Walkthroughs, Rubrics, and Teacher Evaluation—Do They Improve Teaching?

By Jon Saphier

Schools that produce better student learning follow one common route: they pay unrelenting attention to the improvement of teaching and learning in the classroom. What the individual teacher knows, believes, and can do dwarfs all other variables affecting student achievement. Research replicated all over the country since Sanders and Rivers first blockbuster study (1995) confirms that finding. Therefore the improvement of teaching must be at the center of the leader's job, and the leader must be knowledgeable about what it looks and sounds like.

Unfortunately principals cannot spend all their time directly with teachers on the improvement of classroom teaching (though many would like to.) To compensate, there are powerful indirect ways principals can influence the quality of teaching:

- Focus on how the **Instructional Leadership Team** defines itself and spends it's time.
- Develop operation, time use, and interpersonal norms of teams that share content and meet regularly. These teacher teams are often called **Professional Learning Communities (PLCs)** these days. Ensure they know how to do error analysis and design of re-teaching.
- Emulate the best models for the design and operation of the **instructional coach's job** and the coach's relationship with the principal.[1]
- Do **planning conferences** with teachers rather than simple pre-conferences. Planning conferences, in fact, can yield more productive improvements than observations and feedback.[2]

In this piece, however, I want to focus on the principals' direct contact with teachers using such practices as walkthroughs and short visits. Specifically this article asks: 1) Can walkthroughs and short visits produce useful information for teachers that leads to improved practice? 2) Should we rely on written evaluations to provide important input for a teacher's growth? 3) Should teacher evaluation rubrics be used in formal summative teacher evaluations?

All educators need feedback to hone their craft; observation by someone who knows good instruction when they see it should be a prime source for improvement for any classroom teacher. But the best place to start addressing the need for good feedback is not with a procedure like walkthroughs. The place to start is ensuring that every school has leadership personnel who are deeply knowledgeable about good instruction and skillful at communicating about it with colleagues. Any structures and procedures for observation depend on the level of expertise of the observer. Observers must have sufficient time to work with teachers. The principal should certainly be one such person, but especially in a large school, one person is not enough, and the principal will never have enough time, even in a modest size school, to provide all the stimulation and feedback teachers need to reach peak performance.

Principals themselves, as it happens, are rarely sufficiently prepared in classroom observation and analysis. My data for this generalization is that over the past 27 years my colleagues and I have taught courses on observing and analyzing teaching to over 10,000 school administrators all over the country who evaluate teachers. We have found that the

[1] See "How School Leaders Create High-Functioning Teams that do Error Analysis and Plan Re-teaching--redefining instructional leadership" Jon Saphier. Research for Better Teaching 2008. "How to Build a High Functioning Corps of Building-Based Coaches" (Saphier and West 2008) plus books on creating PLCs of teachers who teach the same content and use data to focus teaching and plan re-teaching (DuFour et al. 2005; Love et al. 2008)

[2] See The *Skillful Teacher*, Saphier et al, chapter 17, Planning.

incoming skills of administrators tend to be quite low. This is not a blame statement. Administrators have had little or no preparation for analyzing instruction when they are hired, and almost never had to show they could do so to get their jobs. This is a pity, because the principal above all, as the instructional leader of the building, *needs* to be knowledgeable and skillful at recognizing good teaching and learning and at coaching teachers to the next level, even teachers who are already capable. Developing this instructional knowledge and skill in principals is currently a giant void in their preparation. What's the use of using observation protocols and walkthroughs if the walkers aren't highly developed and reasonably consistent in understanding what they see? A piece of paper in their hand or a checklist does not convert a person into a skilled observer.

Filling this void by providing good training at rigorous standards to principals will be important but insufficient by itself: principals can never reach sufficient numbers of teachers by individual supervision or by themselves, no matter how good they are at it. But the principal *does* need to be deeply knowledgeable about good teaching so he/she can mobilize the people and create the structures and relationships in the school that *will* improve the teaching of every teacher.

Given this situation, several related questions emerge:

1) How often should principals and others do walkthroughs and short visits, do observational write-ups, and use rubrics for evaluating teaching? Suppose principals had a well-developed eye for recognizing good teaching, how valuable would walkthroughs and write-ups be?
2) If their value is marginal, which I am going to argue, then what is more useful? And who should do the work?

Walkthroughs

The problem with walkthroughs is they are easily degraded into superficial acts. When done competently, they can serve useful purposes, which I will describe below, but even at their best there are important questions about teaching and learning that can never be answered in walkthroughs because the visits are too short, questions like the following:

- **What is the objective of the lesson as stated?** Too many principals seek to find the answer to this question on the board, or listen to what is stated to the students, or written in the plan book. This is what I mean by superficial. In order to assess the real or *lived* objective, i.e., the actual objective that is being worked on given what is actually taking place for students, the observer must listen to what the teacher and the students are doing for more than a few minutes.

- **What is the *worthy* objective, i.e., what is the important thing students should be learning given what is called for in the curriculum and especially what is available in the materials they are using?** This is hard to assess in a quick visit without examining the materials carefully by quick-reading them, doing a few problems, or asking yourself questions about the content and what is significant or interesting about it.

These three objectives—stated, lived and worthy—should all be the same. If they're not, there is a problem. These objectives can be assessed by a skilled observer in a moderate amount of time, but not a 5minute walkthrough.[3]

[3] For examples, see *The Skillful Teacher*, chapter 16, Objectives.

Other vital questions walkthroughs can't answer:

- **Can doing the activities the students are doing logically lead to learning what they are supposed to learn?**

- **Do the students know what the objective is?**

One way to find out is to ask some of them, "What are you supposed to learn or get better at today?" Follow up with clarifying and extending questions if they seem to know but haven't answered completely. This is likely, and it will require more than five minutes.

- **Do the students know what the criteria are for good work or for mastery of the content?** Again, this could be assessed through asking, "How will you know if you've done this well?" suggesting a clock ticking past five minutes.

- **Can you see that the teacher is going to get some data/evidence (by collecting class work samples, a tally of who could answer questions,…) about how well the students are learning what they are supposed to be learning?** Probably not if you don't stay for a larger chunk of the lesson or conference afterwards?

- **How well *are* the students doing?** This implies taking the time to look over shoulders, examine student work, listen to conversations if students are in groups. Are there any particular confusions evident?

- **Is the lesson based on what students know and can do or is the teacher mechanistically following a set of curriculum materials? What prior knowledge do they need in order to do the tasks they are asked to do? Do they have that prior knowledge?**

These are fundamental issues that are at the center of productive discussion between educators. Instruction doesn't improve if these questions are not on the table for examination along with evidence of student learning.

Walkthroughs first came to prominence in the 80s and 90s in District 2 in New York City. Anthony Alvarado, the Superintendent at the time, wanted principals to be in classrooms often; he used the walkthrough *with* the principal during his school visits to signal the importance of instruction to his principals and teachers. He used monthly professional development sessions with all principals (which were called "Principals Conferences,") led by literacy and math curriculum directors to teach the principals what to look for to know when good literacy and math instruction were taking place. Principals were expected to practice these walkthroughs on their own, regularly.

Surely some productive individual conversations between principals and teachers were byproducts of this practice. But the main reason for the practice was symbolic. The message to principals was: signal by where you show up and what you do that classroom instruction is important! And show also that you are curious and always learning more about good instruction by doing observations *with the literacy and math coaches* in your own building so they can assist you in determining what you look for and how to support improvement. Such co-observations with coaches are also an important model to staff of constant learning by principals.

Walkthroughs also facilitated attaining common expectations for instructional practices associated with good math and literacy instruction like Word Walls, Shared Reading and many others. Without heavy handed evaluation, certain instructional regularities could be spread more quickly. But the capacity of Walkthroughs to deepen compliance by teachers with agreements about certain instructional practices, while a start, is

still relatively superficial. The presence of a Word Wall doesn't say anything about how well or how consistently it is used. The occurrence of "shared reading" doesn't say anything about how well the students are engaged with the content of the passage, how appropriate it is for this particular class, or whether the significant elements of the passage are discussed with students.

In District 2, which became famous as an urban district that improved student achievement significantly, substantive improvement of teaching and learning itself depended on the intensive work of coaches, not the light touch of walkthroughs. Coaches focused on developing *planning* expertise among all teachers through focusing first on the planning skills of lead teachers, and then, through these lead teachers, developing a culture with groups of teachers for careful planning of lessons together, teaching in front of one another, and then non-defensive critique and examination of practice together.

So should walkthroughs be discarded as a superficial fad? No. They are still useful symbolically and also for gathering certain kinds of information about patterns in classrooms. And when groups of teachers in the same building do them together they can be a powerful vehicle for making teaching more public and for surfacing gross patterns, which are visible in five-minute visits and may need addressing. For example, perhaps we see little dialog between students in which they can make their thinking visible to one another to deepen learning. In addition, walkthroughs can focus on single aspects of instruction picked because a staff wants a baseline to evaluate how it is doing on improving it: focuses like "checking for understanding," "feedback to students," "climate of mutual support among students." For a fine summary of a variety of walkthrough protocols and purposes, see Chapter 7 of *Skillful Leader II* by Platt, Tripp et al, 2008. But the really significant questions about what is going on for students and how well learning is proceeding cannot be asked or answered in five-minute visits. Individual principals may use walkthroughs and short visits to have short conversations and raise good questions for teachers when they exercise the discipline to get into classes often, as Kim Marshall recommends (Marshall 2008.) This will be useful, but will in no way leverage the large scale improvement of instruction we need in most all our schools, especially low performing urban schools.

Short Visits

Short visits by principals who are skillful observers are different from walkthroughs. They are more open-ended since they are not bound by walkthrough protocols that may prescribe what to look for on a given visit. If principals can discipline themselves to do several quick visits daily, they send the message to teachers that the principal believes that what the teacher is doing daily with students is important. If the principal makes most of them "good news" visits with authentic good things he/she has observed, the climate can be non-threatening. And if the principal becomes skilled at framing good reflective questions for teachers based on data from the visit, a climate of inquiry about teaching and its effects can be nurtured. Principals also begin to get a useful sense of what is going on for students, day to day in classrooms. They cannot be leaders for improving teaching if they are locked away in their offices all day. But the powerful vehicles for "raising all boats" in quality teaching and learning lie elsewhere.

We have made the case (Saphier and West, 2008) that the principal and the leadership team of the school should invest their time in 1) partnering with the instructional specialists/coaches in their building so that the teachers have frequent, high-quality conversations about practice, 2) building a workplace culture of continuous dialog, inquiry, and refinement of teaching practice

based on evidence of student results, and 3) making longer visits (15 to 20 minutes or a full period) and asking the important questions of what's going for students to prepare for meaty discussions with teachers.

Write-ups of Classroom Observations

Write-ups have low to medium leverage on influencing teaching practice; they take a long time to do and one can't manage to do them very often. Frequent high-quality conversations, however, with a skillful observer who has evidence about what went on and how it is impacting students can be immensely valuable to teachers. We should focus on that. Therefore principals and also coaches and assistant principals need to become expert at observing, at gathering evidence in notes, and at conferencing with teachers.

Principals, however, do and should do write-ups for formal evaluations, and they *have* to do them to document problems when a teacher is in "improvement" and might be dismissed. When write-ups must be done, they should be done very well, have a healthy balance of claims, evidence, impact and judgments so they create a credible picture of what happened in the class, especially the cause and effect connections between what the teacher is doing and the students are learning.

A **claim** is a generalization about a person's teaching. For example,

"Mr. Powers sticks with students and makes sure they understand hard concepts."

Evidence is a literal piece of data, a quote or literal description that backs up or illustrates the claim.

"Mr. Powers re-explained osmosis and Maria said she understood. "I don't want to drop you just yet, Maria. What would be the next step?"

An **impact statement** explains what was accomplished by the event cited as evidence. It's the "effect" side of a cause and effect equation, that is, the impact on student learning.

"Thus Maria had to show she could apply the concept of osmosis that she had just said she now understood; and Mr. Powers got confirmation she did."

A **judgment**, though not always needed explicitly, is a clear statement of what the observer thought of the behavior.

"This is typical of the thorough checking for understanding and the commitment Mr. Powers has to all students mastering the material[4]."

Write-ups must convincingly support the conclusions, judgments, and recommendations the evaluator is making. It takes considerable training for observers to develop this kind of writing skill. The point of the training, however, is not just the writing; that is a by-product, a secondary effect. The point of the training is to develop a common language and concept system for talking about teaching together, and use it to spot what is important in the class and analyze its significance. As these acute lenses for seeing develop, observers must also learn how to capture classroom events in literal notes, and to talk productively with the teacher about it afterward in a way that is evidence based and productively points toward actionable improvement.

The center of the work our organization has done for 30 years is learning how to analyze the significance of what goes on in class. "Claims, evidence, impact, and judgments" (C, E, I, J) is a way of thinking; it is an evidence-based framework for analyzing teaching and learning. C, E, I, Q where the "Q" stands for "question" is the more frequent behavior we

4 For more examples see Saphier, Jon. *How to Make Supervision and Evaluation Really Work*. Research for Better Teaching: Acton, MA. 1993.

expect our participants to bring from our training to their interaction with teachers. Questioning is a prime lever on the improvement of teaching: the skill of sharing literal data and asking questions. "I noticed that you came back to Isaac three times during the seatwork time, yet he seemed to be working well. What was your thinking?"

In training experiences for evaluators, one should evaluate participants ability to write about teaching with claims and evidence, because it is proof they can do the data gathering and thinking necessary to be good analysts of teaching; these skill are the foundation for good conferences. And, of course, they are also necessary for producing good documentation in written evaluations.

Rubrics for Teaching

Rubrics take criteria for "goodness" in a performance and spell out levels of quality, usually from 1 (lowest) to 4 (highest) on each criterion. Thus rubrics create clear images of what a high quality performance looks like in a field. In this way rubrics are useful to school people in profiling what we want in teaching performance. Good rubrics have been created by the State of California and published by the Beginning Teacher Center, University of California at Santa Cruz, and also separately by Charlotte Danielson and by Kim Marshall. It is a mistake, however, to use these rubrics to score a classroom observation or to summatively evaluate a teacher's overall performance.

First, scores on rubrics get added up, and teachers look at their "grade" (how many 4s, how many 3s. how many 2s) just like students do. The presumption of "scientific rigor" based on the presumed validity of numbers leads to unfounded confidence in marginal scores between teachers and unproductive comparisons. ("She got a 48 overall and I got a 45. He must think she's better.") It also leads to focus on scores rather than substance. One does not need a number to bring a teacher's attention to a gap in their skill or a problem in student learning. Just say what the problem is, and show the evidence that led you to think so. The more serious the problem is with teaching and learning, the more you need to write about it and provide good evidence. But calling it a "2" on a rubric does not make the case more forceful or clear.

Second, and more serious, we do not have the assessment technology to make the fine distinctions between teachers that these rubrics call for. For example, on one of the published rubrics, in the Domain of "Monitoring, Assessment, and Follow-Up," a score of 4 is: "Continuously checks for understanding, unscrambles confusion, gives specific, helpful feedback." A score of 3 is: "Frequently checks for understanding and gives students helpful feedback if they seem confused."

What's the difference? What evidence would warrant a definitive "4" instead of a "3?" What's the difference between "specific, helpful feedback" and just "helpful feedback?" Feedback can't be helpful without being specific.

What's the difference between "continuously" and "frequent?" If checking for understanding is frequent and well interpreted and responded to by the teacher (presumably a "3",) it may be more effective than checking for understanding that is continuous but not effectively analyzed and used (possibly a "4.")

If the difference between a 3 and a 4 is that a 4 "unscrambles confusions" and a 3 doesn't, how can the evaluator tell? Maybe the unscrambling isn't going to happen until a re-teaching session the next day. A small number of observations in the routinized process of teacher evaluation will not produce enough raw evidence for even the most skilled evaluator to make the fine distinction called for between a 3 and a 4.

Here's another example: how can an evaluator on the basis of a few observations distinguish between a teacher who "Is an expert in the area and has a cutting-edge grasp about how student learn it best, " a 4, and a teacher who "Knows the subject matter well and has an up-to-date grasp of how students learn it best," a 3?
Tell me what you would take as evidence to make this distinction and what you would have to do to gather this evidence and convince a third party you were objectively and definitively correct in ranking two teachers differently!

Rubrics are useful, in fact important for generating a shared image of what good teaching and learning looks like and sounds like, but are not valid or reliable tools for evaluating actual teachers' teaching. The "4" point on a rubric is the most useful, because it grounds a standard of excellence in practice and invites debating what it means, what it looks and sounds like in real practice. That in turn can lead to teacher engagement and specificity about what teaching behaviors are expected.

Investing in Human Capacity

In conclusion, let's put walkthroughs, short visits, rubrics, and formal teacher evaluation in perspective. They are tools that can play a small and useful part in larger, more comprehensive efforts to improve teaching and learning. Large-scale improvement of instruction is the goal we need to set out sights on; our children need that. But the foundation for accomplishing that improvement is investing in people, not procedures and structures. The starting point of investing in people is developing school based leaders who can recognize good teaching and learning, be articulate about it, and skillful at developing teachers' capacity to use it. This happens one classroom and one school at a time, but only at a large enough scale if the district and school board have human capacity building as a commitment and then a policy that brings that commitment to life. We'll get a lot more bang for our bucks if we spend our time and resources on growing high functioning professionals instead of procedures, evaluations, and teaching rubrics.

References

DuFour, Rick and Beck DuFour and Robert Eaker, Tom Many. *Learning by Doing: A Handbook for Professional Learning Communities at Work*. Solution Tree: Bloomington, IN 2007.

Danielson, Charlotte. *Enhancing Professional Practice: A Framework for Teaching,* 2nd ed., Association for Supervision and Curriculum Development Alexandria, VA 2007

Love, Nancy, Katherine E. Stiles, Susan Mundry, Kathryn DiRanna. *The Data Coach's Guide*. Research for Better Teaching, Acton, MA 2008.

Marshall, Kim. "Is Supervising the Heck out of Teachers the Answer?" Education Week, May 7, 2008.

Marshall, Kim. "The Why's and How's of Teacher Evaluation Rubrics." Edge Magazine. Oct. 2006.

Professional Teaching Standards. New Teacher Center, University of California, Santa Cruz. Adopted Jan 1997 by California Dept. of Education.

Saphier, Jon, MaryAnn Haley-Speca, Robert Gower. *The Skillful Teacher*. Acton, MA: Research for Better Teaching, 2008.

Saphier, Jon and Lucy West. "How to Establish a High-Functioning Corps of Building-Based Instructional Coaches that Increases Student Learning." Research for Better Teaching, Acton, MA 2008

Sanders, W.L. and Rivers, J.C. "Cumulative and Residual Effects of Teachers on Future Student Academic Achievement." Knoxville, TN: University of Tennessee Value-Added Research and Assessment Center, 1996.