Math (normalized)		ELA (normalized)	
		Percent ELL	Percent GATE	N	Mean	N	Mean
1	30,703	31.7%	5.8%	28,264	-0.76	27,926	-0.74
2	23,286	24.3%	12.3%	22,814	-0.33	22,628	-0.3
3	13,382	21.3%	18.7%	13,236	-0.05	13,194	-0.07
4	17,427	12.4%	26.8%	17,278	0.41	17,227	0.42
5	10,181	6.1%	50.7%	10,107	1.1	10,090	1.03

The general patterns found in Tables 4A and 4B tend to be similar, and unsurprising, across indicators. In particular, students assigned to higher categories are less likely to be classified as ELL, more likely to participate in gifted and talented programs, and tend to obtain higher scores on the Math and ELA SBAC tests.[6] These similarities suggest that the indicators are closely related. Table 5 quantifies this relationship.

[6] We report the sample sizes for the SBAC scores, because we are missing 8th-grade test score data for about two (Math) or four (ELA) of our 9th grade sample. Note that we allow for some missing data in the

Table 5. Distribution of PA Indicator Values, Conditional on Checklist Indicator

9th Grade Checklist Indicator Category	9th Grade Predictive Analytic Indicator Category					Total
	I	II	III	IV	IV	
1	18.4%	71.0%	6.3%	2.6%	1.7%	100%
2	0.6%	56.5%	24.4%	12.7%	5.8%	100%
3	0.0%	21.7%	34.5%	24.4%	19.4%	100%
4	0.0%	2.9%	15.4%	34.8%	47.0%	100%
5	0.0%	0.4%	1.7%	17.0%	81.0%	100%

Here we see that almost 90 (100) percent of students in the lowest (highest) checklist category are in one of the two lowest (highest) probability-based categories. The relationships for the intermediate categories (2-4) are not quite as strong, but the great majority are in the diagonal or adjacent entries.

Table 6A. Distribution of PA Model Probabilities, Conditional on Predictive Analytic Indicator

9th Grade Predictive Analytic Indicator Category	N	Distribution of 9th Grade PA Predictions											
		% Chance Do Not Graduate HS						% Chance Graduate A-G					
		mean	Quantiles					mean	Quantiles				
			0.05	0.25	0.50	0.75	0.95		0.05	0.25	0.50	0.75	0.95
I	5,791	67%	51%	56%	64%	76%	91%	0%	0%	0%	0%	0%	2%
II	38,385	20%	7%	11%	18%	28%	44%	8%	0%	2%	6%	13%	22%
III	15,065	6%	3%	3%	5%	6%	13%	37%	26%	31%	37%	43%	49%
IV	14,833	2%	1%	1%	2%	3%	6%	63%	51%	57%	63%	69%	74%
V	20,905	1%	0%	0%	1%	1%	3%	88%	76%	82%	88%	94%	98%

estimation of our predictive analytic probability models, which include 8th-grade test scores as predictors. The missing data rate for test scores is much higher in the lower categories, ranging from about 20 percent in category I to five percent in category II to two percent or less in categories III, IV, and V.

Table 6B. Distribution of PA Model Probabilities, Conditional on Checklist Indicator Category

9th Grade Checklist Indicator Category	N	Distribution of 9th Grade PA Predictions										
		% Chance Do Not Graduate HS						% Chance Graduate A-G				
		mean	Quantiles					mean	Quantiles			
			0.05	0.25	0.50	0.75	0.95		0.05	0.25	0.50	0.75	0.95
1	30,703	31%	4%	14%	25%	43%	75%	9%	0%	1%	2%	9%	46%
2	23,286	11%	1%	5%	9%	15%	30%	28%	2%	8%	20%	42%	77%
3	13,382	5%	1%	2%	4%	7%	13%	48%	11%	28%	45%	69%	94%
4	17,427	2%	0%	1%	1%	2%	5%	70%	32%	56%	73%	86%	96%
5	10,181	1%	0%	0%	1%	1%	2%	85%	61%	78%	88%	94%	99%

Tables 6A and 6B describe how the distribution of estimated graduation probabilities varies with the probability-based and checklist indicator categories, respectively. In Table 6A, we see how category I is defined by the probability of not graduating from high school on time—at least 50 percent chance, are reported in the left panel. Those with at least a 50 percent chance of graduating are then assigned to one of the four higher categories based on the probability of graduating while meeting A-G requirements, as reported in the right panel: (II) less than 25 percent chance , (III) 25 to 50 percent chance, (IV) 50 to 75 percent chance, or (V) more than a 75 percent chance.

Finally, we see in Table 6B some key differences between the probability-based indicator and the checklist indicator in terms of the composition of students in comparable categories. In particular, half the students in the lowest checklist category—that is, category 1—are estimated to have at least 75 percent chance of graduating from high school, whereas all students in PA category I have less than a 50 percent chance. Further, the top five percent of category 1 students are estimated to have at least a 46 percent chance of graduating A-G, and one-quarter of category-2 students are estimated to have at least a 42 percent chance of graduating A-G.

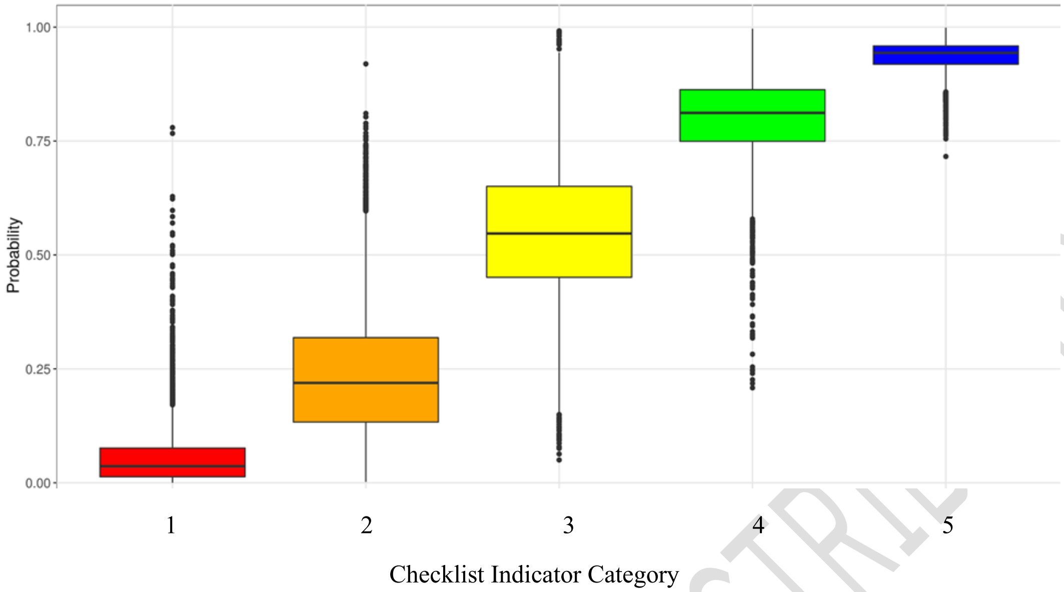
In the sample as a whole, the median estimated probability of A-G graduation is about a 30 percent chance. Thus, sizable shares of the students in the two lowest checklist categories are estimated to have relatively high probabilities of A-G graduation. One naturally wonders whether these chain-linked predictions are valid.

4.2. 2015-16 Cohort of 9th Graders

To assess the predictive validity of our probability models, we need to look at a cohort for whom we have data not only on 9th-grade predictors but also on graduation outcomes. Therefore, we look back to students who were in the 9th grade in the 2015/16 school year and were in the class of 2019, the most recent year for which we have graduation outcomes. Whereas we have the data available to assign these students to the checklist indicator categories, we have more incomplete data on the predictors used to estimate the chain-linked predictive analytic models. To make an illustrative comparison prior to obtaining data on the class of 2020, we have instead directly estimated graduation probability models directly with a reduced set of predictors.

These probability models tend to generate predictions that vary similarly with the checklist indicator categories, as depicted in Figure 1 and may be compared to entries in the right panel of Table 6B.

Figure 1. Distribution of A-G Graduation Probabilities, Conditional on Checklist Indicator



Figures 2 thru 5 show how the average graduation rates ("Actual AG Grad Rate") based on 2019 graduation outcomes vary with both the checklist indicator category and the estimated A-G graduation probability using he probability models with 9th-grade data in 2016. For example, Figure 2 shows that, *within the interquartile range of checklist category 1*, the mean of the estimated A-G probability is 3.9 percent and the realized graduation rate is 3.4 percent. Further, *within the group of checklist category 1 students with the highest estimated probabilities*, the mean of the estimated A-G probability is 33.6 percent and the realized graduation rate is 39.8 percent. Thus, the probability models are picking up some positive predictors of success among these students who are labelled as being in the most at-risk group. Similar inferences may be drawn from Figures 3 thru 5 for the other checklist categories.

Figure 2. Distribution of A-G Graduation Probabilities and Realized Rates, Conditional on Checklist Indicator Category 1

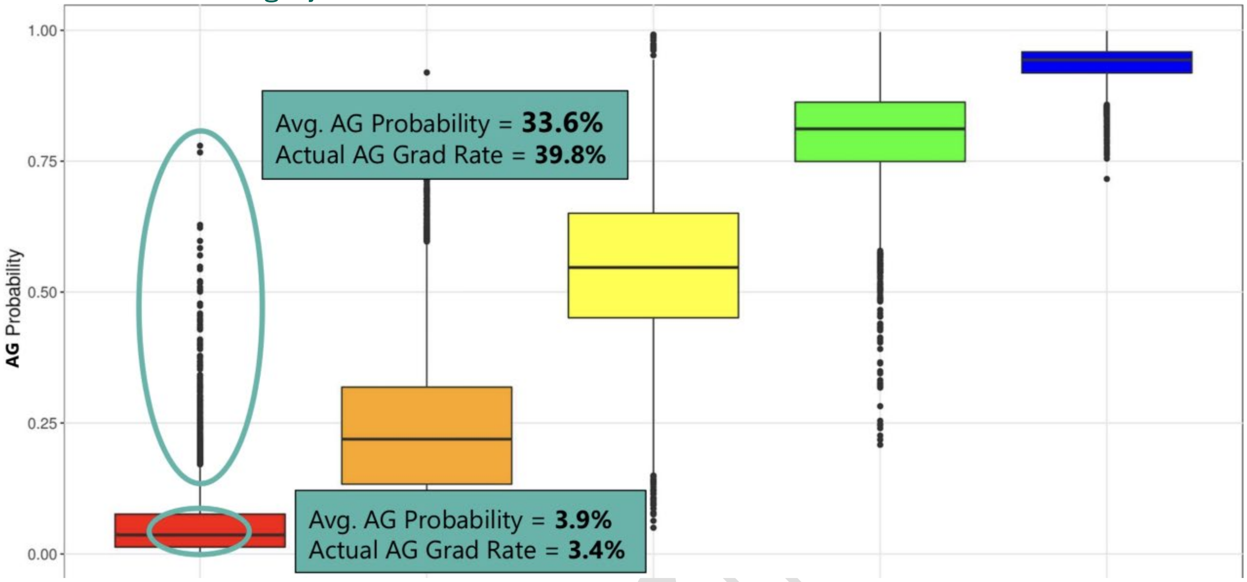


Figure 3. Distribution of A-G Graduation Probabilities and Realized Rates, Conditional on Checklist Indicator Category 2

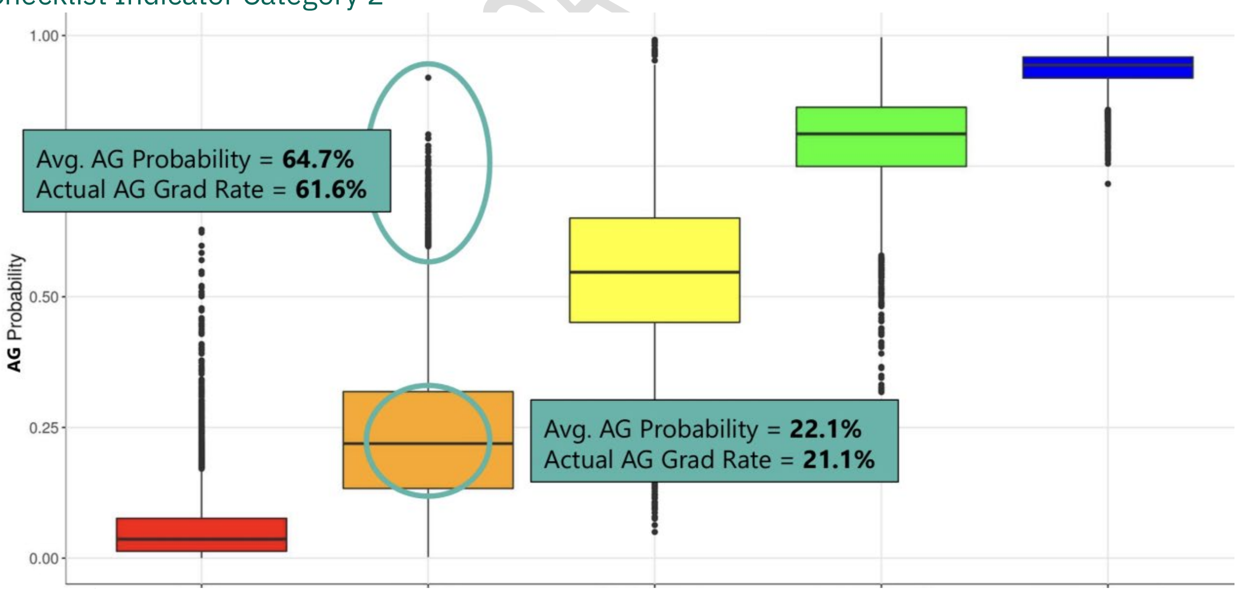


Figure 4. Distribution of A-G Graduation Probabilities and Realized Rates, Conditional on Checklist Indicator Category 4

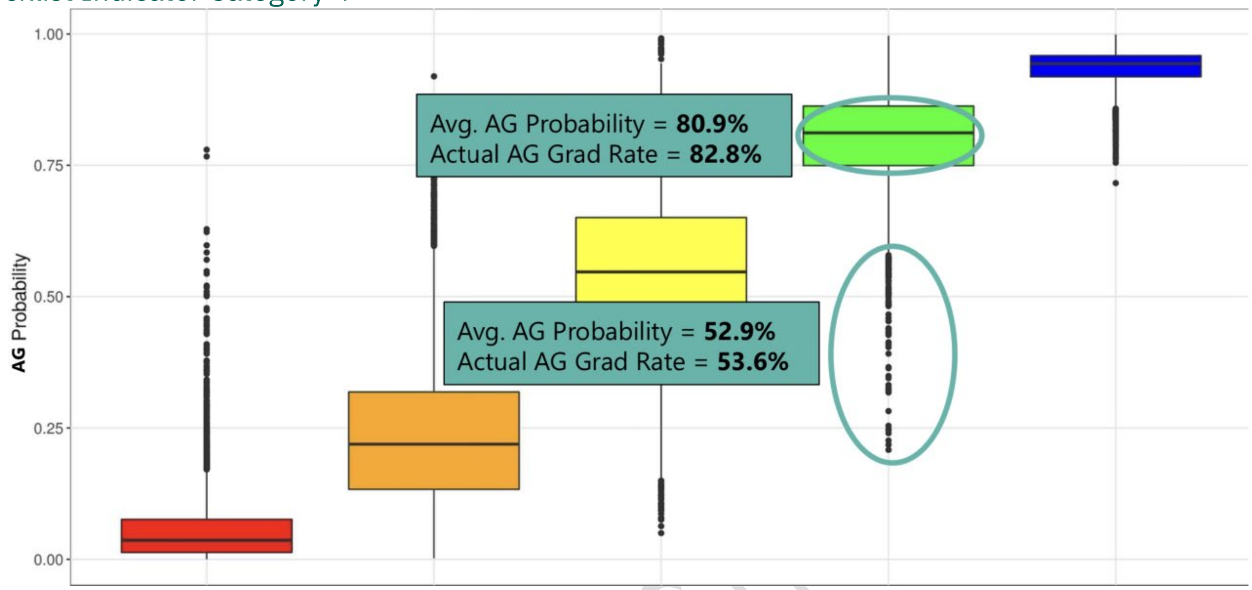
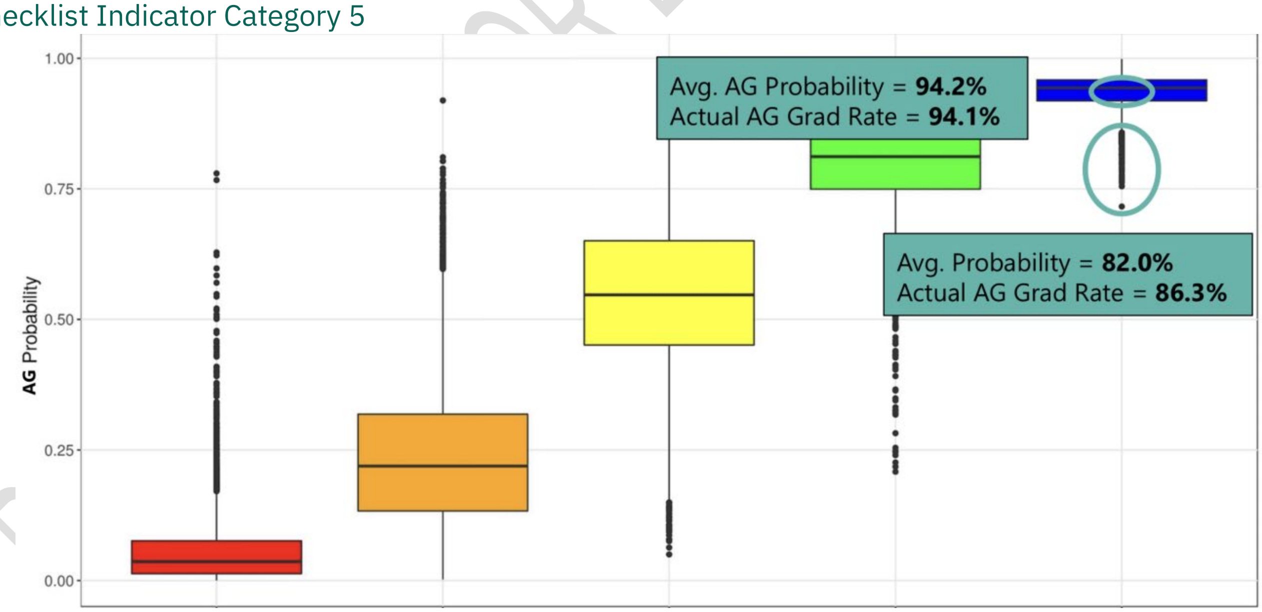


Figure 5. Distribution of A-G Graduation Probabilities and Realized Rates, Conditional on Checklist Indicator Category 5



5. DIRECTIONS FOR FUTURE WORK

Our predictive models result in estimated statistical relationships that are correlational in nature, giving rise to the standard correlation-causation critique raised by Clotfelter et al. (2015). Thus, while our models may predict better outcomes for students who are enrolled in a higher-level math course in ninth grade, one should not leverage those results to encourage specific ninth graders to enroll in more challenging math courses. Yet, this is the very use our partners desire, as has also been documented elsewhere (McMahon & Sembiante, 2019). We therefore seek to build a companion to our predictive models that utilizes contingent analytics to answer questions of "What if?," tailoring the answers to the attributes of different types of students and properly accounting for the determinants of observed variation in outcomes.

As noted above, we have found that practitioners often wish to use qualitative indicators rather than quantitative assessments of risk. An additional contribution of our work will be to clarify statistical concerns about classification error, predictive accuracy, and related issues for the use of qualitative indicators to forecast long-term outcomes, as in Bowers and Zhou (2019). By starting with a well-specified predictive model as the basis for an EWS, we may use model estimates to directly assess these issues and formalize the tradeoffs between different classification methods.

REFERENCES

Allensworth, E., & Easton, J.Q. (2005). *The On-Track Indicator as a Predictor of High School Graduation*. Chicago: Consortium on School Research Report

Allensworth, E., and Easton, J.Q. (2007). *What Matters for Staying On-Track and Graduating in Chicago Public Schools*. Chicago: Consortium on School Research Report.

Allensworth, E., Gwyne, J.A., Moore, P., & de la Torre, M. (2014). *Looking Forward to High School and College*. Chicago: Consortium on School Research Report.

Bowers, A. J. (2010). Grades and Graduation: A Longitudinal Risk Perspective to Identify School Dropouts. *Journal of Educational Research*, *103*(3), 191-207.

Bowers, A. & Sprott, R. (2012). Why tenth graders fail to finish high school: A dropout typology latent class analysis. *Journal of Education for Students Placed at Risk*, *17*(3), 129–148.

Bowers, A., & Zhou, X. (2019). Receiver Operating Characteristic (ROC) Area Under the Curve (AUC): A Diagnostic Measure for Evaluating the Accuracy of Predictors of Education Outcomes. *Journal of Education for Students Placed at Risk (JESPAR)*, *24* (1), 20-46. https://doi.org/10.1080/10824669.2018.1523734

Bruce, M., Bridgeland, J. M., Fox, J. H., & Balfanz, R. (2011). On Track for Success: The Use of Early Warning Indicator and Intervention Systems to Build a Grad Nation. *Civic Enterprises*.

Bruce, Mary, John M. Bridgeland, Joanna Hornig Fox, and Robert Balfanz (2011). "On Track for Success: The Use of Early Warning Indicator and Intervention Systems to Build a Grad Nation," Civic Enterprises & Everyone Graduates Center Report.

Carl, B., Richardson, J. T., Cheng, E., Kim, H., & Meyer, R. H. (2013). Theory and application of early warning systems for high school and beyond. *Journal of Education for Students Placed at Risk (JESPAR)*, *18*(1), 29-49.

Clotfelter, C.T., Ladd, H.F., & Vigdor, J.L. (2015). The Aftermath of Accelerating Algebra: Evidence from District Policy Initiatives. *Journal of Human Resources 50*(1), 159-188. doi:10.1353/jhr.2015.0005.

Dougherty, S. M., Goodman, J. S., Hill, D. V., Litke, E. G., & Page, L. C. (2017). Objective course placement and college readiness: Evidence from targeted middle school math acceleration. *Economics of Education Review*, *58*, 141–161. doi:10.1016/j.econedurev.2017.04.002

Freeman, J. & Simonsen, B. (2015) Examining the impact of policy and practice interventions on high school dropout and school completion rates: A systematic review of the literature. *Review of Educational Research*, *85*(2), 205–248.

Fuller, W. A. (1987). *Measurement Error Models*, New York: John Wiley & Sons.

Kieffer, M., Marinell, W. & Neugebauer, S. (2014). Navigating into, through, and beyond the middle grades: The role of middle grades attendance in staying on track for high school graduation. *Journal of School Psychology*, *52*, 549–565.

McEachin, A., Domina, T., & Penner, A.M. (2020). Heterogeneous Effects of Early Algebra across California Middle Schools. *Journal of Policy Analysis and Management*. https://doi.org/10.1002/pam.22202

McMahon, B. M., & Sembiante, S.F. (2019). Re-Envisioning the Purpose of Early Warning Systems: Shifting the Mindset from Student Identification to Meaningful Prediction and Intervention. *Review of Education*, *8* (1), 266-301. https://doi.org/10.1002/rev3.3183.

Rumberger, R., Addis, H., Allensworth, E., Balfanz, R., Duardo, D., & Dynarski, M. (2017). *Preventing Dropout in Secondary Schools*. IES: What Works Clearinghouse: https://ies.ed.gov/ncee/wwc/Docs/PracticeGuide/wwc_dropout_092617.pdf