The Development of Connecticut's Vertical Scale and Growth Model

The Case for a Connecticut Growth Model

The Connecticut Mastery Test (CMT) has been Connecticut's standardized test of essential skills in mathematics and language arts for more than 25 years. For the first several years, the tests provided snapshots of student performance in grades 4, 6 and 8. Grade level scale scores (range of 100-400) were developed for each grade and for each subject (math, reading and writing). For instance, test results revealed how students in grade 4 in 2003 performed (thumb down) in comparison with students who were in grade 4 the following year (thumb up)—two different groups of students.

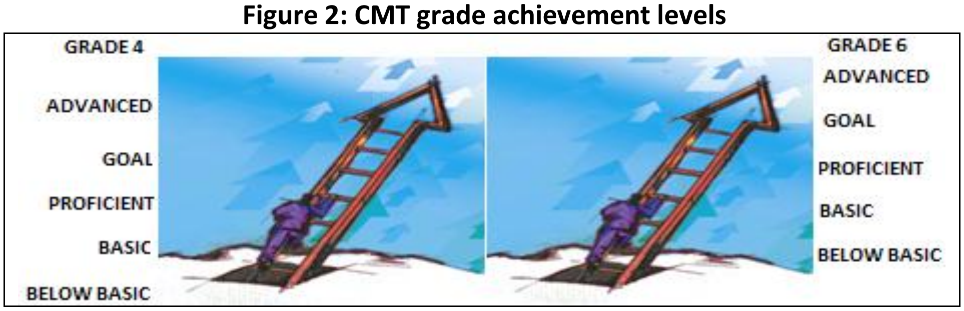
Figure 1: Grade level scale score interpretation

Grade 4: 2003  	Grade-Level Scale Scores = 100 - 400	**Grade 4: 2004**  

However, test results were not designed to measure how much grade 4 students in 2003 grew academically over the years from grade 4 up through the grades. It's true that a grade 5 test did not exist at that time. Even if one had existed, there was no system in place to compare performance of a single group of students from year to year.

The CMT achievement levels (Below Basic, Basic, Proficient, Goal and Advanced) place performance within each grade in a series of buckets. The buckets meant something different in each grade. For example, students in the Below Basic bucket in grade 4 mathematics receive a score of 100-186, and students in the Basic bucket in grade 6 receive a score of 191-214. Like grade-level scale scores, CMT achievement levels describe performance within grades and subjects, but do not facilitate monitoring progress across grades within a single achievement level. Therefore, a student may remain at the Basic level in grades 4 and 6, but progress (or lack of it) is difficult to identify because of the limitations of grade-level scores linked to grade achievement levels.

Figure 2: CMT grade achievement levels



The system in use left many educators asking "How can I tell if my students grew in achievement since last year?" Although the 2002 No Child Left Behind (NCLB) federal legislation was regarded by states with some trepidation, it became the catalyst for the answer to districts' question. In 2006, Connecticut began to administer NCLB-mandated annual testing in math and reading in grades 3 through 8. The new testing infrastructure made it possible to construct a reliable system to measure achievement progress in adjacent grades in two content areas: math and reading.

Growth Model Foundation: Connecticut's Vertical Scale

Like the previous three-grade (4, 6 and 8) configuration of the CMT, each test (grade 3-8) in the new CMT generation was somewhat different in length, difficulty and the content covered. However, the assessments were sufficiently similar in content to support two vertical scales—one for math and the other for reading.

In order to build the scales, a test was administered in 2007 in which students completed the tests for their assigned grades as well as a supplemental exam from one grade higher or one grade lower. For example, grade 5 students completed a supplemental test that contained either grade 4 or grade 6 test items. The supplemental tests were shorter than the exams from students' grade level (students took only one section of the test). As illustrated in Table 1, the diagonal fields represent the on-grade-level (OG) items at a given grade level, while the off-diagonal fields represent the supplemental tests (SU) administered to adjacent grades. The numerals represent students' grade levels and the grade levels of the tests taken. Therefore, SU56 means fifth-grade students took the supplemental exam of grade-six items. This common item and student design permitted statistical linking of performance vertically across grades.

Table 1: Common item and student design

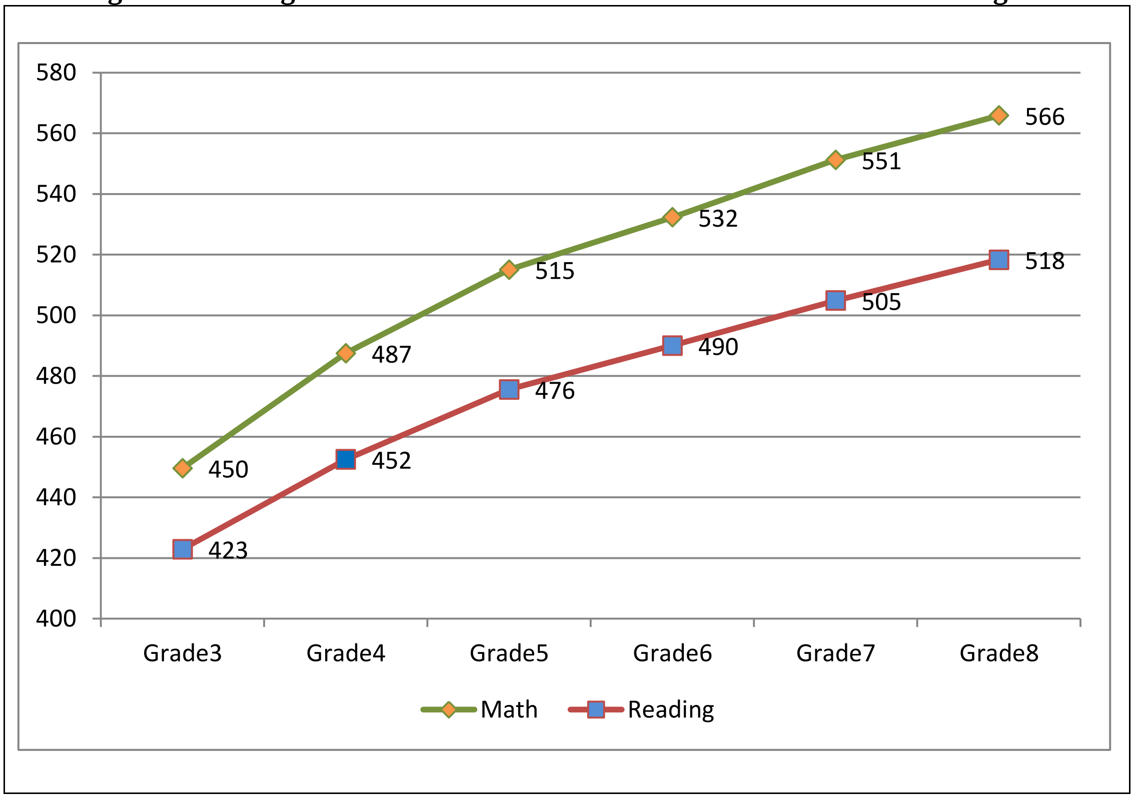
		Items					
		Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8
Students	Grade 3	OG33	SU34				
	Grade 4	SU43	OG44	SU45			
	Grade 5		SU54	OG55	SU56		
	Grade 6			SU65	OG66	SU67	
	Grade 7				SU76	OG77	SU78
	Grade 8					SU87	OG88

Using Item Response Theory, and in consultation with student assessment's Technical Advisory Committee, vertical equating procedures were performed to construct a single scale for mathematics grades 3-8 and a separate one for reading. The results were cross-validated by Hariharan Swaminathan and Jane Rogers of the University of Connecticut.

Characteristics of the Vertical Scale

The range for the math and reading scales is 200-700. Figure 3 displays a graphic depiction of the mean vertical scale scores for math and for reading. As the graphic illustrates the rate of growth changes somewhat from the lower to the upper grades, particularly in reading; the growth pattern suggests that the rate of growth at the lower grades is steeper than at the upper grades beginning with grade 6.

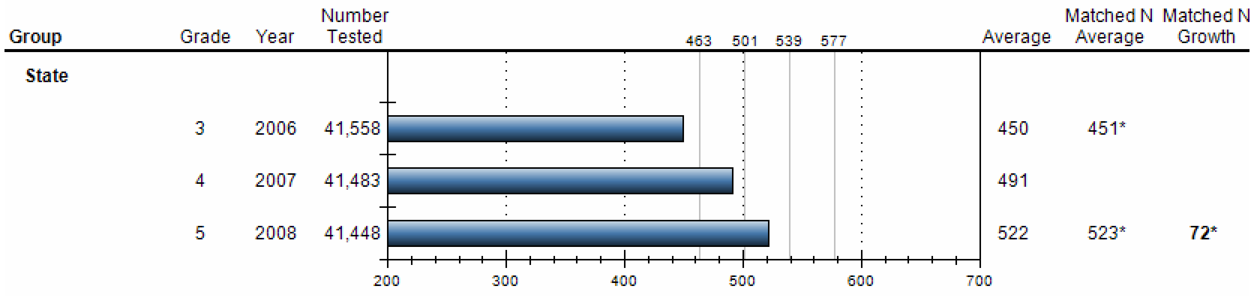
Figure 3: Average CMT vertical scale scores for mathematics and reading



With the availability of a vertical scale in math and reading, students' performance may be compared from grade to grade, beginning with scores attained in 2006. Consequently, a student whose vertical scale score in mathematics was 417 in grade 3 and 486 in grade 5 grew in achievement by 69 vertical scale score points. A grade 4 student with a vertical score in math of 462 and a score of 459 in grade 5 has not grown and has scored below the mean for grade 5 students. Therefore, the vertical scale makes it possible to assess whether or not students have improved in achievement from year to year and grade to grade.

With the construction of the vertical scales, the fundamentals were in place for monitoring growth and vertical scale scores were incorporated into the Department's online and paper reporting system. Beginning in 2007, test reports included students' vertical scale scores in math and reading and shortly thereafter, student and district summary reports became available online from ctreports.com (Figure 4).

Figure 4: Sample average vertical mathematics scale score report from ctreports.com



Constructing Connecticut's Growth Model

Although Connecticut's vertical scales made it possible to track student growth in achievement across grades, achievement growth interpretation issues remained. Those issues are framed below using questions frequently asked by school district personnel:

> How much are students likely to grow based on their previous achievement?
> How much student progress is good enough?
> How can I identify students/schools that are most at risk?
> How can I identify student/school success?

In effect, a growth model was needed that provided tools for improving instruction and improving schools. A model was subsequently constructed to predict the scores of students to:

> Monitor student growth in achievement
> Identify students at risk; and
> Set appropriate student achievement targets;

The model may also be used to make projections at the subgroup, class, school, or district level to:

> Identify groups of students who may be at risk based on mean projected achievement levels;
> Identify achievement targets for student groups based on average student performance targets.
> Evaluate success based on percentage of students achieving performance targets

Two tools were constructed to accomplish those goals:

> A CMT performance prediction model; and
> Achievement expectations and evaluation system.

Constructing the Performance Prediction Model: Method

The CMT performance prediction model was constructed by Swaminathan and Rogers of the University of Connecticut. They used four years of CMT vertical scale scores to construct a performance prediction model and to select five cohorts of students. Each cohort had scores from the same points in time and scores from a maximum of three time points were available for constructing a prediction model. For example, grade 8 scores for Cohort 1 could be projected from scores in grades 7, 6 and 5, a total of three years. Cohort 3 students' grade 6 scores could be predicted from two score predictors—those from grades 5 and 4.

Two prediction models were evaluated using two different statistical methods of analysis. Linear regression models assume a linear trajectory and are most accurate when applied to groups of students. Accordingly, a cohort's grade 8 scores may be projected using their grade 7 and 6 scores. Other variables (e.g., poverty status, genders, or ethnicity) may be included in the regression equation. Additionally, data from as few as a single time point might be used in the prediction model.

Growth models are constructed using mixed-model analysis. These analyses accommodate curvilinear projections, which lead to more accurate projections. In addition, growth projections may be obtained for individual students. However, this model has a major limitation; in order to model curvilinear growth, data from more than three time points are needed. Accordingly, projections can't be examined for curvilinearity unless they are projections for grades 7 or 8.

Regression models were fitted to each student cohort to predict cohort scores using only previous available scores. The number of scores varied from 1 to 3 prior scores. Subsequently, the regression models were refitted using achievement scores and background data (gender and poverty) as predictors.

Growth models were constructed using data from three prior time points (e.g. grades 3, 4, and 5) to project scores in the following grades. Growth models were also formulated using data from four time points (grades 3, 4, 5, and 6).

Regression models were devised for all five cohorts for reading and for math. Accordingly, regression models for cohorts entering grades 4-8 were completed. By contrast, linear growth models were possible for cohorts entering grades 6-7 for reading and math. Growth models were not possible for cohorts with scores from fewer than three time points.

Constructing the Performance Prediction Model: Results

Regression and growth models indicated that prior scores were good predictors of future performance, and therefore, could be used for dependable performance projections:

For both growth and regression models, predicted and actual scores were highly correlated, with correlations between predicted and actual scores ranging from .88 to .90 for reading in grades 6 and 7. Correlations were higher for math, ranging from .90 to .93 in grades 6 and 7. However, regression models produced higher correlations than growth models.

- The proportion of variance accounted for by the previous scores was high as well. For reading in grades 6 and 7, .78 to .81 of the variance was accounted for by the predictor scores. For math, .81 to .87 of the variance was accounted for by prior scores. Here again, regression model predictors accounted for more variance than growth models.
- The regression models remained stable regardless of the number of predictor scores available. Regression models produced correlations between predicted and actual scores that remained relatively high with even one predictor score for reading and for mathematics.

- Including additional demographic variables in the regression or growth model equations did not appear to enhance the prediction of the models appreciably. For example, in the regression model for grade 6 math, the proportion of variance accounted for increased by .001, with the addition of gender and poverty status to three prior vertical scale scores used as predictors. There were no changes in outcome with similar additions to the growth models for math and no changes to the results for either regression or growth models for reading.

- When scores from four time points are available for constructing growth models, they tend to out-perform regression models in their prediction accuracy. Fitting a quadratic growth model to grade 7 math resulted in .91 proportion of variance accounted for by prior scores. Analyses also suggest that reading performance may have a curvilinear trajectory; growth in achievement slows for students in higher grades.

- The study determined that reading is likely to benefit most from use of growth models. Nevertheless with growth model requirements of four data points only projections for grade 7 and 8 are possible. Accordingly, regression based student group performance projections were recommended for the performance prediction model.

Achievement Expectations and Evaluation System

Student group performance predictions were designed to identify at-risk students and schools needing instructional intervention. They also provided indicators of future performance if no substantive intervention occurred. The regression-based student performance prediction model provided the raw material for monitoring growth and risk of failure. By providing projected vertical scale scores in reading and math, districts and schools had the necessities to identify students at risk of failure. But did they have the essentials needed to evaluate success? Is it okay to assume the following statements are true?

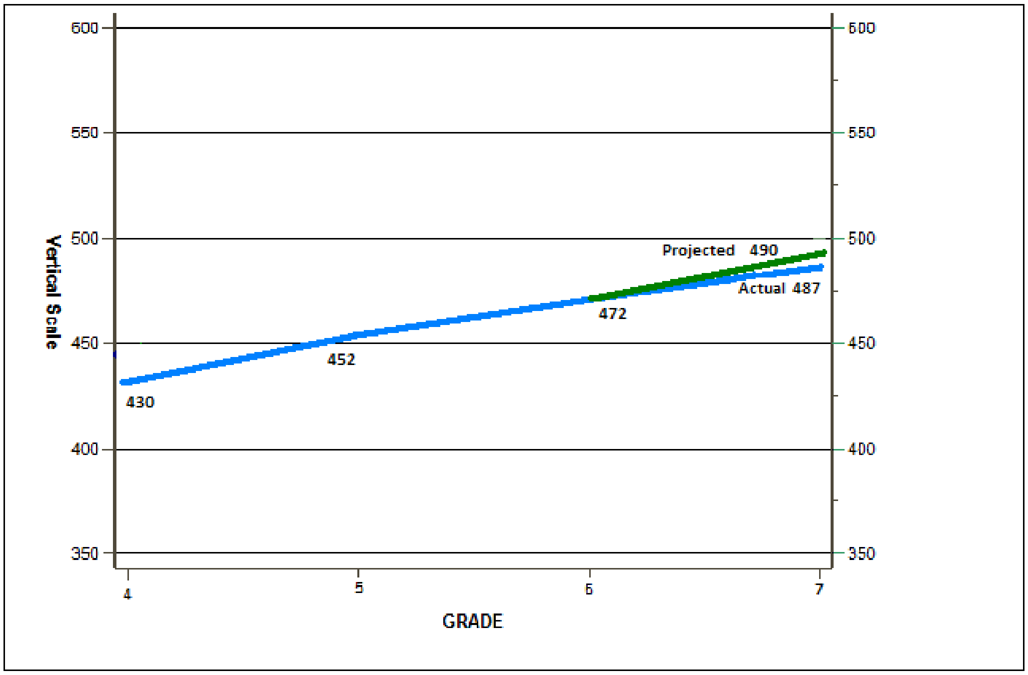
AVERAGE ACTUAL SCORE > PROJECTED SCORE → **SCHOOL SUCCESS?**

AVERAGE ACTUAL SCORE < PROJECTED SCORE → **SCHOOL UNSUCCESSFUL?**

As Figure 5 illustrates, achieving the projected score does not always mean that substantive achievement growth occurred. Achievement could be flat from one year to the next. Achievement could also increase, yet never be enough to help students rise to proficiency or to the state goal; the growth shown in Figure 5 would leave the student at the Basic level after three years. Accordingly districts sought help in not only monitoring achievement growth, but also in evaluating growth or lack of it; therefore, a key and recurrent district question has been how much growth is good enough?

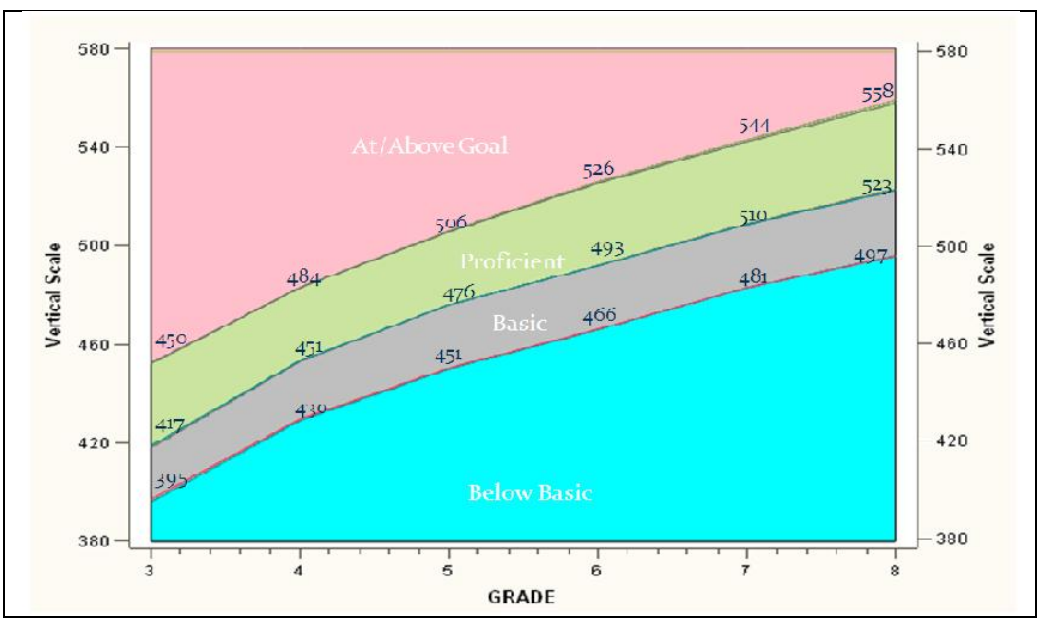
If students achieve, on average, a math vertical scale score of 452 that places them at the Basic Achievement Level in grade 5, and they achieve a higher average scale score of 472 in grade 6, but remain, overall, at the Basic level, have they made adequate progress for one year? What achievement growth attainments are reasonable indicators of successful instructional interventions and achievements?

Figure 5: Sample Actual and projected mathematics vertical scale scores



From the earliest years of statewide assessment in Connecticut, the Department has always asserted that every Connecticut student is expected to achieve the level of performance that has been termed the “state goal”. Districts’ questions suggested that they needed guidelines that would assist teachers and administrators in nudging students toward that goal and, at a minimum, help students who achieved goal level to maintain that level of achievement.

Figure 6: Math achievement level cut scores on the CMT vertical scale



To assist educators to move students toward the state goal, a three-step model was devised that combined the familiar CMT achievement levels with the vertical scales to help students move from one achievement level to the next within three years. In order to accomplish that, the cut points for CMT achievement levels (Figure 2) were located on the vertical scale for math and reading (Figure 3) rather than the grade- specific scale scores. By using the vertical scale, student achievement growth may be compared across grades using the achievement levels overlay on the vertical scale. Achievement level cut scores for mathematics on the vertical scale are displayed in Figure 6.

Using achievement level cut scores on the vertical scale, two equidistant interim cut points were then calculated for achievement level bands basic, proficient and goal. This produced three gain-score targets, low, mid-, and high scores for each of these CMT score bands. Tables 2 and 3 provide cut scores for reading as well as the gain score targets for the same subject.

Evaluation of the effect sizes of average gain score targets indicated that on average a gain score at or above 10 vertical scale score points, results in effect sizes at or above .3. This implies that a gain of 10 vertical score points from one administration to the next reflects a moderate effect size. Effect sizes increased as vertical gain scores increased.

Table 2: Minimum cut scores for grades 3-8: Reading

Grade	LOW BASIC	MID BASIC	HIGH BASIC	LOW PROF.	MID PROF.	HIGH PROF.	LOW GOAL	MID GOAL	HIGH GOAL	ADVANCED
3	381	387	394	401	409	417	425	442	460	478
4	412	417	422	428	435	442	449	468	487	506
5	434	438	443	448	454	460	466	485	504	524
6	440	445	450	456	462	468	475	496	517	539
7	454	459	464	470	475	481	487	507	528	549
8	466	471	476	481	487	493	500	520	541	561

Table 3: Sample one-year vertical scale gain score targets: Reading

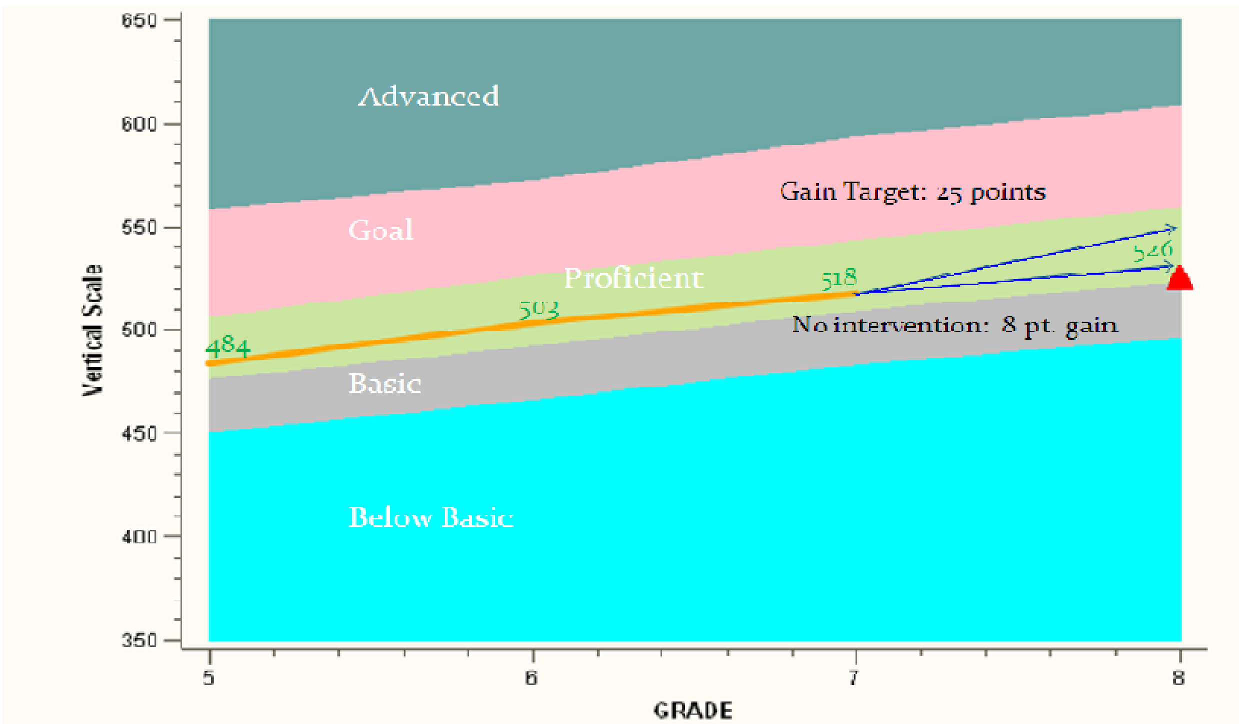
GRADE		CURRENT AND TARGET READING ACHIEVEMENT LEVELS AND GAIN SCORES									
CURRENT	TARGET	BELOW BASIC TO LOW BASIC	LOW TO MID BASIC	MID TO HIGH BASIC	HIGH BASIC TO LOW PROF.	LOW TO MID PROF.	MID TO HIGH PROF.	HIGH PROF. TO LOW GOAL	LOW TO MID GOAL	MID TO HIGH GOAL	HIGH GOAL TO ADVANCED
3	4	L. Basic/36	36	35	34	34	33	32	43	45	46
4	5	L. Basic/26	26	26	26	26	25	24	36	36	37
5	6	L. Basic/10	10	12	13	14	14	15	30	32	35
6	7	L. Basic/19	19	19	20	19	19	19	32	32	32
7	8	L. Basic/17	17	17	17	17	18	19	33	34	33

Target scores are designed to be set at the student level (e.g., for reading) using the vertical scale gain scores in Table 3. As an example, let's assume a student received a score of 430 in reading in grade 4. The student's score places him in the low-proficient score range for grade 4 (See Table 2). At the beginning of the following year when the student enters grade 5 (Target Grade in Table 3), he should be assigned a target gain score of a minimum of 26 vertical gain score points (Low to Mid Prof. column in Table 3)—for students currently in the low proficient range. Should he successfully achieve the target gain score, he should be at the mid-proficient level or higher in grade 5. Should the student achieve his grade 6 target gain score, he should be at the high proficient level. If the student fails to achieve his target gain score, then his target gain score for the subsequent year should be based appropriately on the score he received. The one-year target for students at the Below Basic level is to attain the Low Basic level or the target score indicated for their grade, whichever target is highest. Target scores may be aggregated at the classroom and school level for average classroom or school-level targets.

Target gain scores are also designed to be used in conjunction with score projections as displayed in the report prototype in Figure 7 for School Alpha, designed for the beginning of the school year. As illustrated, School Alpha's report provides the average math vertical scale scores students secured in prior grades (grades 5-7), the average score of 526 projected for the forthcoming test administered in grade 8 (indicated by the red triangle), and the aggregated gain score targets for students in grade 8 in that school—25 points. If students in grade 8 meet their targets, then the average vertical scale score for students in grade 8 would be 543, placing students, on average, at the high proficiency level. An end-of-year school report would be similar in design, but would additionally include the mean of grade 8 students' actual scores as well as whether or not School Alpha met its target.

Figure 7: Prototype mathematics school report

GRADE 8 MATH PROJECTION AND GAIN SCORE TARGET
SCHOOL ALPHA

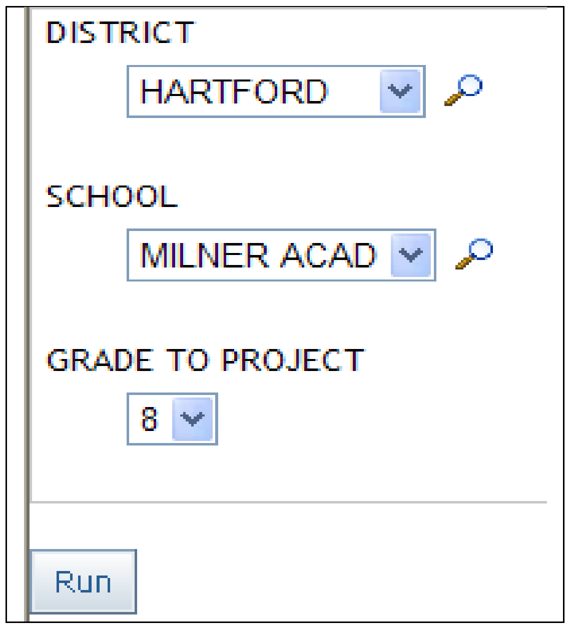


As explained earlier, the flexibility and accuracy of regression-based projections made them superior to multi-linear growth model projections. However, projections are limited to groups of students. How then might teachers plan for individual students? Although scores are not projected for individual students, setting targeted achievement goals are still appropriate. Therefore, educators may use gain score targets for math and reading to set expectations for each student achieving below the state goal. For students at/above the state goal, educators may set targets that will require students to sustain their goal level of achievement through the grades.

Interactive Web-Based Reports

In an attempt to meet district needs, a series of interactive web-based reports were developed to assist educators to monitor and evaluate student achievement growth. The score projection and gain score target-setting is the first of reports accessible through a SAS portal. The application package allows educators to select districts, schools, subgroups and grades of interest to retrieve student-group score projections and average group gain score targets. Additional components of the package provide options to drill-down to individual student-level reports they have permission to access. Figure 8 presents an example of the selection pane in the interactive reporting system.

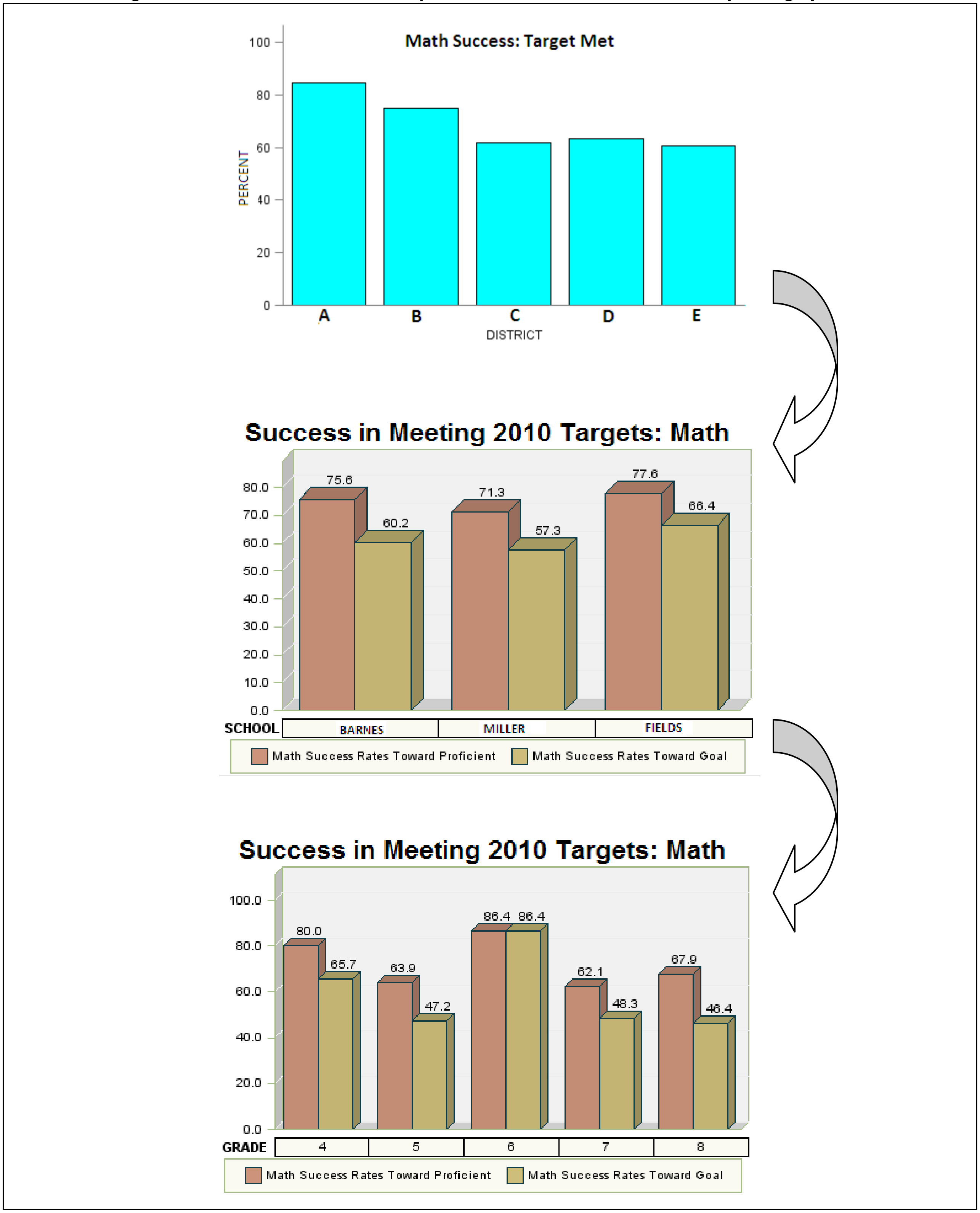
Figure 8: Section pane in interactive web-based reporting system



The score projection chart previously referred to (Figure 7) displays an example of the score projection chart included in the web-based report system. As the chart illustrates, projections in this system, are limited to group-level predictions. However, mathematics and reading gain score targets for individual students are available via a separate report.

Group-level success rates at achieving targeted gain scores are available in a third report. Success rates are available at various institutional levels via drill-through options. Figure 9 displays a sequence of reports of school success in achieving gain score targets. As Figure 9 illustrates, the reports are designed to allow the consumer to drill down from district-level success rate reports to grade-level success rate reports. Reports may also be developed to drill-down to the classroom-level or other subgroups as desired.

Figure 9: Success rate drillable reports in interactive web-based reporting system



Final Comments

The score projection and growth model systems were intended to provide educators with additional tools to assist them in student achievement and instructional improvement. It is not a part of Connecticut's school accountability system. It is part of the Department's ongoing efforts to be responsive to school districts and to provide technical assistance to help them accomplish their goals.

The growth model also represents part of a national trend reflected in the Race to the Top (RTTT) grants and other initiatives, to supplement the status model of measuring achievement stressed in NCLB. Rather than focus on what different groups of students accomplish each year, growth models are intended to emphasize the same student's or groups of students' growth in achievement over time. Our growth model system, like others used around our nation, is designed to help schools use assessments as tools to monitor and enhance students' attainment across the grades.

This research bulletin was prepared by Norma Sinclair and Mohamed Dirir. For Further information contact Norma Sinclair at the Connecticut State Department of Education, P.O. Box 2219, or by email at norma.sinclair@ct.gov.