



District  
Management  
Group

## MANAGER'S TOOLKIT

# How to Design a Multidimensional Teacher Evaluation System

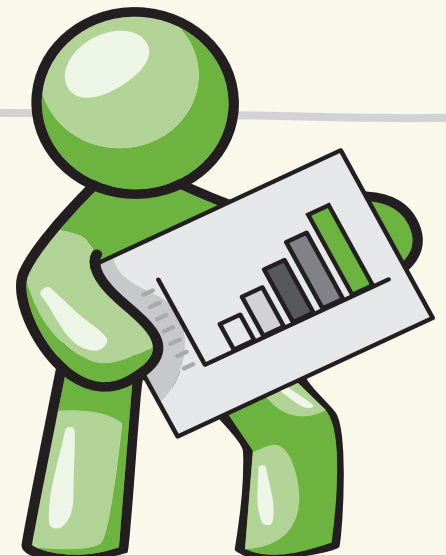
Originally published in *District Management Journal*, v.6, Winter 2011

*Use this toolkit to assess key areas that need to be determined in designing a teacher evaluation system for your district.*

# How to Design a Multidimensional Teacher Evaluation System

With research showing that most teacher characteristics and qualifications have little predictive effect on student achievement outcomes, the national dialogue has shifted from teacher quality to teacher effectiveness. Race to the Top and the ESEA Blueprint ask states and local districts to establish definitions of teacher effectiveness “that are based in significant part on student growth and also include other measures, such as classroom observations of practice.” Reform emphasis is now on robust evaluation systems that tie quantitative and qualitative measures of effectiveness together. Assessing teacher effectiveness along these dimensions requires a multidimensional approach that includes both teacher practice and student learning.

DMC has designed a process to guide districts in creating a teacher effectiveness program that takes into account the specific needs and constraints of each district. The process incorporates the multiple measures and multiple ratings that we at DMC believe need to be included to create a powerful evaluation system. The summary tool on the following pages outlines the key issues that need to be determined in designing a teacher evaluation system for your district. A complete workbook including worksheets to guide the discussion and design of individual measures and to work through the weighting for each measure is available to members at [www.dmcouncil.org](http://www.dmcouncil.org).



*Turn the page for your DMC Toolkit* ▶

# How to Design a Multidimensional Teacher

Category of Teacher to Evaluate (e.g. core/tested, non-core, etc.): \_\_\_\_\_

	Measurement Approach	Description	Pros/Cons
Teacher Practice/ Qualitative Measures	Classroom Observations	Observations of classroom instruction. Observations can take a wide range of formats.	<p><b>Pros:</b></p> <ul style="list-style-type: none"> <li>○ Considered by most to be the “gold standard”</li> </ul> <p><b>Cons:</b></p> <ul style="list-style-type: none"> <li>○ Expensive in time, resources, and manpower</li> <li>○ Announced or planned observations may not be representative</li> <li>○ Possible problems with variability due to observer’s training or other biases</li> <li>○ Potential to be disruptive to the instructional process</li> </ul>
	Analysis/Rating of Classroom Artifacts	A qualitative professional judgment and scoring of a set of classroom artifacts. Classroom artifacts may include the following: lesson plans, curriculum units, student work samples, audio or video and/or classroom discussion transcripts, and more.	<p><b>Pros:</b></p> <ul style="list-style-type: none"> <li>○ Allows asynchronous reflection on teacher’s classroom</li> <li>○ Provides feedback on specific classroom tactics</li> <li>○ Less costly than classroom observations</li> </ul> <p><b>Cons:</b></p> <ul style="list-style-type: none"> <li>○ Sample artifacts may not be representative</li> <li>○ May not correlate with student achievement outcomes</li> </ul>
	Analysis/Rating of Teacher Portfolio	A qualitative professional judgment of a teacher portfolio. A portfolio may include the following: a summary of teaching experience and responsibilities, a reflective statement of teaching philosophy and goals, a discussion of teaching methods and strategies, activities undertaken to improve teaching, and a statement of goals and plans for the future.	<p><b>Pros:</b></p> <ul style="list-style-type: none"> <li>○ Encourages self-reflection and long-range planning</li> <li>○ Less costly than classroom observations</li> </ul> <p><b>Cons:</b></p> <ul style="list-style-type: none"> <li>○ Sample artifacts may not be representative</li> <li>○ May not correlate with student achievement outcomes</li> </ul>
	Teacher Self-evaluation	The teacher completes a self-scoring rubric that may address the following areas: classroom environment, curriculum and instruction, planning and scheduling, documentation and assessment, interactions with families, and more.	<p><b>Pros:</b></p> <ul style="list-style-type: none"> <li>○ Encourages self-reflection</li> </ul> <p><b>Cons:</b></p> <ul style="list-style-type: none"> <li>○ Possibly uncorrelated with student achievement outcomes</li> </ul>
	Stakeholder Input from Standardized Surveys (Peers, Students, and/or Parents)	Solicitation of feedback using standardized survey instruments on specific dimensions of effective teaching (e.g. professional responsibilities or role in the community).	<p><b>Pros:</b></p> <ul style="list-style-type: none"> <li>○ Provides additional voice to evaluation process</li> <li>○ May drive behavior change and encourage a “customer service” mindset</li> </ul> <p><b>Cons:</b></p> <ul style="list-style-type: none"> <li>○ May be culturally difficult to adopt</li> </ul>
Student Achievement/ Quantitative Measures	Achievement or “Improvement” data: Individual Classroom or School-wide	Improvement models are used to measure the change in test results for a teacher or school by comparing status at two points in time—but not for the same students. These models implicitly assume that student populations remain fundamentally similar over time.	<p><b>Pros:</b></p> <ul style="list-style-type: none"> <li>○ Accountability data is readily available</li> <li>○ Easier to understand than VAM or Growth models</li> <li>○ Assumes student population characteristics are similar enough to compare overall achievement levels</li> </ul> <p><b>Cons:</b></p> <ul style="list-style-type: none"> <li>○ Data does not account for uncontrollable factors</li> </ul>
	Growth data: Individual Classroom or School-wide	Growth models track the test scores of the same students from one year to the next to determine the extent of their progress. Growth models usually do not control for student or school background factors, and therefore they do not attempt to address which factors are responsible for student growth.	<p><b>Pros:</b></p> <ul style="list-style-type: none"> <li>○ Data focuses on cohort growth, not static achievement levels</li> </ul> <p><b>Cons:</b></p> <ul style="list-style-type: none"> <li>○ Appropriate data is difficult to assemble</li> <li>○ Data is difficult to understand and use to improve classroom effectiveness</li> </ul>
	Value-added Data: Individual Classroom or School-wide	With most models, the value-added estimate for a school or a teacher is the difference between the observed improvement of the students and the expected improvement (after taking account of differences among students that might be related to their academic achievement).	<p><b>Pros:</b></p> <ul style="list-style-type: none"> <li>○ Data is most precise method of measuring collective performance effect</li> <li>○ School-wide measurement of improvement can build collaboration</li> </ul> <p><b>Cons:</b></p> <ul style="list-style-type: none"> <li>○ Statistical models may be flawed, misused, or misinterpreted</li> </ul>

# Evaluation System



Readiness Level	Who	When
Will you do this? How? What are the potential barriers to implementation?	Who needs to be involved? What will their role be?	By when will this happen? How frequently?
	<input type="radio"/> School Principal? Other qualified, designated reviewer?	<input type="radio"/> Scheduled or unscheduled? Short or long?
	<input type="radio"/> School Principal? Other professionals?	
	<input type="radio"/> School Principal alone? School Principal and other professionals?	
	<input type="radio"/> Teacher. Teacher may choose to supplement their own reflections with input from peers, students, or families through surveys, etc.	
	<input type="radio"/> Peers? Students? Parents?	



Turn the page for more ▶

# Weighting



**It is important to consider how much each measurement approach (e.g., classroom observations, parent input, classroom growth data) will weigh in the total evaluation tool. This worksheet is intended to help you summarize these different weights.**

**How many different types of evaluations will you have, considering data availability and differences in subject matter taught?**

- One for all teachers
- Two (e.g., for tested and non-tested)
- Three (e.g., for tested core, non-tested core, non-core)
- Four or more

**How will you weight the quantitative and qualitative elements chosen?**

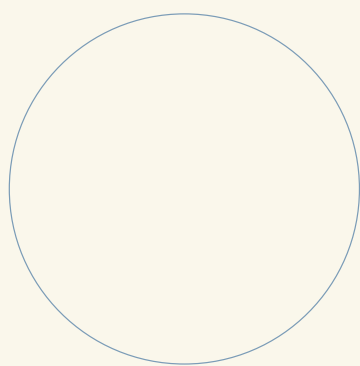
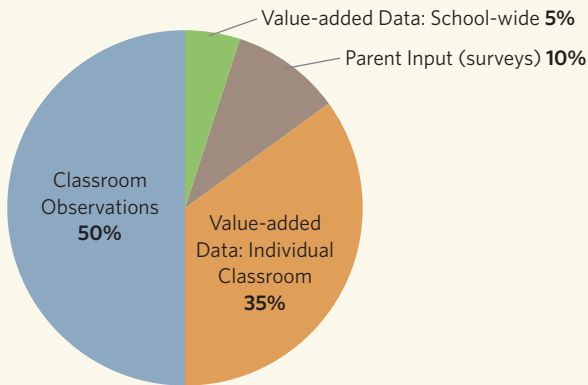
(e.g., 100% qualitative vs. 100% quantitative vs. 80/20 vs. 20/80 vs. 50/50)

Complete the pie chart below, providing labels and weighting for each measure that will be included in your evaluation tool. If there are multiple types of evaluations, fill out a pie chart for each.

**EXAMPLE:**

Type: Core ELA/Math (3rd-11th grade)

Type: \_\_\_\_\_



**What will the sum total of all of this information be used for? (Check all that apply)**

- |   |                                       |
|---|---------------------------------------|
| <input type="checkbox"/> Summative Evaluation     | <input type="checkbox"/> Commendation |
| <input type="checkbox"/> Professional Development | <input type="checkbox"/> Promotion    |
| <input type="checkbox"/> Discipline               | <input type="checkbox"/> Other        |

**Are there potential barriers to the weighting system you have selected?**

---



---



---