Method

Sample Characteristics

Based on the experiences of other states with early warning systems, an extensive literature review, and several meetings with data analysts at CSDE, we decided to focus on a group of variables that have been shown to affect student performance. In order to verify that the measures chosen were significantly related in our data, we extracted six cohort samples of students in grades 1 through 6. Each sample was grade-level specific and contained more than 32,000 longitudinal data records. Table 1 summarizes the sample of students in each model.

Table 1
Sample characteristics for Grades 1-6

Grade	n	% Female	% SpEd	% FRL	% ELL	Mean Attendance % Last Year
1	34,096 (35,478)	48.5 (48.5)	10.8 (10.9)	38.8 (39.4)	11.3 (12.0)	94.98 (94.98)
2	32,588 (35,899)	48.6 (48.5)	11.2 (11.6)	40.3 (40.9)	10.5 (12.3)	95.57 (95.53)
3[a]	32,579 (37,147)	48.6 (48.7)	12.8 (13.5)	40.7 (42.9)	9.2 (11.5)	95.77 (95.75)
4	36,858 (38,062)	48.4 (48.1)	12.7 (14.3)	41.7 (42.0)	10.4 (10.5)	95.66 (95.88)
5	38,087 (39,398)	49.2 (48.8)	13.4 (15.0)	41.2 (41.4)	8.5 (8.6)	96.00 (95.94)
6	36,653 (38,841)	49.1 (48.7)	14.4 (16.2)	39.8 (40.6)	6.3 (7.3)	95.98 (95.92)

Note. Each value in parentheses reflects the population value across all Connecticut students in that grade.
[a] The large disparity between population and sample size in Grade 3 is due to missing data.

Measures

We then ran a series of linear regression, logistic regression, and HLM models to determine which of the variables were consistently significant. Table 2 summarizes the indicators used in each model.

Table 2
EIT Indicators for Grades 1-6 Models

Indicator	Description	Grade-level models in which indicator was included					
		1	2	3	4	5	6
Attendance_pct_LY	Percentage of school days attended last year	X	X	X	X	X	X
Attendance_pct_2Y	Percentage of school days attended two years ago		X	X	X	X	X
ISS_2Y[a]	In-school suspension days during the last two years	X	X	X	X	X	X
OSS_2Y[a]	Out-of-school suspension days during the last two years	X	X	X	X	X	X
School_moves_2Y[a]	Number of times the student changed schools during the last two years	X	X	X	X	X	X
Repeater_Current	Student is repeating the current grade	X	X	X	X	X	X
Repeater_LY	Student repeated the previous grade	X	X	X	X	X	X
KINV_LLN	Aggregate score on Kindergarten Inventory (sum of Language, Literacy, and Numeracy; max score: 9)	X	X	X			
SBAC_Math_LY	Smarter Balanced Assessment Consortium (SBAC) mathematics scale score last year				X	X	X
SBAC_Math_2Y	SBAC mathematics scale score two years ago					X	X
SBAC_ELA_LY	SBAC English Language Arts (ELA) scale score last year				X	X	X
SBAC_ELA_2Y	SBAC ELA scale score two years ago					X	X

Note. [a] Unless a student repeated kindergarten, s/he will only have one year of values for ISS_2Y, OSS_2Y, and School_moves_2Y for the Grade 1 model.

Regression Results

Several explanatory variables and combinations of these variables were tested, and the models that had theoretical support and the best overall explanatory power included attendance, discipline, mobility, and assessment. The kindergarten through 3rd grade model results provide qualitative and quantitative support for the inclusion of the Kindergarten Entry Inventory (KEI) assessment results, first and second grade attendance rate, the number of school moves over the period, the number of in-school suspensions and the number of out-of-school suspensions. All variables had the expected sign and, unless noted, were significant at the .05 level.

Table 3 presents model fit and effect size information. Overall, the K-3 model explained 33 percent of the variation in ELA and 34 percent of the variation in math for 3rd grade Smarter Balanced scale scores. With the addition of Smarter Balanced scale scores as an explanatory variable, the models for the later grade spans showed improved model fit. In fact, R-squared values for predicting 6th grade Smarter Balanced ELA and math scale scores were 0.67 and 0.75, respectively. The HLM model results suggested that school effects were minimal, so the more parsimonious single-level models were retained for all grade spans.

Table 3
Model Fit/Effect Size for EIT Models

	Model Fit/Effect Size	
Model	ELA	Math
Grades K-3	0.33	0.34
Grades 4-6	0.67	0.75

Note. For the ELA and math models, model fit reflects R-squared for the model.

Getting to Targeted Support Levels

To determine the level of targeted support suggested by the variables in the regression models, we explored whether there are interactions among the explanatory variables that could be discerned with a clustering approach. Since the association between indicators does not hold for all students within the population, we used latent profile analysis (LPA), a person-centered, model-based, probabilistic clustering approach, to classify students by targeted support level. Among the benefits of this person-centered approach is the ability to consider indicators simultaneously and describe similarities and differences among individuals with respect to how indicators relate to each other.

In addition, we explored different supervised modeling approaches that can be used for classification tasks. All models were developed using holdout sample validation, with 80 percent of the sample designated to model training, and 20 percent dedicated exclusively to testing. Random forest, a machine-learning approach that creates a prediction model based on an ensemble of hundreds of decision trees (hence “forest”), produced the best classification results of the supervised learning models. The random forest model captures complex interactions between variables and achieves greater accuracy and stability in the classification and prediction of the outcome.

Using the indicators in Table 2, we developed one hybrid model that combined LPA and random forest approaches for each grade from Grade 1 to Grade 3. By going beyond simply forecasting SBAC scores or predicting at-risk designations, this modeling approach expands the universe of students who may be identified for timely and targeted interventions and support. LPA output contains cluster identifiers at the student level as well as probabilities that each student may belong in another cluster. Random forest models produce probabilities that a student will reach a particular outcome (e.g., proficient in 3rd Grade on the ELA Smarter Balanced assessment). The hybrid models integrate these probabilities to assign each student to a targeted support level.

In the case of students that are new to Connecticut public schools, or for which significant data elements are missing for whatever reason, no level of targeted support was assigned.

Enhancements for Grades 7 to 12

The models for grades 7 to 12 focus on two outcomes:

1. On-track in 9th Grade (for students in grades 7, 8, and 9)[1]
2. College and Career Ready (for students in grades 10, 11, and 12)[2]

The models for grades 7 to 12 use a more robust set of predictors, including detailed special education covariates (including primary disability and time with non-disabled peers), course performance data (starting in grade 7), and PSAT scale scores. The *Early Indication Tool – FAQ* document on the Resources tab of the on-line Early Indication Tool provides additional details of the EIT models and report.

[1] For purposes of the EIT, On Track to High School Graduation in Grade 9 means a student who has earned at least five full-year credits and has failed no more than 0.5 credits in a core subject area (i.e., English, Mathematics, Science, or Social Studies) in 9th grade.

[2] For purposes of 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 (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.