

## State Council for Educator Effectiveness Glossary

This glossary contains terminology that is often associated with the development of teacher performance management systems, many of which are used in the State Council for Educator Effectiveness Report. Specifically, the terms defined here are most often used when discussing teacher evaluation and the measures associated with that evaluation. The glossary is divided into three sections; the first contains general terms, the second includes definitions for the various ways of measuring teacher performance, and the final section defines technical aspects of teacher performance. Sources are cited in instances in which the definition has a primary source.

**Aggregation** – combining of multiple measures

**Analyzing student growth for “non-tested” subjects and grades** – methods that should be used to analyze the evidence and data elicited from the measurement tools for courses without at least two years of state summative assessment data.

**Analyzing student growth for “tested” subjects and grades** - methods that should be used to analyze the evidence and data elicited from the measurement tools for courses that have at least two years of state summative assessment (i.e. CSAP) data.

**Classroom Observations**—Used to measure observable classroom processes including specific teacher practices, aspects of instruction, and interactions between teachers and students. Classroom observations can measure broad, overarching aspects of teaching or subject-specific or context-specific aspects of practice.

**Covariate Adjusted Model**—In the covariate adjusted model, the current year’s test score is modeled as a function of prior test score(s) and other student or classroom characteristics. This model is in relatively wide use. This model may be further subdivided into two variant analytic strategies: a fixed-effect strategy yields less bias at the cost of precision, whereas the random-effect strategy produces estimates that are potentially biased but more precise. The potential correlation between teachers’ effectiveness and student characteristics yields biased estimates in the random-effect strategy, whereas including student fixed-effect coefficients to eliminate selection bias in the fixed-effect approach make estimation of the other coefficient less reliable (have higher variance). The model generally focuses only on year-to-year change.

**Criterion-referenced or standards-based** – a test which allows test scores to be translated into a statement about the behavior to be expected of a person with that score or their relationship to a specified subject matter. The objective of these types of tests is simply to see whether a student has learned the material, NOT how he/she compared with other students.

**Elements (of the Quality Standards)** – These are finer-grained descriptions which are used to further explicate/illustrate the skills/knowledge/dispositions that should be demonstrated by educators in their performance of the Quality Standards.

**Expected Growth**—A student’s expected/predicted performance on a current year test given his or her previous year’s test score. This information is obtained by regressing the current year test score on the prior year test score. In other words, estimating expected growth addresses the question, “Compared to students with the same prior test score, is the current year test score higher or lower than would be expected?”

**Formative Assessment**—Formative assessment is a process used by teachers and students during instruction that provides feedback to adjust ongoing teaching and learning to improve students' achievement of intended instructional outcomes.

**Gain Score Model**—A gain score model measures year-to-year change by simply subtracting the prior year score from the current year score. Typically, the gains for all students for a given teacher are averaged, and this average is then compared with the gain for a given teacher.

**Growth Models**—Traditional definitions of growth models indicate that they are models that measure student achievement growth from one year to the next by tracking the same students. This type of model addresses the question “How much, on average, did students’ performance change from one grade to the next?” To permit meaningful interpretation of student growth, the model implicitly assumes that the measurement scales across grades are vertically linked (i.e., that student scores on different tests across grades are directly comparable and represent a developmental continuum of knowledge and skill). An alternate understanding of growth models as put forth by Damian Betebenner is a model that examines performance of students with identical prior achievement scores and computes a percentile for each student indicating the probability of that outcome given the student’s starting point, which can be used to gauge whether or not the student’s growth was atypically high or low (Growth, Standards and Accountability, The Center for Assessment, April 2009; [http://www.nciea.org/publications/growthandStandard\\_DB09.pdf](http://www.nciea.org/publications/growthandStandard_DB09.pdf)).

**Instructional Artifacts**—Used to analyze classroom deliverables in order to determine the quality of instruction in a classroom. Instructional artifacts may include lesson plans, teacher assignments, scoring rubrics, and student work.

**Interim Assessments (form of summative assessment)** - A term generally used to refer to medium scale, medium-cycle assessments currently in wide use. Interim assessments: 1) evaluate students’ knowledge and skills relative to a specific set of academic goals, typically within a limited time frame, and, 2) are designed to inform decisions at both the classroom and beyond the classroom level, such as the school or district level. Thus, they may be given at the classroom level to provide information for the teacher, but unlike true formative assessments,

the results of interim assessments can be meaningfully aggregated and reported at a broader level.

**Layered Growth Model**—The “layered” model simultaneously models scores for multiple years in multiple subjects. Later years of teacher effects build upon estimated effects from earlier years (thus “layering” effects onto one another). This model is in relatively wide use. Covariates such as student background variables are not included. The “persistence” model, a more generalized version of proprietary EVAAS type models, relaxes the assumption of prior teacher effects and offers the possibility of including covariates.

**Multiple Measures of Student Learning**—The various types of assessments of student learning, including for example, value-added or growth measures, curriculum-based tests, pre-/post-tests, capstone projects, oral presentations, performances, or artistic or other projects.

**Multiple Measures of Teacher Performance**—The various types of assessments of teachers’ performance, including, for example, classroom observations, student test score data, self-assessments, or student or parent surveys.

**Non-tested Grades and Subjects**—Refers to the grades and subjects that are not required to be tested under the Federal Elementary and Secondary Education Act.

**Norm-referenced** - a type of test, assessment, or evaluation which yields an estimate of the tested individual’s performance evaluation relative to a predefined population, with respect to the trait being measured. This type of test identifies whether the test taker performed better or worse than other test takers, but not whether the test taker knows either more or less material than is necessary for a given purpose.

**Ordered Instructional Performance Task** – A curriculum-embedded performance task that is part of a teaching and learning cycle that is part of a sequence of tasks that reflects expected progress of students through that course or content area. By observing how students move along the sequence of tasks, teachers can assess student mastery.

**Other Assessments**—The development and/or adaptation of other measures of student growth for non-tested grades and subjects used across schools or districts. These measures may include early reading measures; standardized end-of-course assessments; formative assessments; benchmark, interim, or unit assessments; and standardized measures of English language proficiency. Other assessments may be developed at either the state education agency or local education agency level. Teacher-developed assessments of student learning or growth also may fall into this category when those assessments meet expectations for rigor and comparability across classrooms in a district or across classrooms statewide.

**Parent Surveys**—Questionnaires that usually ask parents to rate educators regarding various aspects of educators’ practice as well as the extent to which they are satisfied with the

educators' instruction, based upon the parent's personal experiences rather than specific expertise.

**Performance Management**—Refers to the entire system that affects a teachers' career continuum. Although teacher evaluation is a large component of the system, performance management refers more to the utilization of teacher evaluation data to inform decisions including hiring, tenure, compensation, and dismissal. In addition, successful performance management systems use data to inform professional development decisions and opportunities for teachers and principals.

**Performance Standards** - Levels of effectiveness established by rule of the state board pursuant to section 22-9-105.5 (10). Also, text-based descriptions that are used to define "effective", "highly effective", "partially effective" and "ineffective" educators. The Council has a proposed definition of "effective" and a work group of the Council will determine the definitions for the other proposed performance standards.

**Policy** – The formal guidance needed to coordinate and execute activity throughout an institution. When effectively deployed, policy statements help focus attention and resources on high priority issues - aligning and merging efforts to achieve the institutional vision. Policy provides the operational framework within which the institution functions.

**Pre- and Post-Tests**—Typically, locally developed student achievement tests that measure the content of the curriculum of a particular course. They are taken at the beginning of a time period (usually a semester or year) and then toward the end of that period to obtain a measure of student growth. Many pre- and posttest models also include mid-year assessments and formative assessments for teachers to adjust instruction throughout the course or year.

**Principal Evaluation Framework** - This outlines the approaches that should be used to measure the quality standards and elements for principals. This includes outlining a general approach for including both test- and non-test based information for the student growth portion of the evaluation.

**Procedure** - The operational processes required to implement institutional policy. Operating practices can be formal or informal, specific to a department or applicable across the entire institution. If policy is "what" the institution does operationally, then its procedures are "how" it intends to carry out those operating policy expressions.

**Quality Standards for Professional Practice**—The behaviors, skills, knowledge, and dispositions teachers should exhibit. Within the Council's proposed framework, these are the 6 (teachers) or 7 (principals) major categories that serve as the basis of judging educators as effective or not.

**Reliability**—The ability of an instrument to measure teacher performance or student achievement consistently across different rates and different contexts.

**Residual Growth**—A student's expected/predicted performance on a current year test given his or her previous year's test score. This information is obtained by regressing the current year test score on the prior year test score. In other words, estimating expected growth addresses whether the current year test score is higher or lower than would be expected compared to students with the same prior test score.

**Shared attribution** – process by which responsibility for student performance and growth may be shared by educators across a school, grade level, instructional department, teams or other grouping of teachers. The measures can take a variety of forms including schoolwide student growth measures, team-based collaborative achievement projects, and shared value-added scores for co-teaching situations.

**Shared Attribution OR Measures of Collective Performance**—The use of measures required by the current provisions of the Elementary and Secondary Education Act and/or other standardized assessments used to measure the performance of groups of teachers. Measures of collective performance may assess the performance of the school, grade level, instructional department, teams or other groups of teachers. These measures can take a variety of forms including schoolwide student growth measures, team-based collaborative achievement projects, and shared value-added scores for coteaching situations.

**State Scoring Framework for Principals** - This framework outlines how the data collected around the professional practice of principals, and the student growth outcomes for students in the school, and teachers in the school, should be combined in order to make a singular judgment about a principal in order to assign him/her to a performance standard.

**State Scoring Framework for Teachers** – This framework outlines how the data collected around the professional practice of teachers, and the student growth outcomes for teachers should be combined in order to make a singular judgment about each teacher in order to assign him/her to a performance standard.

**State Scoring Framework Matrix** – A matrix that will be adopted by all districts statewide to assign teachers and principals to appropriate performance standard ratings based on locally-calculated professional practice and student growth scores. The process of assigning cut scores and determining where performance ratings go within the matrix will be undertaken by CDE and the Council after examining the data gathered during the pilot and rollout phases.

**Student Growth**—The change in student achievement for an individual student between two or more points in time. A state also may include other measures that are rigorous and comparable across classrooms.

**Student Growth Objectives (SGOs)**—A participatory method of setting measurable goals, or objectives, based on the specific assignment or class, such as the students taught, the subject matter taught, the baseline performance of the students, and the measurable gain in student performance during the course of instruction. SGOs can be based on the Elementary and Secondary Education Act or other standardized assessments, but they also may be based on teacher-developed or other classroom assessments if they are “rigorous and comparable across classrooms.” The general method of SGOs draws on both effective pedagogical practices and approaches to goal setting and evaluation and task motivation found in multiple professions. In some instances, SGOs are shared by a team of job-alike teachers.

**Student Growth Percentile Model**—The student growth percentile model analyzes students’ progress from one year to the next in comparison with their academic peers with similar test score histories. It uses quantile regressions, which places students’ current performance relative to prior performance into a percentile metric. Currently implemented models (typically used for school accountability or reporting) do not include student background characteristics in the model, although it is possible to do so.

**Student Growth Percentile Model (e.g. Colorado Growth Model)**—The student growth percentile model analyzes students’ progress from one year to the next in comparison with their academic peers with similar test score histories. It uses quantile regressions, which places students’ current performance relative to prior performance into a percentile metric. Currently implemented models (typically used for school accountability or reporting) do not include student background characteristics in the model, although it is possible to do so.

<http://www.schoolview.org/>

**Student Surveys**--Questionnaires that usually ask students to rate educators regarding various aspects of educators’ practice as well as the extent to which they are satisfied with the educators’ instruction based the student’s personal experiences rather than specific expertise.

**Summative Assessment**—Summative assessment is a process that concerns final evaluation to ask if the project or program met its goals. Typically the summative evaluation concentrates on learner outcomes rather than only the program of instruction. It is a means to determine a student’s mastery and understanding of information, skills, concepts, or processes. Summative assessments occur at the end of a formal learning/instructional experience, either a class or a program.

**Teacher-Developed Test**—Assessments of student achievement or growth that have been developed by a teacher or teachers.

**Teacher Effect**— A teacher’s contribution to student performance growth compared with that of the average (or median, or otherwise defined) teacher in the district or the state. In essence, teacher effect is the difference between the observed student achievement growth and the expected student achievement growth (controlling for confounding factors, such as prior



student achievement and sometimes student background factors), which are interpreted as representing differences in student achievement growth due to differences in teacher effectiveness. Note that the description of “school effect” or “principal effect” is less straightforward because it will depend on decisions about how to aggregate grade- or subject-level estimates based on the specific model employed to determine teacher effects.

**Teacher Evaluation Framework** - This outlines the approaches that should be used to measure the quality standards and elements for teachers. This includes outlining a general approach for including both test- and non-test based information for the student growth portion of the evaluation.

**Teacher Portfolios and Evidence Binders**—A collection of materials that exhibit evidence of teaching practice, school activities, and student progress. They are usually compiled by the teacher and may include teacher-created lesson or unit plans, descriptions of the classroom context, assignments, student work samples, videos of classroom instruction, notes from parents, and teachers’ analyses of their students’ learning in relation to their instruction. Evidence binders often have specific requirements for inclusion and involve a final teacher-led presentation of the work to an evaluation team.

**Teacher Self-Assessments**—Surveys, instructional logs, or interviews in which teachers report on their work in the classroom; the extent to which they are meeting standards; and in some cases, the impact of their practice. Self-assessments may consist of checklists, rating scales, and rubrics and may require teachers to indicate the frequency of particular practices.

**TELL Survey** – The Teaching Empowering Leading and Learning Survey is a statewide survey of school based educators to assess teaching conditions at the school, district and state level.  
<http://www.tellcolorado.org/>

**Unique Identifier**—Numbers that are assigned to each individual student and teacher in a school and are matched to data about that student’s or teacher’s performance.

**Validity**—The ability of an instrument to measure the attribute that it intends to measure.

**Value-Added Estimate**—A comparison of teacher effects with the counterfactual (sometimes referred to as a “typical” teacher). If the teacher effect is higher than the counterfactual, then the teacher may be perceived as effective (i.e., positive value-added). Conversely, if the teacher effect is lower than the counterfactual, then the teacher may be perceived as ineffective (i.e., negative value-added). The number or rating produced in the comparison is the value-added estimate.

**Value-Added Models (VAMs)**—Complex statistical models that attempt to determine how specific teachers and schools affect student achievement growth over time. This model generally uses at least two years of students’ test scores and may take into account other student- and school-level variables, such as family background, poverty, and other contextual

factors. VAMs attempt to determine the extent to which changes in student performance can be attributed to a specific school and/or teacher compared with that of the average school or teacher.

**Weighing** – the process of determining how the data elicited from an evaluation should be interpreted or used to measure performance on a particular Quality Standard.

**Weighting** – how much a particular measurement tool, student growth measure or quality standard determines overall performance when multiple measures are combined into a single rating.