Determining Final Effectiveness Ratings Using the Colorado State Model Evaluation System for Teachers
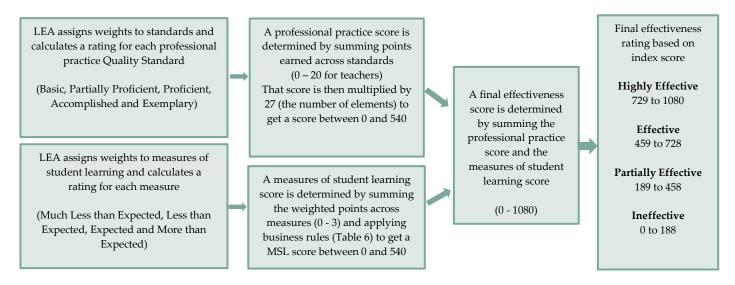
The Colorado State Model Approach and Method for Combining Professional Practices and Measures of Student Learning

Senate Bill 10-191, the Great Teachers and Leaders Act, requires that 50 percent of an educator's evaluation be based on professional practices and 50 percent be based on multiple measures of student learning. Educators will earn a professional practice score based on the rubric and a measures of student learning score based on multiple measures. The professional practices score and the measures of student learning scores are combined to determine an overall effectiveness rating of Ineffective, Partially Effective, Effective or Highly Effective.

There are several approaches and methods for combining these measures to arrive at a final effectiveness rating for educators (see Appendix C [i]). This document provides information on the approach and method used in the Colorado State Model Educator Evaluation System (state model).

The state model uses an additive approach expressed through an index score to arrive at a final effectiveness score. Figure 1 illustrates the series of steps taken to move from scores earned on each component to a final effectiveness score and rating. The process of combining measures starts with the final scores from professional practices and the measures of student learning. Once the professional practice score and measures of student learning score are determined, they are added together to create a single effectiveness, or index score. A final effectiveness rating is assigned to an educator based on the total number of points reported.

Figure 1. Process for Assigning Effectiveness Ratings for teachers

LEA assigns weights to standards and calculates a rating for each professional practice Quality Standard (Basic, Partially Proficient, Proficient, Accomplished and Exemplary)	A professional practice score is determined by summing points earned across standards (0 – 20 for teachers) That score is then multiplied by 27 (the number of elements) to get a score between 0 and 540	A final effectiveness score is determined by summing the professional practice score and the measures of student learning score (0 - 1080)	Final effectiveness rating based on index score **Highly Effective** 729 to 1080 **Effective** 459 to 728 **Partially Effective** 189 to 458 **Ineffective** 0 to 188
LEA assigns weights to measures of student learning and calculates a rating for each measure (Much Less than Expected, Less than Expected, Expected and More than Expected)	A measures of student learning score is determined by summing the weighted points across measures (0 - 3) and applying business rules (Table 6) to get a MSL score between 0 and 540		

Understanding Professional Practices Ratings

In the Colorado State Model Evaluation System for teachers, five professional practice Quality Standards are evaluated using a rubric. These standards are content, environment, instruction, reflection and leadership. Each of

the Quality Standards has a varied number of associated elements that are scored individually to evaluate each professional practice Quality Standard. To view the professional practice Quality Standards for teachers and their 27 elements, click here.

Table 1 shows the point values assigned to each element in the Colorado State Model Evaluation System. Each educator earns a professional practice rating based on the accumulation of points on the 27 elements of the model system rubric. Appendix A provides an example of how points earned on each standard are rolled up to an overall professional practice score and rating.   In addition, the vertical axis or y-axis in Figure 2 shows how the 540 point scale rating for the professional practices component is divided into segments that correspond to five different performance ratings.

Table 1: Point Value of Professional Practices Ratings for Teachers

Professional Practices Ratings	Point Value per Rating	Professional Practices Rubric Scale 0-20 (Rounded to the nearest hundredth)	Scale When Converted to 0-540 (Rounded to the nearest whole number)
Exemplary	4	17.01 to 20	460 to 540
Accomplished	3	12.01 to 17	325 to 459
Proficient	2	7.01 to 12	190 to 324
Partially Proficient	1	2.01 to 7	55 to 189
Basic	0	0 to 2	0 to 54

Understanding Measures of Student Learning Ratings

Local school districts identify the different measures of student learning comprising an educator's body of evidence for the 50 percent measures of student learning portion of their evaluation. Districts determine the best approach for combining these measures. In the state model, each measure is awarded points that range from zero to three. Table 2 shows how the point values correspond to measures of student learning ratings. The measures of student learning are weighted and combined and then converted to a score between zero and 540. For more specific information on how a score between zero and 540 is obtained on this component, please refer to Appendix B or Step 6 in the Measures of Student Learning Guidance document. The horizontal axis or x-axis in Figure 2 shows how the 540 point scale rating for the measures of student learning component is divided into segments that correspond to four different ratings.

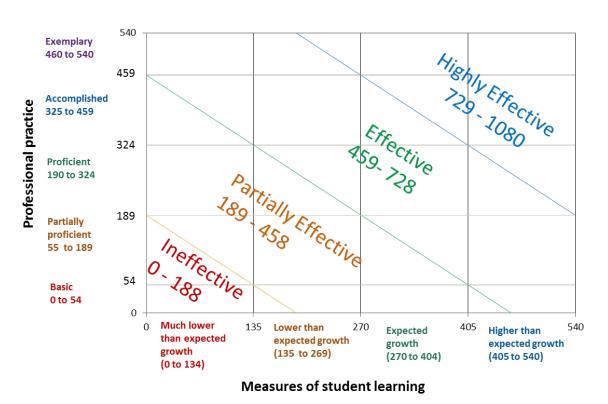**Table 2: Point Values for Measures of Student Learning**

Measure of Student Learning Rating	Point Value per Measure	Cut Point Scores
More than Expected	3	405 to 540
Expected	2	270 to 404
Less than Expected	1	135 to 269
Much less than Expected	0	0 to 134

Determining a Final Effectiveness Score and Rating:

Figure 2. Illustrating Three Cut Points Considered in the Colorado State Model Evaluation System

Teacher's cut points for the Colorado Model Evaluation System

To arrive at the final educator effectiveness score, the professional practice score is simply added to the measures of student learning score. Translating the final effectiveness score into a rating entails locating the score earned in the range of scores presented in Table 3. The next section explains how the cut points in Table 3 were established.

Table 3: Cut Points Set for Teacher Effectiveness Ratings

Rating Category	Ineffective	Partially Effective	Effective	Highly Effective
Cut points	0 to 188	189 to 458	459 to 728	729 to 1080

Setting the Cut Points for Measures of Student Learning

The graph in Figure 2 shows the professional practices and measures of student learning cut points. The vertical axis or y-axis displays the professional practices scale of 540 points and is divided into five sections of varying points each. Moving from the bottom of the y-axis to the top, each of these sections corresponds to a professional practices rating of Basic, Partially Proficient, Proficient, Accomplished or Exemplary. The horizontal axis or x-axis, which displays the

measures of student learning scale of 540 points, is divided into four sections of 135 points each. Moving from left to right along the x-axis, each of these four sections corresponds to ratings of Much Lower than Expected Growth, Lower than Expected Growth, Expected Growth or Higher than Expected Growth. The third set of cut points considered are in Table 3.

The first cut point is established by determining the maximum score for Basic on professional practices (54) and the minimum score for Lower than Expected Growth on measures of student learning (135). With 54 + 135 = 189, 189 is the first cut point for a partially effective rating. To determine the cut point for Effective, the maximum score for Partially Proficient on professional practices (189) is added to the minimum score for Expected Growth on the measures of student learning (270). The cut point for an Effective rating is 459 (189+270). The cut point for Highly Effective is determined by adding the maximum score for Proficient on the professional practices (324) to the minimum score for More than Expected Growth on the measures of student learning (405). The cut point for a Highly Effective rating is 729 (324+405). An educator's final effectiveness rating is determined after the professional practice score and measures of student learning score have been combined. For example, an educator will earn an Effective rating if his/her combined scores are between 459 and 728.

Summary

The Colorado State Model Evaluation System is structured to ensure that professional practices and measures of student learning are equally represented in the determination of a final effectiveness rating. The index approach was selected to clearly demonstrate that an equal number of points are being distributed and combined across the two sides of the system. Each of the two components represents 50 percent of a teacher's final evaluation.

Each component of the Colorado State Model Evaluation System has specific cut points to determine the professional practice and measures of student learning ratings. Districts using the state model system will use the cut points established for the combined scores to assign one of the following four final effectiveness ratings to educators: Ineffective, Partially Effective, Effective and Highly Effective.

The state will use the 2013-14 year to study assigning final effectiveness ratings using this approach. Revisions may be made to the approach based on what is learned from the first year of implementation.

Appendix A. Determining the Final Measures of Student Learning Score

Note: This information is from Step 6 in the [Measures of Student Learning Guidance document](). While it is included in this document to provide additional context, it is recommended that the Measures of Student Learning Guidance document be reviewed in its entirety.

By assigning weights to each score associated with the multiple measures in educator evaluations, districts are signaling which results or measures in the system are deemed to have more value than others, are better aligned with learning goals, are more appropriate for measuring educator impact or may signal that all results should be weighted equally. After each of the measures of student learning are scaled (e.g., on a zero-three scale), the next step would entail assigning weights to each and applying an approach to calculate a total score earned by teachers on measures of student learning. Districts may wish to preliminarily weight the results from each measure as it is selected at the beginning of the school year. Districts are encouraged to continuously evaluate the impact of weighting decisions and revise as needed.

Although districts can decide how to weight the scores from each of the multiple measures, districts may want to keep things simple by selecting weighting percentages that sum up to 100 percent. Multiplying the scores earned by the assigned weight yields the weighted score for each measure. The composite score in this example represents a compensatory approach, which was selected as a design choice to ensure that each measure included in an educator's body of evidence can have a measureable influence on the student learning score. Table 4 provides an illustration of how districts *may* consider distributing the weights assigned to each score for their teachers, and how a single index score is computed.

Table 4: Weighting and Combining Scores Example

Measures/Results from Colorado Growth Model and Student Learning Objectives (SLO)	Score Earned (Expected Growth)	Weight Assigned	Weighted Score
TCAP Reading MGP (collective school)	2 (typical)	.15	.30
TCAP Writing MGP (collective school)	2 (typical)	.15	.30
SLO 1 Results (collective grade level reading)	2	.35	.70
SLO 2 Results (individual teacher)	1	.35	.35
Sum of Weights		**1**	**1.65**

In this example, the assumption is made that the district has agreed to attribute Colorado Growth Model results from reading and writing (total of six points possible) to all teachers in the school. Further, Table 4 illustrates that all teachers will have two additional measures based on targets yielding two scores (total of six points possible) for attainment of expected targets. The first column is the measure that is included. The second column reflects the rating earned - Much Less than Expected (zero points), Less than Expected (one point), Expected (two points) and More than Expected (three points) - by a hypothetical teacher with all these measures relevant to his/her goals.

To assign weights to scores, a district can allocate smaller or higher percentages to each rating and ensure that the weights assigned across all measures sum up to 1 or a 100 percent as shown in the third column. In this example, the district has decided that each of the results from their SLO targets and the set of combined TCAP growth results should have about the same weight. The third column shows that each SLO result has a weight of .35 and the set of combined TCAP growth scores has a total weight of .30. The fourth column shows the weighted scores. These are computed by multiplying the score earned for each measure (column 2) by the assigned weight (column 3). In this example, it is determined that the raw score for measures of student learning is 1.65.

The sum of all weighted scores (1.65) in Table 4 represents the composite student learning score earned by the teacher. Table 5 translates the composite score range into qualitative judgments about student learning for a given teacher. The cut points in Table 5 for raw composite scores are based on scores of zero for Much Lower than Expected, one for Lower than Expected, two for Expected and three for Higher than Expected. When numbers in the four ranges in this table are combined and rounded to the nearest whole number, they are placed in the four categories as shown. The fractions are produced when teachers have multiple assessment scores which are weighted and averaged together.

Table 5. Cut Points for Composite Measures of Student Learning Scores

Composite Rating	Much Lower than Expected	Lower than Expected	Expected	Higher than Expected
Total RAW Composite Score Ranges (0-3)	0.0 to 0.49	0.50 to 1.49	1.50 to 2.49	2.50 to 3.0

In Figure 3 the raw composite score of 1.65 in Table 4 (above) is converted to a measures of student learning score between zero and 540. The measure of student learning score will be added to an educator's professional practices score in order to determine an overall effectiveness rating.

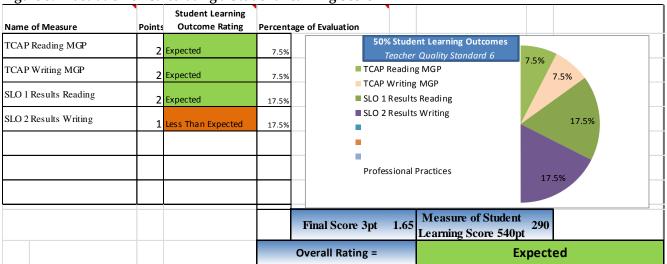**Figure 3: Illustration of Calculating a Student Learning Score**

Table 6 describes the method for converting the measures of student learning raw composite score into a measure of student learning score. Note: the model system Excel rubrics will do this math for users.

Table 6: Rules for Converting a Measure of Student Learning Raw Score to the 540 Point Scale

Measures of Student Learning Raw Composite Score		Computing a Measures of Learning Score
Much Lower than Expected	(0 < score < .5)	(score − .0) * 270
Lower than Expected	(.5 <= score < 1.5)	(score − .5) * 135 + 135
Expected	(1.5 <= score < 2.5)	(score − 1.5) * 135 + 270
Higher than Expected	(2.5 <= score <= 3.0)	(score − 2.5) * 270 + 405

Using the example of 1.65 above as the weighted average of four measure ratings, we can use Table 6 to convert 1.65 to the 540 scale with the Expected Growth formula:

(1.65 – 1.5) * 135 + 270 = 290, which would be the final measures of student learning score for this teacher.

Appendix B. Using Sample Reports to Determine Final Effectiveness Rating

1. **Sample report illustrating the professional practices score based on the observation rubric when standards are weighted differently**

TEACHER EVALUATION REPORT

Name: _Alex P. Keaton_____   **School:** _Mountain Top MS_____   **District:** _Durango_____

Professional Practices

Quality Standard	Element	Rating				
		Not Evident	Partially Proficient	Proficient	Accom-plished	Exemplary
I. Content and Pedagogical Knowledge	a. Provides aligned instruction		1			
	b. Demonstrates knowledge of literacy development		1			
	c. Demonstrates knowledge of mathematics (All)			2		
	d. Demonstrates knowledge of content			2		
	e. Develops interconnected lessons					4
	f. Makes instruction and content relevant to students	0				
	Overall Rating for Quality Standard I	*Proficient*				
II. Learning Environment	a. Fosters a predictable, caring learning environment				3	
	b. Demonstrates a commitment and respect for diversity					4
	c. Engages students			2		
	d. Adapts teaching to meet individual needs		1			
	e. Works and communicates with families		1			
	f. Creates a well managed learning environment		1			
	Overall Rating for Quality Standard II	*Proficient*				
III. Instruction	a. Demonstrates knowledge of current developmental science					4
	b. Plans and delivers data driven instruction				3	
	c. Demonstrates knowledge of effective instructional practices			2		
	d. Integrates and utilizes technology		1			
	e. Establishes high expectations for students			2		
	f. Provides opportunities to develop leadership qualities			2		
	g. Communicates to students effectively (objectives)	0				
	h. Uses appropriate methods to assess		1			
	Overall Rating for Quality Standard III	*Proficient*				
IV. Reflection	a. Analyzes student data and applies to instruction		1			
	b. Links professional growth to professional goals		1			
	c. Is able to respond to a complex, dynamic environment			2		
	Overall Rating for Quality Standard IV	*Partially Proficient*				
V. Leadership	a. Demonstrates leadership			2		
	b. Contributes to the teaching profession		1			
	c. Advocates for schools and students		1			
	d. Demonstrates high ethical standards					4
	Overall Rating for Quality Standard V	*Proficient*				
Overall Rating		**Proficient**				

Overall Rating For Professional Practice

Total Points for all Five Standards	Overall Rating on Professional Practices	Scores on This Evaluation		Number of points	Percent-age
		Standard	Rating		
0 to 2	Basic	I.	Proficient	1.85	11.11%
2.01 to 7	Partially Proficient	II.	Proficient	2.22	11.11%
7.01 to 12	Proficient	III.	Proficient	2.78	14.81%
12.01 to 17	Accomplished	IV.	Partially Proficient	0.74	5.56%
17.01 to 20	Exemplary	V.	Proficient	1.48	7.41%

Select Preset Weights		Total Points for 20 point scale:	Total Points for 540 point scale:	50.00%
⦿	Weight Standards - based on the # of elements within each standard	9.07	245	
○	Weight Standards - equally	Overall Rating on Professional Practices for All Standards		
○	Custom Weight	Proficient		

2. **Measures of student learning score based on the combination of multiple measures**

Quality Standard	Assessment Measure	Growth Rating				Percent weight	Raw Points Earned	Final Student Outcomes
		Much Lower Than Expected	Lower Than Expected	Expected	Higher Than Expected			
VI. Measures of Student learning	a. TCAP Reading MGP (collective school)			2		15%	0.30	(Total scaled points earned - .5) * 135 + 270
	b. TCAP Writing MGP (collective school)			2		15%	0.30	
	c. SLO 1 Results (collective grade level reading)			2		35%	0.70	
	d. SLO 2 Results (individual teacher)		1			35%	0.35	
	Overall Rating for Quality Standard VI	*Expected Growth*				Total points	1.65	290

3. **Final effectiveness rating for teacher with a total effectiveness score of 535 (score earned by adding the two components together: 245 + 290)**

Score Range	Final Rating
729 to 1080	Highly Effective
459 to 728	Effective
189 to 458	Partially Effective
0 to 188	Ineffective

Professional Practices	245
Measures of Student Learning	+ 290
Final Score	535
Final Rating	**Effective**

Appendix C. Approaches to Combining Scores for a Final Rating

[i]The table below describes two other common approaches besides the Compensatory model with accompanying methods used to combine scores earned across components. As indicated by the table, the methods can be adjusted to reflect the rules governing each selected approach. Districts are encouraged to use an approach and method that meets values expressed by stakeholders and reflects equal consideration of data from the professional practices (50 percent) and the measures of student learning (50 percent) portions of the evaluation system.

Common Approaches and Methods for Combining Scores to Achieve a Final Rating

Approaches	Stakeholder Values Expressed by Approach	Methods*
Disjunctive	This approach assigns a score or rating based on the highest performance achieved on a measure or component by the individual. This type of approach addresses concerns with over-identifying teachers labeled as Ineffective.	Decision matrix, profile, index (based on a simple or weighted average or summed points across measures)
Conjunctive	Requires a minimum level of performance on each measure to qualify for a given performance rating. Not meeting a specific threshold on one component or measure means that the rating would default to the lowest score achieved on either measure or component. This type of approach addresses concerns with over-identifying teachers labeled as Effective.	Decision matrix, profile, index (based on a simple or weighted average or summed points across measures)
Compensatory (Colorado State Model Evaluation System approach)	This approach allows performance on selected measures to be weighted so that they have the desired influence on the overall rating.	Decision matrix, profile, index (based on a simple or weighted average or summed points across measures)

*Note: The cut points set on performance distributions using any of the three methods noted can yield compensatory, disjunctive or conjunctive outcomes.