

Supporting State Efforts to Design and Implement Teacher Evaluation Systems

A Workshop for Regional Comprehensive Center Staff Hosted by the
National Comprehensive Center for Teacher Quality
With the Assessment and Accountability Comprehensive Center

December 13–14, 2010

Glossary

This glossary contains terminology that is often associated with the development of teacher performance management systems. Specifically, the terms defined herein 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.

General Terminology

Effective Teacher—“Teacher whose students achieve acceptable rates (e.g., at least one grade level in an academic year) of student growth. A method for determining if a teacher is effective must include multiple measures, and effectiveness must be evaluated, in significant part, on the basis of student growth. Supplemental measures may include, for example, multiple observation-based assessments of teacher performance” (Race to the Top Application for Initial Funding; CFDA Number: 84.395A).

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).

Highly Effective Teacher—“A teacher whose students achieve high rates (e.g., one and one-half grade levels in an academic year) of student growth. A method of determining if a teacher is highly effective must include multiple measures, provided that teacher effectiveness is evaluated, in significant part, on the basis of student growth. Supplemental measures may include, for

example, multiple observation-based assessments of teacher performance or evidence of leadership roles (which may include mentoring or leading professional learning communities) that increase the effectiveness of other teachers in the school or LEA” (Race to the Top Application for Initial Funding; CFDA Number: 84.395A).

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.

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

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.

Standards for Professional Practice—A set of ideals for what behaviors, skills, knowledge, and dispositions teachers should exhibit.

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.

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.

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.

Measures of Teacher Performance

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.

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.

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.

Other Assessments—The development and/or adaptation of other measures of student growth for nontested 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 teachers on an extent-scale regarding various aspects of teachers’ practice as well as the extent to which they are satisfied with the teachers’ instruction.

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.

Student Learning Objectives (SLOs)—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. SLOs 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 SLOs draws on both effective pedagogical practices and approaches to goal setting and evaluation and task motivation found in multiple professions. In some instances, SLOs are shared by a team of job-alike teachers.

Student Surveys—Questionnaires that typically ask students to rate teachers on an extent-scale regarding various aspects of teachers’ practice as well as how much students say they learned or the extent to which they were engaged.

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

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.

Technical Terminology

Some of the terms in this section are not *specific* to education; however, they are used in education conversations and are important to know.

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.

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?”

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.

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.

Reliability—The ability of an instrument to measure teacher performance 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.

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.

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.

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.