Connecticut Mastery Test
FOURTH GENERATION

Data Analysis Guide

Summer 2009

Connecticut State Department of Education
Bureau of Student Assessment

Purpose

This guide provides Connecticut Mastery Test (CMT) data users with interpretation rules to consider when analyzing CMT data. Data users may include educators, the general public, and members of the media. The rules apply to the *most commonly used data elements* when analyzing CMT scores; these elements include strand scores, performance level scores, and vertical scale scores. This guide also provides information about comparing scores between generations of the CMT.

Each year CMT score reports are sent to school districts. Additionally, scores are available on the Connecticut CMT Online Reports Web site (www.ctreports.com). School, district and state results are available on the Public Summary Performance Reports portion of that Web site. School district personnel who have a username and password can also access the Individual Student Performance Reports Web site.

Contact the CMT Student Assessment Office at (860) 713-6860 for further guidance on how to interpret CMT scores.

Table of Contents

Comparing Scores between Generation Three and Generation Four

The Third Generation CMT (CMT3) was last administered in the fall of 2004. The Fourth Generation CMT (CMT4) was first administered in March 2006. Direct comparisons of performance level scores and strand scores should **not** be made across generations of the CMT because of changes to the census test made during CMT4 development.

The table below summarizes the primary changes:

Change between Generations	CMT3	CMT4
Test Administration	September/October	March
Test Booklets	Separate test and answer booklets	Consumable – students record responses to questions directly in test booklets
Reading Comprehension	One 75-minute test session with 3 reading passages	Two 45-minute test sessions with 2 reading passages in each session
Degrees of Reading Power®	Administration time of 70 or 75 minutes (depending on grade)	Administration time of 45 minutes for all grades
Editing & Revising	Editing strands assessed in separate passages from composing/revising strands	Each passage assesses both editing and composing/revising strands
Mathematics	Three strand 25 items in Grades 5 through 8	Two Strand 25 items in Grades 5 through 8
Reading, Writing, Mathematics Open-ended Items		Changes to the scoring criteria for open-ended items

The 2006 test data should be considered a benchmark year for CMT4. Comparing data from 2006 to 2009 is one recommended method when looking at single-grade data longitudinally. CMT4 is tentatively scheduled to remain in place until 2014 so this benchmark should remain for several years as a reference point.

Analyzing Strand Scores

Within the Mathematics, Reading Comprehension, and Editing & Revising tests, scores are reported by strand. The strands reported in these content areas are listed below.

Mathematics Strands
1. Place Value
2. Pictorial Representations of Numbers
3. Equivalent Fractions, Decimals and Percents
4. Order, Magnitude and Rounding of Numbers
5. Models for Operations
6. Basic Facts
7. Computation with Whole Numbers and Decimals
8. Computation with Fractions and Integers
9. Solve Word Problems
10. Numerical Estimation Strategies
11. Estimating Solutions to Problems
12. Ratios and Proportions
13. Computation with Percents
14. Time
15. Approximating Measures
16. Customary and Metric Measures
17. Geometric Shapes and Properties
18. Spatial Relationships
19. Tables, Graphs and Charts
20. Statistics and Data Analysis
21. Probability
22. Patterns
23. Algebraic Concepts
24. Classification and Logical Reasoning
25. Mathematical Applications

Reading Comprehension Strands
A. Forming a General Understanding
B. Developing Interpretation
C. Making Reader/Text Connections
D. Examining the Content and Structure

Editing & Revising Strands
1. Composing/Revising
2. Editing

Strand scores are useful when analyzing data about discrete parts of the content area being assessed. For example, the Place Value strand in mathematics is tested every year from Grades 3 to 8. Analyzing these scores is helpful in determining the status of student understanding in this area. Each strand has a criterion that is the standard for "mastery." Comparison of strand level performance can be made by comparing the number or percentage of students who are at the mastery level.

The tables below outline the raw score points necessary for mastery in each strand on the Mathematics, Editing & Revising, and Reading Comprehension tests. Student performance on the CMT Science test is reported as raw score points by content strand and dimension. There are no mastery criteria established for Science.

Strand Mastery Criteria

Mathematics	**Grade**					
Content Strand	**3**	**4**	**5**	**6**	**7**	**8**
Total Strands	18	21	23	23	23	21
	3 of 4	3 of 4	3 of 4	3 of 4	3 of 4	3 of 4
	4 of 6	4 of 6	4 of 6	4 of 6	4 of 6	4 of 6
	6 of 8	6 of 8	6 of 8	6 of 8	6 of 8	6 of 8
				7 of 10	7 of 10	7 of 10
						8 of 12

Editing & Revising	**Grade**					
Content Strand	**3**	**4**	**5**	**6**	**7**	**8**
Composing/Revising	12 of 16	12 of 16	13 of 18	13 of 18	15 of 20	15 of 20
Editing	12 of 16	12 of 16	13 of 18	13 of 18	15 of 20	15 of 20

Reading Comprehension	**Grade**					
Content Strand	**3**	**4**	**5**	**6**	**7**	**8**
Forming a General Understanding	10 of 14	8 of 12	7 of 11	7 of 11	7 of 10	7 of 10
Developing Interpretation	6 of 9	7 of 10	7 of 10	7 of 10	7 of 10	7 of 10
Making Reader/Text Connections	6 of 8	6 of 8	6 of 8	6 of 8	6 of 8	6 of 8
Examining the Content and Structure	6 of 9	7 of 10	7 of 11	7 of 11	8 of 12	8 of 12

Analyzing Strand Scores Across Years within the Same Grade

An analysis of strand scores across years within a grade may be used to gauge student performance. Within the same test form, the overall strand difficulty remains constant because the same test items are used in multiple years.

For example, in 2008 and 2009 the same test items were used to assess students in a specific strand because Form S was used both years. Differences in performance could be attributed to true differences in student performance as opposed to minor strand difficultly differences across different test forms.

Caution should be used when analyzing strand scores across years. Strand difficulties are pre-equated across the forms, but there is no post equating to account for any possible differences in actual item performance. For example, between 2007 and 2008, different test forms were administered. In this case, there may be changes in strand performance because of true student performance differences. However, this change could also be attributed to minor differences in the overall difficulty of the strands between test forms. Comparing district results to state results is helpful when making decisions about trends in student performance. For example, if a district's performance on a particular strand improved significantly, but state-level performance remained relatively the same, the district's gain may be attributed to true student performance differences. If, on the other hand, the state numbers kept pace, it may be due to a difference in test forms.

Analyzing Strand Scores Across Years and Grades

Analyzing strand scores across grades will result in **invalid** data analysis. The overall test design of CMT4 does not include equated strands across grades. This means that the average difficulty of one strand in a grade does not equal the average difficulty of the same strand in other grades. For example, the percentage of students achieving mastery in Place Value in Grade 3 would have no relationship to the percentage of students achieving mastery in Place Value in Grade 4 (or any other grade) because the difficulties were not made equivalent.

Analyzing Performance Level Scores

There are five performance levels reported for each content area. The performance levels are Below Basic, Basic, Proficient, Goal, and Advanced. The number and percentage of students at each level can be compared as an indicator of performance. For example, looking at the number and/or percentage of students at the goal level in one year versus another year is a method for analysis. However, there are some important characteristics of the performance levels that should be considered when analyzing the score band data.

Analyzing Performance Level Scores Across Years within the Same Grade

Mathematics, science, reading, and writing performance level results can be compared across years within the same grade. For example, Grade 4 mathematics performance levels in 2006 can be compared to Grade 4 mathematics performance levels in 2007. Since the grade level test forms are equated in difficulty across years and represent the same "measuring stick," it is valid to use this technique when comparing results across years. The amount of achievement required for a student to reach a particular performance level in a specific grade and content area is equal across the years within the generation, regardless of which test form is used.

Performance level results may also be compared across years within the same grade at the subgroup level (i.e., English language learner status, special education status, gender status, free/reduced meal status, or ethnic background status). For example, the performance of Grade 6 girls in writing can be compared to the performance of Grade 6 boys in writing across years.

Analyzing Performance Level Scores Across Grades (Within a Year and Across Years)

Making comparisons of performance levels across grades within a content area will result in **invalid** data analysis**.** For example, one cannot legitimately compare Grade 7 performance in the goal range in 2007 to Grade 8 performance in the goal range in 2007. In addition, one cannot legitimately compare Grade 7 performance in the goal range in 2007 to Grade 8 performance in the goal range in 2008. This will result in invalid data analysis because the standards across the grades are not identical. For example, the established goal range in Grade 7 mathematics is not the same as the goal range in Grade 8 mathematics. Although the scale score ranges for each performance level are similar, a specific scale score in one grade is not equivalent to the same scale score in another grade. When analyzing data from the Performance Level Summary, Longitudinal Roster, Matched Comparison, and/or Cohort Comparison reports be aware that differences in performance across grades may be the result of different standards between the grades. A better use of the data would be to use the data comparatively. For example, it is appropriate to compare Grade 7 performance in the goal range in 2007 and Grade 8 performance in the goal range in 2008 for one group compared to another group. This analysis can be used to compare student groups, schools, and districts to each other and/or to state data. Note: Page 12 presents information about the vertical scales which is an even better tool to look at longitudinal data.

Analyzing Performance Level Scores Across Content Areas

Making comparisons of performance levels across content areas will result in **invalid** data analysis. For example, one cannot legitimately compare Grade 6 performance in the goal range in reading and Grade 6 performance in the goal range in mathematics. The performance levels were not equated across content areas. These performance levels represent different standards, making it inappropriate to compare across the content areas.

Analyzing Performance Level Scores Compared to Content Strand Mastery

Generally, students who score at the higher performance levels (Goal and Advanced) master more content strands than students in the lower performance levels (Proficient, Basic and Below Basic). However, it is possible for two students to master the same number of strands but score within different performance levels. It is also possible for two students with the same scale score to master a different number of content strands.

A student's performance level is determined by the scale score for each content area. Scale scores are calculated from the student's total raw score (the number of raw score points across all strands). Refer to the CMT Technical Bulletin for detailed information about converting raw scores to scale scores for each content area.

As the following table illustrates, two students can master the same number of mathematics strands, yet be reported in different performance levels. Students A and B mastered every Grade 8 strand, however, Student A met the minimum mastery criteria for each content strand while Student B met the maximum number of raw score points for each content strand. The raw scores for these students are not the same, which results in different scale scores and performance levels.

Note: Not all 25 mathematics strands are tested at Grade 8.

Mathematics Strand Mastery vs. Scale Score and Performance Level

Strand	Mastery Criteria	Student A	Student B
1	3 of 4	3	4
3	4 of 6	4	6
4	4 of 6	4	6
5	4 of 6	4	6
7	4 of 6	4	6
8	4 of 6	4	6
9	6 of 8	6	8
11	6 of 8	6	8
12	6 of 8	6	8
13	4 of 6	4	6
15	4 of 6	4	6
16	6 of 8	6	8
17	6 of 8	6	8
18	8 of 12	8	12
17	4 of 6	4	6
20	6 of 8	6	8
21	4 of 6	4	6
22	4 of 6	4	6
23	4 of 6	4	6
24	7 of 10	7	10
25	4 of 6	4	6
Content Strands Mastered		21	21
Total Raw Score Points		102	146
Scale Score		257	400
Performance Level		Goal	Advanced

As the following table illustrates, two students with the same raw scores, scale scores and performance levels may master different numbers of content strands. For example, Student A met the minimum mastery criteria for each Grade 8 mathematics strand for a total of 102 raw score points. Student B achieved the same number of raw score points while mastering about half of the mathematics strands.

Mathematics Strand Mastery vs. Scale Score and Performance Level

Strand	Mastery Criteria	Student A	Student B
1	3 of 4	3	4
3	4 of 6	4	6
4	4 of 6	4	5
5	4 of 6	4	6
7	4 of 6	4	0
8	4 of 6	4	0
9	6 of 8	6	8
11	6 of 8	6	8
12	6 of 8	6	8
13	4 of 6	4	2
15	4 of 6	4	2
16	6 of 8	6	8
17	6 of 8	6	8
18	8 of 12	8	12
17	4 of 6	4	3
20	6 of 8	6	8
21	4 of 6	4	2
22	4 of 6	4	0
23	4 of 6	4	0
24	7 of 10	7	10
25	4 of 6	4	2
Content Strands Mastered		21	12
Total Raw Score Points		102	102
Scale Score		257	257
Performance Level		Goal	Goal

Analyzing Vertical Scale Scores

The CMT vertical scales are designed to measure growth (or change) across grades (i.e., from Grade 3 to Grade 4, from Grade 4 to Grade 5, etc.) on tests that have different characteristics and items but have similar content. Vertical scales have been established in the content areas of mathematics and reading. The vertical scales were constructed so that each vertical scale score represents the same theoretical achievement level whether derived from a Grade 3, Grade 4, Grade 5, Grade 6, Grade 7, or Grade 8 CMT scale score. Each grade-level CMT scale score (range 100 - 400) in mathematics or reading corresponds to a specific value on a common mathematics or reading vertical scale score (range 200 - 700). Thus, students in different grades taking different tests can have the same vertical scale score representing the same level of achievement defined by the vertical scale. This vertical scale score allows for valid interpretations of growth across time using tests differing in content, length, and item difficulty.

The vertical scales are not meant to replace the usual year-to-year comparisons based on the percentage of students scoring at each achievement level. Instead, the vertical scales are intended to enhance the CMT program by providing a means to assess achievement growth across grades Every Grade 4 through 8 student who has a valid CMT grade-level score in mathematics and/or reading for at least two years also has a corresponding valid vertical scale score in mathematics and/or reading. Although there are now four years worth of vertical scale data, the state recommends that analysis be limited to three consecutive years.

Vertical scale scores (like all other CMT scores) are based on the performance of individual students on the day of testing. When interpreting growth, care should be taken not to base important educational decisions solely on vertical scale results. CMT results can best be used in conjunction with classroom assessments and classroom work to identify potential strengths and needs of students in the content areas assessed.

Interpretation of Vertical Scale Results

Growth by an individual student from one year to another is defined as:

(Vertical Scale Score Year 2) minus (Vertical Scale Score Year 1)

Growth for groups of students from one year to another is defined as:

(Mean Vertical Scale Score Year 2) minus (Mean Vertical Scale Score Year 1)

Growth over time at the individual level or group level can be positive or negative. Growth is positive when the vertical scale score is larger in Year 2 than in Year 1. Negative growth occurs when the vertical scale score is smaller in Year 2 than in Year 1.

Note: While "negative growth" is uncommon overall, it is less likely at the group level than at the individual student level because group mean scores are more stable than individual student scores.

One way to use the vertical scale results is through a normative interpretation. With normative comparisons, the growth of an individual or group of students is viewed in terms of a larger, representative reference group. For example, the growth within a particular school over time may

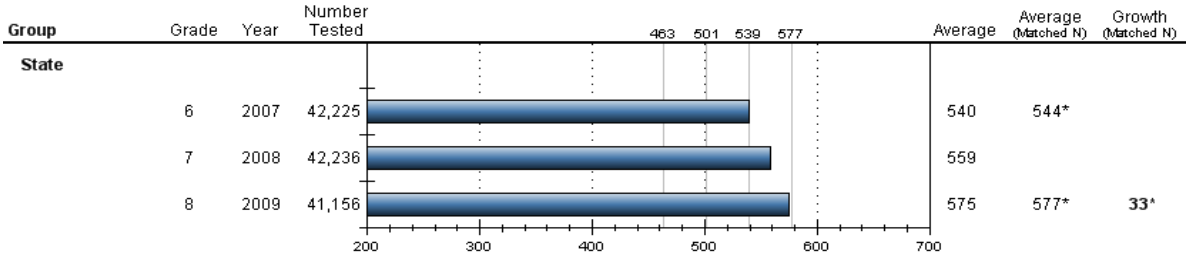
be compared to the growth at the district and state levels over the same time period to determine the school's relative performance in relation to the district and state groups.

The reading and mathematics vertical scales are divided into five Levels of Understanding (Stages I - V), with a description of typical performance for each level as shown in Charts 1 and 2. Each of the five Levels of Understanding corresponds to content-referenced descriptions of student performance. The vertical scale descriptors do not describe grade-level performance. The descriptors span the six grades and are different from the grade specific Performance Level Descriptors (PLD) that describe the five performance levels (i.e., Below Basic, Basic, Proficient, Goal, and Advanced) for each content area. The vertical scale Levels of Understanding do not represent performance standards or categories of student expectations but rather are meant to provide a general (not grade-level specific) description of what typical students know and can do within certain score ranges or levels of understanding across the vertical scales.

Suppose, for example, one wants to measure the reading growth from Grade 5 to Grade 6 for a matched group of students. Suppose further that the mean reading vertical scale score for the group is 480 in 2006 (Grade 5) and 500 in 2007 (Grade 6). Using Chart 2, the performance of the Grade 5 group in 2006 is similar to students in the upper part of Stage III, while one year later the group performs more like students in the lower part of Stage IV.

Below is a sample report of the Vertical Scale Score Analysis Report from the Connecticut CMT Online Reports Web site (www.ctreports.com). Understanding how the fields are calculated will be helpful when analyzing the data. The "Average" column represents the average vertical scale score of all students who had a valid score in the defined year. The "Matched N Average" column represents the average vertical scale score of only those students who had a score in the first and last years. Finally, the "Matched N Growth" column represents the change in the average of those students who were matched in the first and last years.

Average Vertical Scale Score: Mathematics



When generating growth reports based on the matched groups, each student's demographic/student information group identification is determined by the student's status in the most recent year. For example, when generating a report for ELL students, a student reported as a regular education student in the first year who is identified as an ELL student in the second year would be considered a member of the ELL subgroup for both years in this report.

For additional information about the CMT Vertical Scales, please refer to the *CMT Vertical Scales 2009 Interpretive Guide* and *CMT Vertical Scales 2009 Frequently Asked Questions*.

Chart 1

Connecticut Mastery Test Vertical Scale Descriptors

Mathematics

Level of Understanding	Description	Vertical Scale Score
Stage V	Generally, students who perform at this level demonstrate *extensive* knowledge of all assessed content areas.	578+
Stage IV	Generally, students who perform at this level demonstrate *extensive* knowledge of most of the assessed content areas. These students demonstrate adequate knowledge of some of the assessed content areas including, measurement, ratio, proportion and percent.	540 – 577
Stage III	Generally, students who perform at this level demonstrate *adequate* knowledge of most and *extensive* knowledge of some of the assessed content areas. These students demonstrate limited knowledge of estimating solutions to problems, measurement, ratio, proportion and percent.	502 - 539
Stage II	Generally, students who perform at this level demonstrate *adequate* knowledge of some of the assessed content areas. These students demonstrate *limited* knowledge of rational number concepts and computation, estimation, and measurement.	464 - 501
Stage I	Generally, students who perform at this level demonstrate *limited* knowledge of most of the assessed content areas. These students demonstrate adequate knowledge of basic number concepts and computation skills.	Less than 464

Chart 2

Connecticut Mastery Test Vertical Scale Descriptors

Reading

Level of Understanding	Description	Vertical Scale Score
Stage V	Students who perform at this level are likely to demonstrate an *exceptional* ability to read and respond to informational and literary texts without assistance. Students at this level effectively use sophisticated strategies before, during and after reading to understand, interpret and evaluate text.	529+
Stage IV	Students who perform at this level are likely to demonstrate a *consistent* ability to read and respond to informational and literary texts with minimal assistance. Students at this level use strategies effectively before, during, and after reading to understand, interpret and evaluate text.	495 - 528
Stage III	Students who perform at this level are likely to demonstrate an *adequate* ability to read and respond to informational and literary texts with some assistance. Students at this level use some strategies effectively before, during, and after reading to understand, interpret and evaluate text.	460 - 494
Stage II	Students who perform at this level are likely to demonstrate a *limited* ability to read and respond to informational and literary texts, and require assistance to complete many reading tasks. Students at this level use some strategies inconsistently before, during and after reading to understand and interpret text.	425 - 459
Stage I	Students who perform at this level are likely to demonstrate a *very limited* ability to read and respond to informational and literary texts and require significant assistance to complete most reading tasks. Students at this level are not able to use strategies to understand and interpret text.	Less than 425