

Multiple Ways to Measure Teacher Effectiveness

- **Planning:** Instructional planning provides a window into a teacher's preparation to deliver content, scaffold the development of student skills, and manage the classroom learning environment. Current evaluations rarely link planning to practice.
- **Classroom Observation:** The classroom observation is the most commonly used tool for evaluating teachers and it requires highly trained and skilled observers to recognize and understand the complexities of a well developed lesson and effective instruction.
- **Self Assessment and Reflection:** Self Assessment and reflection is a process in which teachers analyze their own instruction retrospectively. It can occur in a variety of ways; professional conversations with other teachers during grade or subject-area meetings, pre-observation and post-observation debriefings, the development of a portfolio, or an individual professional development plan.
- **Perception Survey Data:** Generally speaking, perception survey data (usually collected through surveys) comes from students, parents, and adults who are connected with the school. Research indicates that perception survey data is best used as formative assessment (evaluation meant to improve instruction and student learning) and developed by those connected to the results of the survey.
- **Peer Evaluation:** Peer review is a practice in which experienced teachers mentor new and veteran teachers to improve their knowledge and teaching skills. Participants in peer review programs are identified as experiencing pedagogical problems, or problems related to classroom management or curriculum design. Consulting teachers observe classrooms, sharing ideas and skills and recommending study materials to assist struggling teachers and improve teacher quality. This method of collecting data works well as formative evaluation and creates a support system for inexperienced teachers and veteran teachers who are struggling. Some emerging peer evaluation models are being used for summative purposes and for adjudication.
- **Portfolio Assessments:** Portfolio assessments tend to comprise several pieces of evidence of teacher classroom performance, including lesson or unit plans, a video of classroom teaching, reflection and self-analysis of teaching practices, examples of student work, and examples of teacher feedback given to students and parents (Andrejko, 1998).
- **Student Achievement Data:** Student achievement data ranges from standardized student test scores all the way to classroom-based assessments. In some cases, states and districts are adopting systems that use statistical techniques and models to try to isolate effects of a teacher on student learning. Two models that are getting lots of attention lately (both positive and negative) are value-added models and the application of growth models. Use of student achievement data falls into two general categories; using the results of student assessments to determine teacher quality (summative) or using the student assessments to modify instructional practice (formative). Both formative and summative methods are being used in teacher evaluation.
- **Student Work Samples:** An emerging view is that there may be alternative ways to measure the effect of instruction on student learning, including the analysis of student work samples (Mujis, 2006). This method is intended to provide a more insightful review of student learning results over time and can be used in classes not connected to district or state assessments (visual and performing arts, PE, choir, band, orchestra, social studies, career and technical education, early elementary, etc).

| Multiple Measures | Strengths | Limitations |
|---------------------------------------|--|--|
| Planning | <p>One aspect of teaching correlated with student learning is the level of planning used to drive instruction (Stronge, 2007). Lesson plans are more likely to be positively related to improved student outcomes when plans are able to (1) link student learning objectives with teaching activities, (2) describe teaching practices to maintain students' attention, (3) align student learning objectives with the district and state standards, and (4) accommodate students with special needs (Stronge, 2007)</p> | <p>A lesson plan is indeed a "plan," and once it is implemented, the plan may need to be adjusted. The quality and appropriateness of the adjustments a teacher makes in the implementation of the plan in the classroom cannot be evaluated solely from a lesson plan scoring rubric.</p> |
| Classroom Observation | <p>Classroom observations capture information about teachers' instructional practice (Mujis, 2006). Observations can be used in formative and summative evaluations. When used in formative evaluations, the observation can track a teacher's growth and suggest needed professional development – the results of which can then be assessed in subsequent observations.</p> | <p>Despite the frequent use of classroom observations for the purpose of evaluating teacher performance, this measure is not without its limitations. Poorly trained observers and inconsistent, brief observations can create biased results (Shannon, 1991; Shavelson, Webb, & Burstein, 1986). Research suggests that when observations occur more frequently, their reliability improves (Denner, Miller, Newsome, & Birdson, 2002) and similarly, when observations are longer, their validity improves (Cronin & Capie, 1986).</p> |
| Self assessment and reflection | <p>Requiring reflection as a part of an evaluation process may encourage teachers to continue to learn and grow throughout their careers (Uhlenbeck, et al., 2002.) To encourage reflection, some evaluation systems include videotaping teachers in the classroom. The videotaped class sessions may be rated as classroom observations, but these videotapes also allow teachers to review their performance so they can reflect and engage in in-depth conversations with their evaluators about the behaviors and practices observed.</p> | <p>Reflection requires both time and a cultural norm that supports this type of evaluation practice in a school or district. When reflection is not typically used for evaluative purposes, making the time for teachers to engage in this practice is a low priority for administrators (Peterson & Comeaux, 1990; Schon, 1983).</p> |
| Perception survey data | <p>While opinions on instructional practice and student learning are not direct measures of instructor or course effectiveness, they are legitimate indicators of student satisfaction, and there is substantial research linking student satisfaction to effective teaching (Theall and Franklin, 2001). In the area of student surveys as formative assessment, a meta-analysis of 41 research studies provides the strongest evidence for the validity of student ratings since these studies investigated the relationship between student ratings and student learning. There are consistently high correlations between students' ratings of the "amount learned" in the course and their overall ratings of the teacher and course (Gaubatz, 2000).</p> | <p>Teachers may alter their teaching to get higher ratings including weakening the difficulty of the course or giving higher grades. Ratings may be used in a summative fashion and not as a formative method. The content of the student rating form may drive what is taught. Students reward poor teaching by giving high ratings in exchange for high grades. Ratings are used to make discriminations between teaches that are not supported by other data. The data becomes meaningless because of the lack of use and control.</p> |

| | | |
|--|--|---|
| <p>Peer Evaluation</p> | <p>Peer evaluation engages evaluators who are most connected to instruction; fellow teachers. Teachers in this role tend to be very skilled in the art of instruction and can easily identify areas of strength and weakness with teacher practice. Peer evaluations should be used as an opportunity for teachers to discuss with colleagues how effectively they are accomplishing their goals in the classroom. Peer evaluation is best viewed as a formative evaluation technique designed to improve teacher quality.</p> | <p>Peer evaluation can be problematic as colleagues are evaluating fellow educators. Depending on how the results of peer evaluation are used, circumstances can arise which may cause teachers to feel hesitation in writing out completely honest evaluations of their colleagues. Knowing how the results will be used could also cause the teacher to feel the need to go to extraordinary lengths to produce a lesson which is not representative of his/her typical classroom instruction. This technique is better utilized in a formative manner.</p> |
| <p>Portfolio</p> | <p>Teachers and administrators often favor the use of portfolios because they enable teachers to reflect on their own practice, allow evaluators to identify teachers' instructional strengths and weaknesses, and encourage ongoing professional growth (Attinello, Lae & Source, 2006; Tucker, Stronge, & Gareis, 2002). Portfolios are useful evaluation tools because they allow evaluators to review non-classroom aspects of instruction as well as provide teachers with opportunities to reflect on their teaching by reviewing documents contained in the portfolio (Danielson, 1996). Student assessment can be collected in portfolios to demonstrate student learning. Portfolios also promote the active participation of teachers in the evaluation process (Attinello et al., 2006)</p> | <p>The necessary time to develop and review a portfolio is a frequently cited concern among educators. Currently, there are no conclusive findings on the reliability of portfolio assessments as part of an objective teacher evaluation system (Attinello et al., 2006). Existing research has raised questions about whether portfolios accurately reflect what occurs in classrooms and whether the process of developing a portfolio and being evaluated through that process leads to improvements in teaching practices (Attinello, 2006, Tucker, 2002)</p> |
| <p>Student Achievement Data</p> | <p>The use of student test scores enables schools to measure the impact that instruction is having on student performance and builds on existing investment in student testing. While the quality of state and local assessments differ widely, the items on well-developed assessments have been tested for issues of fairness and appropriateness through the application of various statistical models. If assessments are designed for evaluation purposes, schools have an opportunity to examine the relationship between changes in student achievement gains, teachers, and schools.</p> | <p>Student test scores measure only a portion of the curriculum and teachers' effects on learning (Berry, 2007). Most statistical models are not able to differentiate which elements of teaching relate to positive student achievement test outcomes. Teacher value-added effects on test scores are meaningful only in relation to one another, rather than to established teaching proficiency criteria. Some models compare different cohorts of students. Some models don't have vertical scales to measure absolute growth. Most state assessments are not designed for teacher evaluation purposes. Not all teachers can be assessed using state assessments.</p> |

| | | |
|------------------------------------|---|---|
| <p>Student work samples</p> | <p>Using student work samples as the basis for a review of teacher practice is a much stronger indicator of student proficiency than using state assessments for this purpose (Schwabacker, 1993). The research suggests that student work samples may help to better identify which elements of teaching relate more directly to increased student learning than standardized test scores.</p> | <p>One drawback to using student work samples in evaluations is that reviewing these samples can be time-consuming. In addition, the review of student work samples as a means of evaluating teacher effectiveness is more prone to issues of validity and reliability. To reduce subjectivity and address issues of reliability, experts should develop a research-informed scoring rubric that outlines criteria for rating student work samples. Those using the rubric should be trained so that the process is consistent across all student sample evaluations.</p> |
|------------------------------------|---|---|