

Building a Common Language: Glossary

Attainment model is a method for measuring how students perform at one point in time. For example, the percent of fourth graders were scoring at-proficient or above in 2010.

Classroom observations are used by evaluators to make judgments of teachers' practice in the classroom. Classroom observations are the most common form of teacher evaluation and vary widely in how they are conducted and what they assess. High-quality classroom observation instruments are based on standards and contain well-specified rubrics that delineate consistent assessment criteria for each standard of practice. Evaluators should be trained to ensure accuracy and consistency in scoring. A transparent system ensures that all educators who will be observed know how the process will be conducted and how the findings will be used.

Comparability is the extent to which student learning is being measured consistently across classrooms, schools, and districts so that scores have the same meaning both within and across subject areas and grade levels.

Disaggregate is to separate into different categories or to show individual results.

Effective teacher is an individual teacher who produces substantial student outcomes. Effective teacher is often narrowly defined as a teacher who contributes more to student test scores than other teachers. This focus implies evaluating teacher performance by treating teachers and students as single units rather than interconnected members of a learning community. Not to be confused with effective teaching (which addresses the complexity of teaching by recognizing contextual factors).

Effective teaching consists of instruction that enables all students to meet or exceed ambitious goals for student learning (adapted from Darling-Hammond, 2010). Effective teaching is in part a function of individual teacher talent, knowledge, and skills, but it is also highly influenced by the conditions in which teaching takes place—the school leadership, the quality of curriculum materials and resources, the opportunities teachers have for professional growth and learning, the size of their workloads, and the time teachers have to prepare, among other factors.

Evaluation of student artifacts and work judged according to rubrics means that evaluators rate lesson plans, teacher assignments, teacher-created assessments, scoring rubrics, and student work on particular criteria, such as rigor, authenticity, intellectual demand, alignment to standards, clarity, and comprehensiveness. Evaluators typically use an evaluation tool or rubric to make judgments about the quality of student artifacts.

Evaluation tools are models, rubrics, instruments, and protocols that are used by evaluators to assess teachers' performances.

Formative teacher evaluation is the assessment of teachers' practices for the purposes of supporting or improving teachers' practices.

Formative assessments are assessments used while the learning process is underway or recently completed, and designed to inform current or future instruction.

Goal-driven professional development plans are evaluation tools that offer teachers the opportunity to set their own ambitious but feasible objectives for their professional growth in collaboration with their principal or other colleagues. Some tools require teachers to specify the professional development in which they will participate to ensure their students achieve their growth objectives.

Growth measures are assessments of students' improvements in learning from one point in time to another point in time.

Growth models or **individual student growth models** generally refer to models that measure progress by tracking the achievement scores of the same students from one year to the next, with the intent of determining whether the students in a given group are making progress (Goldschmidt, Roschewski, Choi, Auty, Hebbler, et al., 2005). For example, the model can compare the performance of this year's eighth graders with the performance of the same students the previous year in the seventh grade. Growth models control for mobility of students between schools from year to year as well as students' prior achievement and the effects of their family background. (Blank, 2010) Types of growth models include growth-to-proficiency, linear growth, projection, transition table, and value-added growth models.

Growth to proficiency models measure whether students are on track to meet standards for proficient and above.

Multiple measures of student learning are the various types of assessment of students' learning, for example, value-added or growth measures, curriculum-based tests, pre- and posttests, capstone projects, oral presentations, performances, and artistic or other projects.

Multiple measures of teaching effectiveness are the various types of assessments of teachers' performance, for example, classroom observations, student test-score data, self-assessments, and student or parent surveys.

Parent surveys are questionnaires that typically ask parents to rate teachers on a scale from 1 to 5, where 1 = very effective, and 5 = not at all effective) for various aspects of teachers' practice (e.g., course content, usefulness of feedback, quality of homework, quality of communication) as well as the extent to which they are satisfied with the teacher's instruction.

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

Reliability is a measure of the degree to which an instrument measures something consistently. A validated instrument must be evaluated for how reliable the results are across raters and contexts. Discussion of methods for measuring teaching effectiveness often make reference to rater reliability—whether or not raters have been trained to score reliably. Scoring reliably means being able to do the following: rate consistently with standards, rate consistently with other raters (referred to as interrater reliability), and rate consistently across observations and contexts. Ratings should not be influenced by factors such as the time of day, time of year, or subject

matter being taught, and they should be consistent across observations of the same teacher (from Little, Goe, & Bell, 2009).

Self-assessments are 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, or rubrics, and they may require teachers to indicate the frequency of particular practices.

Standards for professional practice are a set of ideals for what behaviors, skills, knowledge, and dispositions teachers should exhibit.

Student growth data are information about change in students' performance on some measure such as a test between two or more points in time.

Student progress is the extent to which individual students are moving through the learning process.

Student surveys are questionnaires that typically ask students to rate teachers on a scale from 1 to 5, where 1 = very effective, and 5 = not at all effective, for various aspects of teachers' practice (e.g., relevance of course content, usefulness of feedback) as well as how much students say they learned or the extent to which they were engaged.

Summative teacher evaluation is the assessment of teachers' practice for the purpose of making high-stakes personnel decisions.

Teacher growth and development system is a comprehensive performance assessment system that incorporates multiple measures of both teacher evaluation and student learning and has the intent of improving the knowledge, skills, dispositions, and classroom practices of professional educators. Beyond a simple evaluation system, a teacher growth and development system is connected closely to other key aspects of the teaching continuum (induction, professional development, promotion, etc.) and recognizes the importance of formative assessment for the improvement of teaching.

Teacher portfolios and evidence binders are a collection of materials that exhibit evidence of teaching practice, school activities, and student progress. The teacher usually compiles them him- or herself. They 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 the teacher's analyses of their students' learning in relation to their instruction. Evidence binders often provide specific requirements for inclusion and require a final teacher-led presentation of the work to an evaluation team.

Teacher quality refers to teachers who are certified, have a bachelor's degree, and have a major in their subject area. Qualified teachers hold credentials certifying that they have successfully completed a state-approved (often nationally accredited) teacher preparation program, have demonstrated their good character (usually through a criminal background check), and hold a bachelor's degree. Moreover, states almost always require an examination of content and pedagogy for state certification.

Unique identifiers are numbers that are assigned to each student and teacher in a school; they can be matched to data about that student's or teacher's performance.

Validity is a measure of the degree to which an interpretation of an evaluation score is supported by evidence. For a measure of teaching effectiveness to be valid, evidence must support the argument that the measure actually assesses the dimensions of teaching effectiveness it claims to measure and not something else. It also is essential to have evidence that the measure is valid for the purposes for which it will be used. Instruments cannot be valid in and of themselves; an instrument or assessment must be validated for particular purposes (Gordon, Kane, & Staiger, 2006; Messick, 1989, as cited in Little, Goe, & Bell, 2009).

Value-added student test scores are a measure of the contribution that a teacher makes on the growth of his or her students' test scores, taking into account those students' test scores in prior years and also often taking into account socioeconomic and other factors that might affect growth in scores.

DRAFT