EIT for Teachers: Methodology and Rationale for Estimating Smarter Balanced Scale Scores for the 2019-20 Academic Year

Background

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 Connecticut State Department of Education (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. Ultimately, CSDE wants more students to meet academic milestones and graduate from high school.

Summative assessments play a key role in the evaluation of student learning, and test scores are among the critical factors in determining an *EIT* Support Level for each student. Moreover, Smarter Balanced Assessment Consortium (SBAC or Smarter Balanced) exam results are a big part of Connecticut's Next Generation Accountability System. Since the SBAC assessments were not administered in 2019-20, there is a gap in all areas impacted by summative assessment reporting for that academic year.

Rationale for Estimating Missing Assessments and Providing the *EIT for Teachers*

The Connecticut State Department of Education sought to address assessment-related gaps by incorporating cross-domain, longitudinal data and advanced modeling methodologies to estimate student scale scores for the 2019-20 SBAC assessments and set growth targets for 2020-21 SBAC assessments. In addition, CSDE developed a specialized report (the *EIT for Teachers)* to provide teachers with these estimated scores and growth targets—along with historical data and *EIT* results—for their incoming students, to inform instruction for the upcoming school year. The ultimate goal of the *EIT for Teachers* is to facilitate sharing of these data with teachers in order to decrease the amount of time spent on formal assessment upon the return to school in accordance with Connecticut's Sensible Assessment Practices guidance.

Methods

The methods used to create models to estimate Smarter Balanced scale scores involved data preparation in addition to model training, testing, and comparison before the working models were established. Each model developed to estimate 2019-20 scale scores was fit using training data collected from the population of Connecticut public school students who took the SBAC assessments in the 2018-19 academic year.

To be included in the training sample for grades 4 through 7, each student record was required to have a value for the dependent variable (i.e., SBAC scale score) in both English language arts (ELA) and mathematics in 2018-19 in the model grade, as well as attendance data and SBAC scores in the previous grade. To be included in the training sample for grade 3, each student record was required to have a value for the dependent variable in both ELA and math in 2018-19 in grade 3, as well as grade 2 attendance. All continuous predictor variables were standardized using student-level standard deviations prior to model development.

Missing assessment scores (SBAC or Kindergarten Entrance Inventory) and attendance values from two or more years ago were imputed using multiple imputation. The missingness of assessment scale scores was treated as a predictor: Imputed scores were flagged, and the flag and imputed scores 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. The grade-level mean was used for all other missing values. 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.

Scores were estimated using multilevel linear regression modeling techniques. These multilevel models best captured the hierarchical structure of the data, as all students are “nested” within schools. Furthermore, multilevel linear growth models were developed for grades 6 and 7. These growth models captured an additional level of nesting, since students in the upper grades have taken multiple annual Smarter Balanced assessments. Since the Smarter Balanced assessments are first administered in grade 3, more than 90% of students in grades 6 and 7 have the requisite three (or more) test scores on which to build a growth model. Modeling the statistical effects occurring at these multiple levels of aggregation helped improve scale score estimates. Figures 1 and 2 below present the general data structure captured by the models.

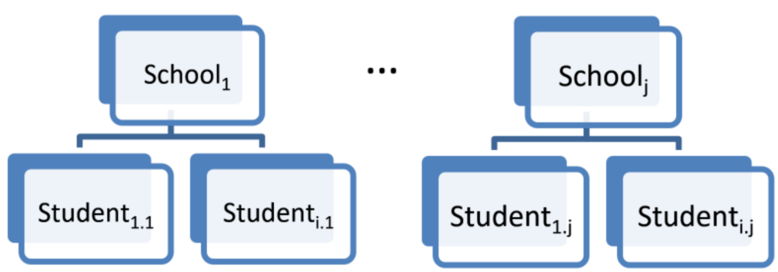

Figure 1. Multilevel regression model (used to estimate grade 3, 4, and 5 scores)

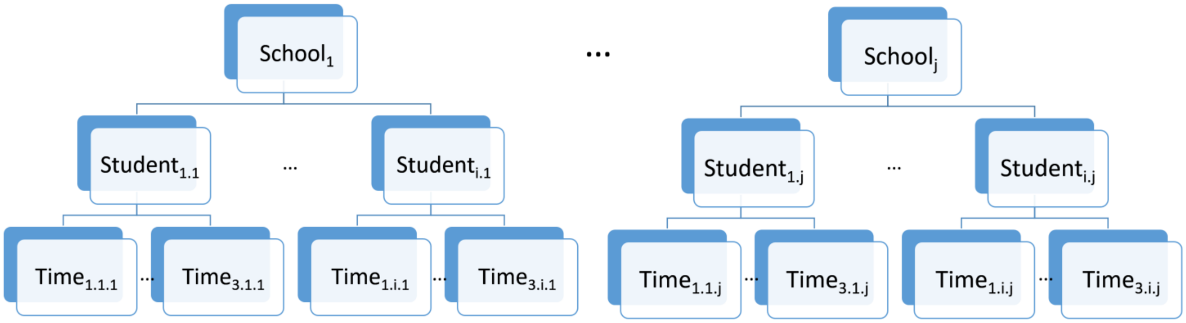

Figure 2. Three-level linear growth model (used to estimate grade 6 and 7 scores)

Unstable parameter estimates due to overfitting and collinearity are concerns in regression models. Overfitting occurs when the estimated model performs well with the original data, but poorly when applied to other data sets. Collinearity (or multicollinearity) occurs when two (or more) predictors are highly correlated. Consequently, elastic net linear regression models—which use regularization methods to simultaneously do (1) automatic variable selection; (2) continuous shrinkage of regression coefficients; and (3) group selection of correlated variables—were developed as part of this analysis. Results from these elastic net models complemented multilevel methods by helping to ensure that important predictors were not excluded from consideration.

Grade-specific models for grades 3 through 7 were developed for ELA and mathematics. Since linear growth models based only on prior test scores were found to over-estimate student scores, the final models include a range of student predictors (e.g., mobility, discipline, attendance, and achievement), as well as school- and district-level predictors, including indices from the Next Generation Accountability system.

In Connecticut public schools, student enrollments vary over time (i.e., students do not necessarily remain in the same school across all grades included in the growth models). To correctly account for this reality, acute-effects cross-classified random-effects models (CCREM) were developed for grades 6 and 7. The use of CCREM that allow for time-varying group membership instead of hierarchical models that assume constant group membership over time provided a truer fit and representation of the data.

Table 1 provides an overview of the predictor fields across all grades. For all models, student-level predictors were limited to those from previous grades. In other words, for a third grade student, data through the end of 2nd grade was used to estimate their 3rd grade scores. School- and district-level predictors included a one-year and two-year look back at enrollment, poverty rates, and Next Generation Accountability results. In addition, the four-year trailing school mean of SBAC scale scores was found to be a statistically significant predictor. Gender and race/ethnicity were not included as predictors in any of the models. The Resources page for the *EIT for Teachers* report includes a document with a glossary of terms and data definitions.

Table 1
Overview of Predictors Used in Models to Estimate Smarter Balanced 2019-20 Scale Scores

Domain	Elements	Grade level model in which predictor was used: 3	4	5	6	7
Student demographics	English learner (EL) status; Special Education status; free and reduced lunch (FRL) eligibility	X	X	X	X	X
Attendance	Percentage of school days attended	X	X	X	X	X
Behavior	Total in-school and out-of-school suspension incidents; total sanction days	X	X	X	X	X
Mobility	Schools and districts attended; number of school and district moves outside of the natural progression (e.g., elementary to middle school)	X	X	X	X	X
Special Education	Primary disability (if applicable); percentage of time with non-disabled peers (TWNDP)	X	X	X	X	X
Retention	Total grade repeats	X	X	X	X	X
Assessments	Smarter Balanced Assessment Consortium (SBAC) mathematics and English Language Arts (ELA) scale scores; school means over the prior four academic years		X	X	X	X
Assessments	Kindergarten Entrance Inventory (KEI)	X				
Performance Index	Performance index values for school and district	X	X	X	X	X
School and district demographics	Enrollment; percent poverty	X	X	X	X	X

Within a single dataset and modeling technique, Akaike information criterion (AIC) and Bayesian information criterion (BIC) values were used to assess the relative quality of the statistical models. When comparing the best models resulting from single-level and multilevel techniques, the model with the lowest root mean square error (RMSE; a measure of how spread out the estimation errors are) and highest R-squared (i.e., percentage of the variance in the dependent variable that the independent variables explain collectively) was selected as doing the best job. Lastly, models were applied to estimate 2020 SBAC scores for grades 3 to 7.

Results

Table 2 presents the RMSE values for the best-performing models. This table also includes an Estimate Interval value for each model that is equal to 1.6*RMSE. During model development, 90% of score estimates fell within 1.6*RMSE of the actual scores. The estimate intervals are displayed in the *EIT for Teachers* report to better reflect the uncertainty in the estimates.

Table 2
RMSE and Estimate Intervals by Grade and Subject

Grade	Subject	RMSE	Estimate Interval	Scale Score Range
3	ELA	64.9	104	2114-2623
4	ELA	48.5	78	2131-2663
5	ELA	45.0	72	2201-2701
6	ELA	31.3	50	2210-2724
7	ELA	33.6	54	2258-2745
3	Math	58.1	93	2189-2621
4	Math	36.7	59	2204-2659
5	Math	38.3	61	2219-2700
6	Math	26.4	42	2235-2748
7	Math	28.4	45	2250-2778

Discussion

The estimated Smarter Balanced scale scores for the 2019-20 academic year include a measure of uncertainty. To be clear: These estimates should not be used as support for any high-stakes decision. Prior assessments and attendance are the two strongest predictors in these estimates, and the year-to-year correlations for these variables are strong. Since SBAC testing begins in grade 3, third graders do not have prior SBAC scores on which their estimates can be based. Consequently, it is not surprising that the Grade 3 models have the highest degree of uncertainty (as evidenced in the comparison of estimate intervals in Table 2). Still, since they were built using a student's entire education record—including attendance, behavior, mobility, detailed special education data, standardized assessment scores, English learner status and family income status—the estimated Smarter Balanced scale scores for the 2019-20 academic year will provide teachers with a more complete picture from which to make more informed decisions regarding the timing, type, and target of interventions to implement.

For additional information on SBAC achievement levels and growth targets, districts are encouraged to consult the CSDE paper that describes the development of Connecticut's Growth Model for the Smarter Balanced summative assessments in ELA and mathematics. This paper includes tables that show the achievement level cut scores (see Page 5) and achievement level ranges and growth targets (see Page 9).

Conclusion

By incorporating cross-domain, longitudinal data and advanced modeling methodologies to estimate student scale scores for the 2019-20 SBAC assessments and set growth targets for 2020-21 SBAC assessments, CSDE is addressing assessment-related gaps. By sharing these data in the *EIT for Teachers*, CSDE is providing educators with a wealth of information on incoming students, informing instruction, and decreasing the amount of time spent on formal assessment upon the return to school.