

Making Teacher Evaluation Meaningful

Dr. David E. Bartz

Professor Emeritus

Department of Educational Leadership
Eastern Illinois University
Charleston, IL

Dr. Cliff Karnes

Chairperson

Department of Educational Leadership
Eastern Illinois University
Charleston, IL

Dr. William Allan Kritsonis

Distinguished Alumnus

College of Education and Professional Studies
Central Washington University
Ellensburg, WA

Abstract

This decade has witnessed an ongoing quest for school districts throughout the U.S. to improve teacher evaluation systems. In many districts this pursuit was prompted by the passage of state laws that emphasize student achievement (growth) and provide detailed models to be used by principals as the foundation for classroom observation ratings. This article focuses on how principals may effectively observe and rate teachers on classroom performance. We also address how to effectively apply feedback in a principal's post-observation conferencing with teachers in the context of making feedback action-oriented.

Keywords: teacher evaluation, principal classroom observations

Context

“Improving teacher evaluation is one of the most pressing and contested contemporary educational policy issues” (Cohen & Goldhaber, 2016, p. 378).

The quest in the U.S. for teacher evaluation to truly differentiate effective from ineffective teachers has existed for decades—and probably centuries. Commencing in 2010, and primarily driven by former U.S. Secretary of Education Arnie Duncan with the support of private interest groups, many states and school districts throughout the nation established new teacher

evaluation systems with a research-based framework that can incorporate student achievement (student growth) tied to uniform learning standards (Hamlin & Peterson, 2018). States and districts have also been encouraged to include student growth as part of their redesigned teacher evaluation systems because of the perceived chance of an increased likelihood of Race To The Top funding from the federal government (Neumerski, et al., 2018). The accountability of teachers for their performance is also a driving force for evaluation reform (Grissom & Youngs, 2016, p. 1).

In addressing the need to develop a new teacher evaluation system that is comprehensive, and to which student achievement growth could be incorporated, many school districts throughout the nation adopted a *framework* for determining effective teaching. Numerous districts use frameworks such as the Danielson model (2014) or the Classroom Assessment Scoring System—CLASS (Center for Advanced Study of Teaching and Learning, n.d.). Marzano's (2017) and Stronge's (2018) approaches are also credible research-based sources which may be used to develop a teacher evaluation system. These frameworks of effective teaching furnish principals with criteria, often in rubric form, for them to use for evaluating the classroom performance of teachers. Such frameworks are applicable across grade levels and content areas. The focus is also on principals effectively using the ratings flowing from observations to provide meaningful feedback to teachers in the post-observation conferences.

Factors of Effective Teaching

According to Stronge, "The elusive concept of teacher effectiveness" has been pursued for centuries (as cited in Bartz, 2018, p. ix).

Before dealing with the specifics of a particular framework for the classroom observation portion of teacher evaluation to use as a focal point, a holistic review was completed to answer the question "What are the state-of-the-art factors for teacher effectiveness?" Expert sources were reviewed to identify credible factors of teacher effectiveness. The significant factors associated with effective teaching from these expert sources are listed in Table 1. The table provides information that allows for comparisons of the content from the experts' sources.

Table 1

Major Factors Associated with Effective Teaching by Expert Sources

Experts	Factors of Effective Teaching
Stronge	<ul style="list-style-type: none"> ● professional knowledge ● instructional planning ● instructional delivery ● assessments ● learning environment ● professionalism
Danielson’s Framework of Teaching	<ul style="list-style-type: none"> ● planning and preparation (1) ● classroom environment (2) (Called Domains) ● instructing (3) ● professional responsibilities (4)
Marzano	<ul style="list-style-type: none"> ● providing and communicating clear learning targets and goals ● assessments ● direct instruction ● structured practicing, examining similarities and differences, and errors in reasoning ● conducting knowledge application lessons ● using strategies that appear in all types of lessons (e.g., reviewing content and organizing students to interact) ● using engagement strategies ● implementing rules and procedures ● building relationships and communicating high expectations ● making system changes (e.g., collaborative teaming)
Handbook of Research on Teaching	<ul style="list-style-type: none"> ● classroom organization and management ● positive emotional climate and support ● engaging and challenging instruction
Classroom Assessment Scoring System (CLASS)	<ul style="list-style-type: none"> ● <i>emotional support</i> via positive relationships among teacher and peers ● <i>classroom organization</i> via well-managed classrooms that provide students with frequent engaging learning activities ● <i>instructional support</i> through interactions that teach students to think, provide on-going feedback and support, and facilitate language and vocabulary (Bartz, 2018, p. 3)

Using the Danielson Framework as an Example for the Classroom Observation Component of Teacher Evaluation

Observations continue to be the foundation of teacher evaluations, even in newer systems that incorporate other measures of success. They are the only part of an evaluation system that nearly every teacher in every grade and subject experiences, and they are often afforded the greatest weight in determining a teacher's final evaluation rating. (The New Teacher Project, 2013, p. 1)

Context of Principals' Observational Ratings of Teachers

Cohen and Goldhaber (2016) advocate that the general perceptions of principals are that teachers' classroom performances vary significantly in effectiveness. However, observational ratings by principals often reveal that most teachers are at the highest rating possible. Kraft and Gilmour (2016, as cited in Cohen & Goldhaber, 2016, p. 383) found "that few teachers are rated below *proficient* (a level where there are consequences associated with the rating), even in states that have recently implemented major changes to their performance evaluation systems to make them more rigorous and accurate" (p. 383). Cohen and Goldhaber (2016) summarize the issue as: "One thing we do know is that it is generally difficult to implement observational performance evaluation systems that actually differentiate among teachers" (p. 384).

This lack of variability is referenced as the *Widget Effect* (Weisberg, Sexton, Mulhern, & Keeling, 2009, as cited in Cohen & Goldhaber, 2016), which is defined as: "The failure of evaluation systems to provide accurate and credible information about individual teachers' instructional performance sustains and reinforces a phenomenon that describes the tendency of school districts to assume classroom effectiveness is the same from teacher to teacher" (p. 2). In the context of rating scales, the Widget Effect is labeled as the *positive tendency effect* (Bartz & Bartz, 1995a).

In numerous instances it is evident that principals need to do a better job of differentiating between the performance levels of teachers based on classroom observations. However, because teacher observation rating scales are a criterion-type measure, it is conceivable that all teachers in a building could be performing at the highest possible rating on the measurement scale. It is also possible for none of the teachers to receive the highest rating. The starting point for principals in effectively observing teachers' classroom performances is understanding the criteria in the district's teacher evaluation system on which to base their observational ratings.

Principals Understanding the Criteria on Which to Base Classroom Observational Evidence

The four domains of Danielson's framework for teachers are: (1) planning and preparation, (2) classroom environment, (3) instructing, and (4) professional responsibilities. This article uses Domain 2—*classroom environment*—as the specific focus for principals' observation skills necessary to collect sufficient evidence of teacher effectiveness. Danielson's domains of *planning and preparation*, *instruction*, and *professional responsibilities* (Domains 1, 3, and 4) will be discussed in a forthcoming article. How principals can give meaningful feedback to teachers regarding their effectiveness in post-observation conferences is also addressed.

Figure 1 presents the components and elements for Danielson's *classroom environment* (Domain 2) and *instruction* (Domain 3). These are the two domains that serve as the basis for principals observing teachers' classroom performance. The classification system from holistic to specifics regarding the content for Danielson framework is:

Domain → Component → Element → Rating Rubrics

DOMAIN 2: The Classroom Environment

2a Creating an Environment of Respect and Rapport

- Teacher interaction with students
- Student interaction with students

2b Establishing a Culture for Learning

- (1) *Importance of content*
- (2) Expectations for learning and behavior
- (3) Student pride in work

2c Managing Classroom Procedures

- Instructional groups
- Transitions
- Materials and supplies
- Non-instructional duties
- Supervision of volunteers and paraprofessionals

2d Managing Student Behavior

- Expectations
- Monitoring behavior
- Response to misbehavior

2e Organizing Physical Space

- Safety and accessibility
- Arrangement of furniture and resources

DOMAIN 3: Instruction

3a Communicating with Students

- Expectations for learning
- Directions and procedures
- Explanations of content
- Use of oral and written language

3b Using Questioning and Discussion Techniques

- Quality of questions
- Discussion techniques
- Student participation

<p>3c Engaging Students in Learning</p> <ul style="list-style-type: none"> • Activities and assignments • Student groups • Instructional materials and resources • Structure and pacing <p>3d Using Assessment in Instruction</p> <ul style="list-style-type: none"> • Assessment criteria • Monitoring of student learning • Feedback to students • Student self-assessment and monitoring <p>3e Demonstrating Flexibility and Responsiveness</p> <ul style="list-style-type: none"> • Lesson adjustment • Response to students • Persistence <p>(Colonial Intermediate Unit 20, 2019, p. 465)</p>
--

Figure 2. The Danielson framework for teaching—classroom environment and instruction.

Component 2b—*establishing a culture of learning*—for Domain 2 (classroom environment) is used as an example for reviewing principals’ observational skills for collecting evidence to rate teachers’ performance. As noted in Figure 1, *establishing a culture of learning* focuses on three elements: (1) importance of content, (2) expectations for learning and behavior, and (3) student pride in work. *Importance of content* is depicted as an example in progression on a rating scale rubric as:

Unsatisfactory	Basic	Proficient	Distinguished
Teacher or students convey a negative attitude toward the content, suggesting that it is not important or has been mandated by others.	Teacher communicates the importance of the work, but with little conviction and only minimal apparent buy-in by the students.	Teacher conveys genuine enthusiasm for the content, and students demonstrate a consistent commitment to its value.	Students demonstrate through their active participation, curiosity, and taking the initiative that they value the importance of the content (Danielson, 2007, p. 96)

Extensive training and development are necessary for principals to master the understanding and application of classroom observational criteria in the Danielson framework. Much of the training and development for principals is available in an interactive digital format. Micro-credentialing is an applicable vehicle for certifying a principal’s mastery of content. Face-to-face training and development are still frequently used because they better allow for quick interactive “give and take” among principals and between principals and trainers.

It is a formidable task for principals to master classroom observational criteria and differentiate performance levels of teachers progressively through a rubric rating scale. To have a

fair, reliable, and valid rating system of teachers' classroom performance, principals must master differentiating those performances on a rating scale rubric. Data reinforcing their conclusions must validate their results.

Training materials for principals provide numerous examples for the 15 elements for Domain 2 (classroom environment) and the 18 elements for Domain 3 (instruction) in the Danielson framework through rating rubrics. These detailed examples represent a *behavior classification system* for principals so they can anchor observational ratings to specific evidence via observed behaviors of teachers, students—or both—collectively. Standardization factors for principals collecting teacher observational data such as time length, number of observations, time of day, and pre-conferencing are also essential to a fair system of evaluation (Pianta & Hamre, 2016, as cited in Grissom & Youngs, 2016, p. 24).

Software is plentifully available from commercial companies so principals can input observational behaviors via laptops, iPads, or other digital input devices as they make classroom observations. It can be a challenge for a principal to effectively digitally input their observations *and* visually observe students, the teacher, and contextual clues in understanding holistically what is transpiring. The principal's pre-observation conference with a teacher to be observed is key to (1) the principal gaining the context of the activities to be observed and (2) the teacher gaining an understanding of the criteria the principal will use for the observational rating.

Feedback

It is essential for principals to use effective feedback practices in a post-observation conference with teachers. Behavioral science research and modern-day management theories have consistently indicated the critical role feedback plays in behavior change. As used here, feedback is the activity of providing information to teachers about their performance in relation to what is expected based on the criteria (e.g., Danielson framework) in the teacher evaluation system's classroom observation component. Feedback addresses performance which has occurred. From the staff member's perspective, feedback answers the question, "How am I doing?" (Hillman, Schwandt, & Bartz, 1990).

Feedback can either (1) support or reinforce desirable performance or (2) indicate a need to improve performance. A principal's feedback needs to be specific and given in such a manner that the teacher feels free to interact with the principal in joint exploration of what the feedback means. Creating an environment in which the teacher can authentically discuss what the feedback means, without fear of reprisal, helps to build a trusting relationship which fosters meaningful discussion about feedback between the teacher and principal (Shea, 1984).

Generally, it is better for the principal to initially be descriptive, as opposed to evaluative, when giving teachers observational feedback. An example of evaluative versus descriptive feedback is:

Evaluative

"You do a poor job of getting students actively engaged in the learning activity."

Descriptive

"Frequently students do not appear to be actively engaged in the learning activity." (Bartz & Bartz, 1995b, p. 34)

A feedback checklist that is beneficial to principals in preparing for giving observational feedback to teachers is found in the Appendix.

Cautions Regarding Principal's Observational Ratings of Teachers

There are many cautions that principals should consider when conducting observational ratings of teachers. Some of them are:

- Teachers who have students with higher prior achievement often receive higher observation ratings (Chaplin, Gill, Thompkins, & Miller, 2014, as cited in Cohen & Goldhaber, 2016, p. 381; Steinberg & Garrett, 2016, as cited in Cohen & Goldhaber, 2016, p. 381; Whitehurst, Chingos, & Lindquist, 2014, as cited in Cohen & Goldhaber, 2016, p. 381).
- Teachers' instructional approaches differ depending on their students' needs. Hence, responsive teaching would likely vary, depending on a teacher's students. This variance is sometimes at odds with the standardization of quality practice underlying observational instruments (Cohen & Goldhaber, 2016, p. 381).
- Good teaching likely varies in response to contextual factors, including school and district leadership, curricula, and collegial support (Little, 2001, as cited in Cohen & Goldhaber, 2016, p. 381; McLaughlin & Talbert, 2006, as cited in Cohen & Goldhaber, 2016, p. 381).
- Research has demonstrated that raters struggle to keep multiple dimensions of quality in mind during observations and that content-specific aspects of instruction are especially cognitively demanding and subject to rater biases (Bell et al., 2014, as cited in Cohen & Goldhaber, 2016, p. 382; Park, Chen, & Holtzman, 2014, as cited in Cohen & Goldhaber, 2016, p. 382).
- Despite extensive efforts to train raters (principals), check for scoring quality over time, and provide ongoing feedback on scoring (often termed *calibration*), a two-year study revealed that there was still substantial *drift* or movement away from master scored lessons (Casabianca, Lockwood, & McCaffrey, as cited in Cohen & Goldhaber, 2016, p. 383).
- Minimizing *rater effects* requires minimizing the more subjective biases that different observers (principals) bring to observations. A good first step is establishing a rigorous system for training raters and ensuring that new raters' scores correspond with those given by "expert" or master raters, which is often called *certification* (Cohen & Goldhaber, 2016, p. 383).
- Systems developed to minimize rater bias based on the characteristics of the rater, teacher, and classroom were useful in the context of video observations when *raters were not personally connected to the teachers* (Park et al., 2014, as cited in Cohen & Goldhaber, 2016, p. 383).
- Because principals have existing relationships with the teachers they observe, and also have multiple competing demands on their time, they may make different strategic decisions about rating teachers that result in less accurate scores (Cohen & Goldhaber, 2016, p. 383).

Closing Thoughts

Teacher effectiveness frameworks such as Danielson's and Classroom Assessment Scoring System's (CLASS) that delineate criteria for principals' classroom observational ratings of teachers can be complex, *data-intensive*, and time demanding for principals. Principals must be provided with ongoing training and development to fully understand the classroom observational criteria, know how to collect data and assign ratings accurately, and give meaningful feedback to teachers. Because principals have extreme demands on their time, it is imperative that they establish—and carry out—a meticulous schedule for completing all required teacher classroom observational activities and providing appropriate feedback to their teachers. Principals must guard against the demands of implementing a teacher evaluation system from depleting their energy and leading to job burnout (Superville, 2018).

School district personnel must provide teachers with the same rigorous training and development they utilize to understand the criteria on which classroom performance will be judged. In that it is logical to assume the criteria used for teachers' classroom observations represent behaviors indicative of effective teaching, it is also logical to assume those teachers receiving high ratings will enhance students' learning. There must be a constant focus by principals to ensure the enhancement of learning for each student.

References

- Bartz, D. E., & Bartz, R. H. (1995a). Improve the summative evaluation conference. *Principal*, 74(4), 24-26.
- Bartz, D. E., & Bartz, R. H. (1995b). Using feedback and coaching to enhance the performance of staff members. *Illinois Schools Journal*, 74(2), 29-41.
- Bartz, D. E., & Quick, J. (2006). Effective use of feedback to enhance staff members performance. *Perspective*, (August-October), 6-7.
- Bartz, D. E. (2018). Factors of effective teaching. *National Forum of Teacher Education Journal*, 28(3), 1-10.
- Center for Advanced Study of Teaching and Learning. (n.d.). *Measuring and improving teacher-student interactions in PK-12 settings to enhance students' learning*. Charlottesville, VA: Curry School of Education, University of Virginia. Retrieved from <http://www.curry.virginia.edu/castl>
- Cohen, J., & Goldhaber, D. (2016). Building a more complete understanding of teacher evaluation using classroom observations. *Educational Researcher*, 45(6), 378-387.
- Colonial Intermediate Unit 20. (2019). Charlotte Danielson's framework for teaching [professional development seminar]. Retrieved from <https://www.ciu20.org/Page/465>
- Danielson, C. (2007). *Enhancing professional practice: A framework for teaching* (2nd ed.). Alexandria, VA: ASCD.
- Danielson, C. (2014). *The framework for teaching: Evaluation instrument* (2013 ed.). Retrieved from <http://www.danielsongroup.org>
- Danielson, C. (2016). *Talk about teaching! Leading professional conversations* (2nd ed.). Thousand Oaks, CA: Corwin.
- Gitomer, D. H. & Bell, C.A. (Eds.). (2016). *Handbook of research on teaching* (5th ed.). Washington, DC: American Educational Research Association.

- Grissom, J. A., & Youngs, P. (Eds.). (2016). *Improving teacher evaluation systems: Making the most of multiple measures* (pp. 1-7). New York, NY: Teachers College Press.
- Hamlin, D., & Peterson, P. (2018). Have states maintained high expectations for students? *Education Next*, 18(4), 42-49.
- Hillman, L. W., Schwandt, D. R., & Bartz, D. E. (1990). Enhancing staff members performance through feedback and coaching. *The Journal of Management*, 19(3), 29-40.
- Marzano, R. J. (2017). *The new art and science of teaching*. Bloomington, IN: Solution Tree Press.
- Neumerski, C. M., Grissom, J. A., Goldring, E., Rubin, M., Cannata, M., Schuermann, P., & Drake, T. A. (2018). Restructuring instructional leadership: How multiple-measure teacher evaluation systems are redefining the role of the school principal. *The Elementary School Journal*, 119(2), 1-24.
- Shea, G. F. (1984). *Building trust in the workplace*. New York, NY: American Management.
- Stronge, J. H. (2018). *Qualities of effective teachers* (3rd ed.). Alexandria, VA: ASCD.
- Superville, D. R. (2018). Teacher evaluations have dramatically changed the principal's job. *Education Week*. Retrieved from <https://www.edweek.org/ew/articles/2018/11/14/teacher-evaluations-have-dramatically-changed-the-principals.html>
- The New Teacher Project. (2013). *Fixing classroom observations*. New York, NY: Author.

Appendix

Feedback Checklist

Feedback should:

- _____ (1) Occur as soon as possible.
- _____ (2) Allow ample time for discussion.
- _____ (3) Be done in a private setting.
- _____ (4) Be given only when the staff member or principal is not upset, frustrated, or tired.
- _____ (5) Initially be descriptive as opposed to evaluative.
- _____ (6) Be specific.
- _____ (7) Focus on behaviors and not personality.
- _____ (8) Demonstrate interest in the teacher.
- _____ (9) Use factual information.
- _____ (10) Use open-ended and probing questions.
- _____ (11) Be given in relation to observational criteria.
- _____ (12) Demonstrate effective preparation by the principal.
- _____ (13) Ensure that the principal and teacher each understand what has been discussed.
- _____ (14) Reflect adequate collection of information and preparation by the principal.
- _____ (15) Encourage input from the teacher.
- _____ (16) Address both effective performance and areas for needed improvement.
- _____ (17) Note subsequent steps that may need to be taken for a performance issue, if it is not adequately addressed.
- _____ (18) Make use of effective verbal and non-verbal communication skills.
- _____ (19) Encourage self-evaluation by the teacher.
- _____ (20) Respect the dignity and opinions of the teacher.
- _____ (21) Exhibit a trusting climate by the principal.
- _____ (22) Include suggestions, resources, information, and timelines to address areas of needed improvement.
- _____ (23) Address professional development (learning) activities for:
(a) areas of needed improvement and (b) talent development for career growth and interest (Bartz & Quick, 2006).