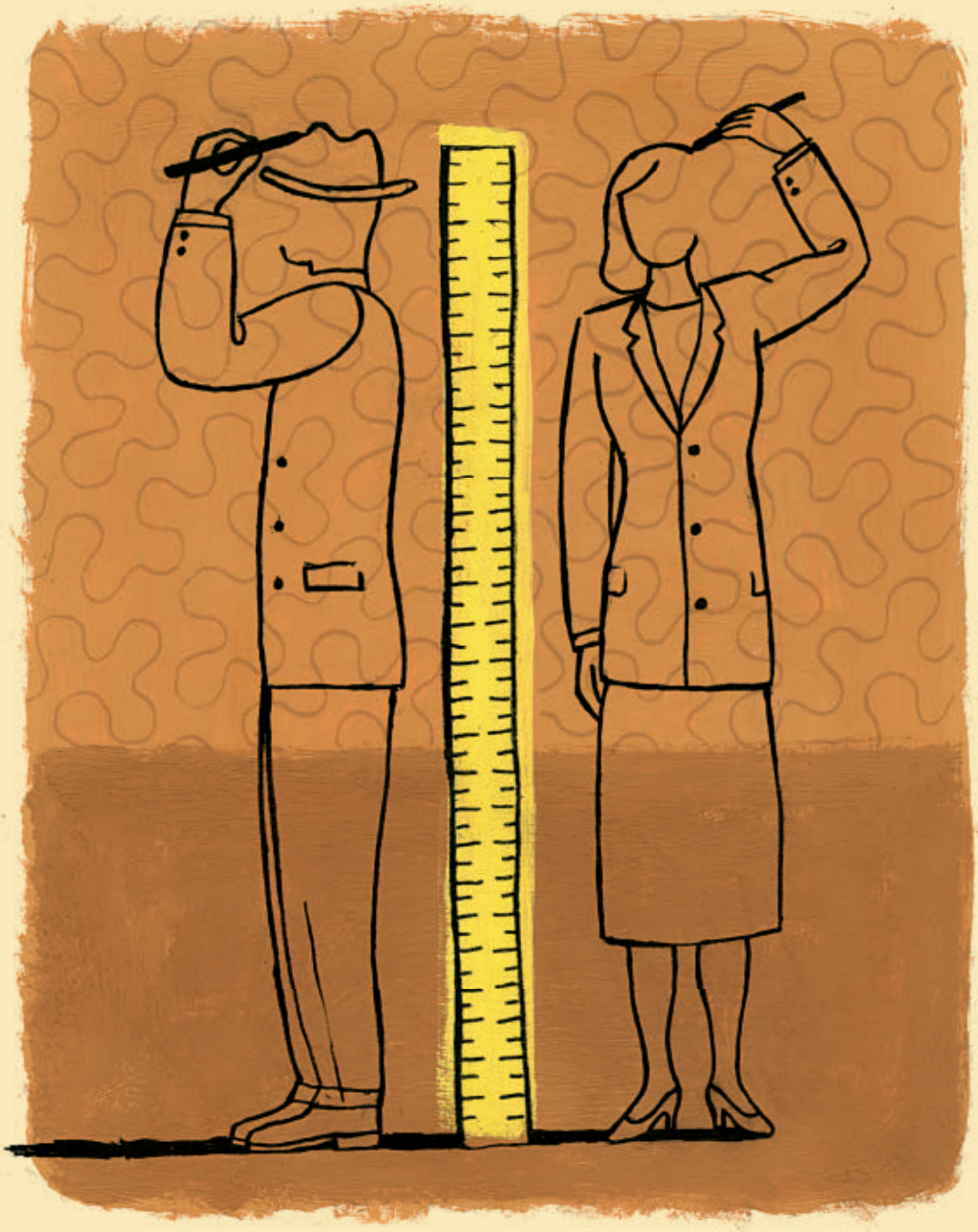
The Two Purposes of Teacher

An evaluation system that fosters teacher learning will differ from one whose aim is to measure teacher competence.

Robert J. Marzano



Evaluation

States, districts, and schools all across the United States are busy developing or implementing teacher evaluation systems. One can trace this flurry of activity to a variety of reports and initiatives that highlight two failings of past efforts: (1) Teacher evaluation systems have not accurately measured teacher quality because they've failed to do a good job of discriminating between effective and ineffective teachers, and (2) teacher evaluation systems have not aided in developing a highly skilled teacher workforce (Bill and Melinda Gates Foundation, 2011; Toch & Rothman, 2008; U.S. Department of Education, 2009; Weisberg, Sexton, Mulhern, & Keeling, 2009).

Although efforts to move quickly in designing and implementing more effective teacher evaluation systems are laudable, we need to acknowledge a crucial issue—that *measuring* teachers and *developing* teachers are different purposes with different implications. An evaluation system designed primarily for measurement will look quite different from a system designed primarily for development.

Teacher evaluation systems have not accurately measured teacher quality and **have not aided in developing a highly skilled teacher workforce.**

Which Is Best?

Over the last year, I've asked more than 3,000 educators their opinions about these two basic purposes by presenting them with a scale that has five values. If educators think that measurement is the sole purpose of teacher evaluation (that is, that development should not be a purpose of teacher evaluation), they select 1. If educators think that development is the sole purpose of teacher evaluation (that is, that measurement should not be a purpose of teacher evaluation), they select 5. If they believe that the purpose of teacher evaluation should be half measurement and half development, they select 3. A value of 2 indicates that measurement and development should be dual purposes but that measurement should be dominant. Finally, 4 indicates that measurement and development should be dual purposes but that development should be dominant.

To date, educators have responded in the following way: No one selected 1, 2 percent selected 2, 20 percent selected 3, 76 percent selected 4, and 2 percent selected 5. Stated differently, the vast majority of respondents believe that teacher evaluation should be used for both measurement and development but that development should be the more important purpose. Although the 3,000 educators I queried do not constitute a representative sample, their responses do raise the issue of what teacher evaluation looks like when its primary purpose is development.

Systems That Focus on Development

Teacher evaluation systems that are designed to help teachers improve have three primary characteristics.

The System Is Comprehensive and Specific

Comprehensive means the model includes all those elements that research has identified as associated with student achievement. *Specific* means the model identifies classroom strategies and behaviors at a granular level. Figure 1 contains 41 classroom strategies and teacher behaviors, all of which have research supporting their relationship with student achievement (Marzano, 2007).

Figure 1 includes three categories of strategies: routine strategies, content strategies, and strategies enacted on the spot. *Routines* involve five types of strategies (Elements 1–5) organized into two subcategories: those that involve communicating learning goals, tracking student progress, and celebrating success and those that involve establishing and maintaining rules and procedures.

Content strategies fall into three subcategories: those used for new content, those used when students are practicing and deepening their knowledge of new content, and those used when students are asked to apply knowledge by generating and testing hypotheses. There are 18 types of content strategies (Elements 6–23).

Strategies enacted on the spot are those that a teacher might not have planned to use in a given lesson or on a given day but that he or she must be prepared to use if needed. These strategies fall into four categories: strategies for engaging students, strategies that acknowledge adherence to or lack of adherence to rules and procedures, strategies that build relationships with students, and strategies that communicate high expectations for all students. There are 18 types of strategies enacted on the spot (Elements 24–41).

I believe these 41 elements represent the diversity of strategies that a comprehensive model of teacher evaluation should include. However, many of the 41 elements are unnecessary if the sole purpose of teacher evaluation is measurement. For example, the Rapid Assessment of Teacher Effectiveness (RATE) was designed with an explicit measurement purpose—to effectively and efficiently determine teacher competence in the classroom (Strong, 2011). The model includes only 10 categories of teacher behavior that appear sufficient to rank teachers in terms of pedagogical skill. Those categories are

- Providing clear lesson objectives.
- Understanding students' background and comfort with the material.
- Using more than one delivery mechanism.
- Providing multiple examples.
- Providing appropriate nonexamples (illustrations of the wrong way to do something).
- Maintaining an effective pace.
- Providing students with feedback about their learning.
- Engaging in timely use of guided practice.
- Explaining important concepts clearly.
- Keeping students actively engaged throughout a lesson.

Measuring teachers and _developing_ teachers are different purposes with different implications.

Studies on the RATE system indicate that it discriminates between effective and ineffective teachers much better than some popular teacher evaluation models do (Strong, 2011).

Conspicuously missing from RATE's list are references to such commonly cited elements as the teacher-student relationship and classroom management. These elements are recognized in virtually every major review of the literature on classroom correlates of effective teaching. For example, in their review of the research on 228 variables identified as having measurable relationships with student achievement, Wang, Haertel, and Walberg (1993) listed classroom management at the top. Over the years, classroom management has continued to be considered an important aspect of effective teaching (Good & Brophy, 2003). Likewise, the teacher-student relationship is prominently positioned in the theory and research regarding student behavior (Evertson & Weinstein, 2006). Indeed, Sheets and Gay (1996) identified poor teacher-student relationships as the root cause of many, if not most, discipline issues.

How does one reconcile this apparent contradiction? How could variables like management and teacher-student relationships, which have research supporting their connections to important student outcomes, not be good discriminators of teacher quality?

The answer is that these elements are important correlates with student achievement—up to a point. If a teacher has not achieved a certain level of competence in these areas, student

achievement will suffer. However, once a teacher reaches an acceptable level of competence in these areas, further skill development will not have a commensurate positive influence on student achievement.

A number of other strategy areas listed in Figure 1 correlate with student achievement but do not necessarily discriminate well between teachers who represent a wide range of competence. For example, consider academic games (Element 25), which are certainly a useful tool in enhancing student achievement (Hattie, 2009; Walberg, 1999) but only up to a certain point. Indeed, a teacher can produce dramatic gains in student learning without using games at all.

If we wished to use the model presented in Figure 1 to rapidly rate

FIGURE 1. A Model of Classroom Strategies and Behaviors

I. Routine Strategies

A. Communicating Learning Goals, Tracking Student Progress, and Celebrating Success

1. Providing clear learning goals and scales to measure these goals

2. Tracking student progress

3. Celebrating student success

B. Establishing and Maintaining Classroom Rules and Procedures

4. Establishing classroom rules and procedures

5. Organizing the physical layout of the classroom

II. Content Strategies

C. Helping Students Interact with New Knowledge

6. Identifying critical information

7. Organizing students to interact with new knowledge

8. Previewing new content

9. Chunking content into "digestible bites"

10. Processing new information

11. Elaborating on new information

12. Recording and representing knowledge

13. Reflecting on learning

D. Helping Students Practice and Deepen Their Understanding of New Knowledge

14. Reviewing content

15. Organizing students to practice and deepen knowledge

16. Using homework

17. Examining similarities and differences

18. Examining errors in reasoning

19. Practicing skills, strategies, and processes

20. Revising knowledge

E. Helping Students Generate and Test Hypotheses about New Knowledge

21. Organizing students for cognitively complex tasks

22. Engaging students in cognitively complex tasks involving hypothesis generation and testing

23. Providing resources and guidance

III. Strategies Enacted on the Spot

F. Engaging Students

24. Noticing when students are not engaged

25. Using academic games

26. Managing response rates

27. Using physical movement

28. Maintaining a lively pace

29. Demonstrating intensity and enthusiasm

30. Using friendly controversy

31. Providing opportunities for students to talk about themselves

32. Presenting unusual or intriguing information

G. Recognizing and Acknowledging Adherence or Lack of Adherence to Rules and Procedures

33. Demonstrating "withitness"

34. Applying consequences for lack of adherence to rules and procedures

35. Acknowledging adherence to rules and procedures

H. Establishing and Maintaining Effective Relationships with Students

36. Understanding students' interests and backgrounds

37. Using verbal and nonverbal behaviors that indicate affection for students

38. Displaying objectivity and control

I. Communicating High Expectations for All Students

39. Demonstrating value and respect for low-expectancy students

40. Asking questions of low-expectancy students

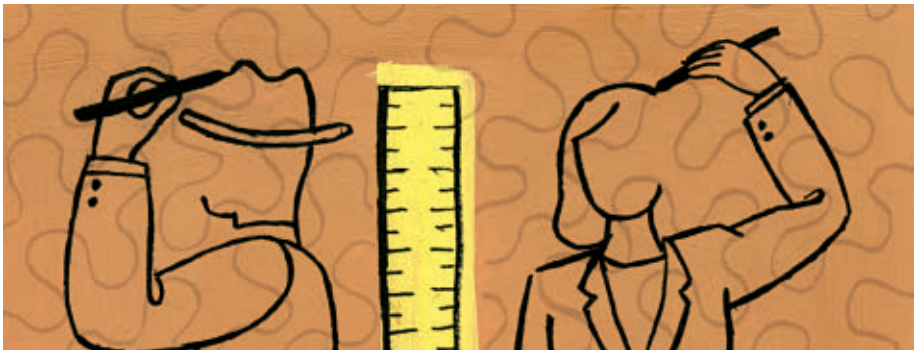
41. Probing incorrect answers with low-expectancy students

Note: Items **highlighted in blue** may be used to rapidly rate teacher competence in the classroom—that is, as a measurement tool as opposed to a development tool.

Source: From *Effective Supervision: Applying the Art and Science of Teaching* (pp. 62–63), by Robert J. Marzano, Tony Frontier, & David Livingston, Alexandria, VA: ASCD. Adapted with permission.

teachers, we'd only need to consider 15 elements (these are highlighted in the figure). In other words, if our goal is efficient measurement, following Strong's model, which appears to discriminate between teachers better than many previous models, we would need only a relatively small subset of elements and could leave out some variables that have historically been associated with effective instruction.

However, if we wished to help teachers *develop* instead of just measuring them, we'd obtain ratings on all 41 elements so teachers could identify



areas of strength and weakness and then systematically begin improving those areas of weakness. Teachers don't need to be scored on each of the 41 elements yearly. Rather, they should gradually work through the elements over time as they seek to improve their competence in the classroom.

The System Includes a Developmental Scale

A second characteristic of a teacher evaluation system that focuses on development is that it employs a scale or rubric that teachers can use to guide and track their skill development. Such a scale would articulate developmental levels, such as *not using*, *beginning*, *developing*, *applying*, and *innovating* (Marzano, Frontier, & Livingston, 2011).

At the *not using* level, a teacher is not even aware of a particular strategy or is aware of it but has not tried it in his or her classroom. For example, if a teacher were unaware of strategies for engaging students in friendly controversy (Element 30 in Figure 1), he or she would be at the *not using* level.

At the *beginning* level, a teacher uses a strategy but with errors and omissions. For example, a teacher who simply asks students to state their opinions about a topic with the goal of generating disagreement would be at the beginning level because errors and omissions are in play. Although students are, in fact, stating their opinions, they need to learn how to support their opinions using evidence and how to disagree respectfully with others.

At the *developing* level, the teacher doesn't make such mistakes. Rather, he or she uses the strategy without significant error and with relative fluency.

Although using a strategy at the *developing* level is a step in the right direction, it's at the *applying* level and above that a strategy starts to produce positive returns in student learning. At the *applying* level, a teacher monitors the class to ensure that the strategy is having its desired effect—in this case, that students are backing up their opinions with evidence and expressing disagreement in a controlled and respectful manner.

Finally, at the *innovating* level, the teacher not only monitors the class to ensure a strategy is having its desired effect with the majority of students but also makes necessary adaptations to ensure that all student populations represented in class are experiencing its positive effects. For example, to help English language learners better understand new content, a teacher might adapt a previewing strategy by using pictures downloaded from the Internet.

These five levels are designed to enable teachers (usually with the aid of a supervisor or instructional coach) to pinpoint their current level of performance for a specific strategy and set goals for operating at higher levels within a given period of time.

Contrast this scale with one designed primarily for measurement. To illustrate, consider the scale for one of the elements in the RATE system: understanding students' backgrounds and comfort with the material (Strong, 2011). This element involves three parts: intentionally sequencing the material based on knowledge of where students are in the instructional process, relating new knowledge to content that students have already mastered, and conveying to students that they are able to reach the learning goal in a manner that instills confidence.

The scale for this element involves three levels. A teacher receives a score of 1 if he or she exhibits none or only one of these elements or does a poor job trying to execute these elements. A teacher receives a score of 2 if two of the three elements are present. A teacher receives a score of 3 if all three elements are present at levels that clearly influence students in a positive way.

Although this type of scale is efficient and effective for measurement purposes, it provides little guidance to teachers, instructional coaches, or administrators regarding how to improve.

The System Acknowledges and Rewards Growth

The third characteristic of an evaluation system designed for teacher development is that it acknowledges and rewards teacher growth. In a

developmental model, each year teachers identify elements on which to improve and then chart their progress throughout the year. A teacher might select one strategy from each of the three major categories depicted in Figure 1: for example, establishing classroom rules and procedures, chunking content into digestible bites, and asking questions of students for whom he or she may have had low expectations in the past. Presumably these strategies would be ones for which the teacher was at the *beginning* or *not using* level.

The teacher would then select specific growth targets to accomplish during the year. To illustrate, assume a teacher was at the *beginning* level for all three target strategies and set a goal to reach the *applying* level on all three by the end of the year. In addition to scoring teachers on their current level of proficiency on the various elements within the evaluation model—we refer to these ratings as "status" scores—teachers would be scored on the extent to which they reached their growth goals. Attaining all three growth goals would earn the highest growth score, attaining two of three goals would earn the next highest growth score, and so on.

At the end of the year, teachers would have two scores: an overall status score and an overall growth score. Both of these scores would be considered when assigning teachers to a summative category at the end of the year—for example, *advanced*, *proficient*, *needing improvement*, or *not acceptable*. Such a system would communicate to teachers that the school expects—and rewards—continuous improvement.

The Best of Both Worlds

Both measurement and development are important aspects of teacher evaluation. When measurement is the primary purpose, a small set of elements is sufficient to determine a teacher's skill in the classroom. However, if the emphasis is on teacher development, the model needs to be both comprehensive and specific and focus on the teacher's growth in various instructional strategies. These distinctions are crucial to the effective design and implementation of current and future teacher evaluation systems. EL

An evaluation system designed for **teacher development acknowledges** and rewards **teacher growth.**

Watch the Interview

Robert J. Marzano talks with *EL* editor in chief Marge Scherer about the purposes of teacher evaluation at www.ascd.org/el1112marzano.

References

Bill and Melinda Gates Foundation. (2011). *Learning about teaching: Initial findings from the Measures of Effective Teaching project*. Bellevue, WA: Author. Retrieved from www.gatesfoundation.org/college-ready-education/Documents/preliminary-findings-research-paper.pdf

Evertson, C., & Weinstein, C. S. (Eds.). (2006). *Handbook of classroom management: Research, practice, and contemporary issues.* Mahwah, NJ: Erlbaum.

Good, T. L., & Brophy, J. E. (2003). *Looking in classrooms* (9th ed). Boston: Allyn and Bacon.

Hattie, J. A. C. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. New York: Routledge.

Marzano, R. J. (2007). *The art and science of teaching: A comprehensive framework for effective instruction.* Alexandria, VA: ASCD.

Marzano, R. J., Frontier, T., & Livingston, D. (2011). *Effective supervision: Supporting the art and science of teaching.* Alexandria, VA: ASCD.

Sheets, R. H., & Gay, G. (1996, May). Student perceptions of disciplinary conflict in ethnically diverse classrooms. *NASSP Bulletin*, *80*(580), 84–93.

Strong, M. (2011). *The highly qualified teacher: What is teacher quality and how do you measure it?* New York: Teachers College Press.

Toch, T., & Rothman, R. (2008). *Rush to judgment: Teacher evaluation in public education*. Washington, DC: Education Sector.

U.S. Department of Education. (2009). *Race to the Top program executive summary.* Washington, DC: Author. Retrieved from www2.ed.gov/programs/racetothetop/executive-summary.pdf

Walberg, H. J. (1999). Productive teaching. In H. C. Waxman & H. J. Walberg (Eds.), *New directions for teaching and practice research* (pp. 75–104). Berkeley, CA: McCutchan.

Wang, M. C., Haertel, G. D., & Walberg, H. J. (1993). Toward a knowledge base for school learning. *Review of Educational Research*, *63*(3), 249–294.

Weisberg, D., Sexton, S., Mulhern, J., & Keeling, D. (2009). *The widget effect: Our national failure to acknowledge and act on differences in teacher effectiveness.* Brooklyn, NY: New Teacher Project. Retrieved from http://widgeteffect.org/downloads/TheWidgetEffect.pdf

Robert J. Marzano (robert.marzano@marzanoresearch.com) is cofounder and CEO of Marzano Research Laboratory in Denver, Colorado. His latest book, coauthored with Tony Frontier and David Livingston, is *Effective Supervision: Supporting the Art and Science of Teaching* (ASCD, 2011).