Administrator Academy #1448
Teacher Evaluator Competency Skill Building
Assessing Danielson Domains 2 & 3

Materials Developed by:
Dr. Catherine Siele Berning
Dr. Diane Cody
As the rubber hits the road in the implementation of states’ revamped teacher-evaluation systems, new research illuminates a troubling source of bias. School principals—when conducting classroom observations—appear to give some teachers an unfair boost based on the students they’re assigned to teach, rather than judging them solely on their instructional savvy.

Observers tended to give the best marks to teachers whose students already were high-performing, while those teachers working with academically struggling students were penalized, according to an analysis of thousands of observation scores. The report, released today by the Brown Center on Education Policy at the Brookings Institution, a Washington think tank, raises a host of new concerns about the nation’s evolving systems for grading teachers. And it suggests that, in trying to manage the technical and political challenges posed by test-score-based approaches to evaluation, such as “value added” methods, policymakers may be missing problems in other features of the systems. “It’s very worrisome. It’s a huge bias,” said Grover J. “Russ” Whitehurst, the director of the Brown Center. “The criticism about value-added is certainly something we need to attend to, but a lot of work has helped reduce or eliminate that bias. None of that’s being done for observation scores.”

The report recommends that districts try to level the playing field by adjusting teachers’ observation scores based on the demographics of the students they instruct. Among other things, the report also recommends scrapping policies that permit teachers to be judged based on the progress of all students in the school.

District Data
Spurred largely by federal efforts, such as the Race to the Top competition, dozens of states rushed to introduce new teacher-evaluation systems, most based on a combination of test scores and classroom observations.
Both in news stories and in research annals, most of the ink spilled on teacher evaluation has focused on value-added approaches, which estimate each teacher’s ability to boost his or her students’ standardized-test progress. Observations, in which administrators visit classrooms and rate the quality of a teacher’s instruction against a framework, are comparatively understudied.

For their analysis, Mr. Whitehurst and two colleagues examined teacher-evaluation data from four urban school districts ranging in size from 25,000 to 110,000 students. They looked at one to three years of data, analyzing the relationships among the various evaluation components, teachers’ overall scores, and the demographics of the students they taught. The specific weights assigned to each component varied across the districts, but classroom observations counted for at least 40 percent of each teacher’s overall score. In most cases, it was more heavily weighed: More than three-quarters of the teacher sample taught in grades or subjects not assessed with standardized tests, and for such teachers, observations are typically weighted more heavily.

Overall, the researchers found that the components’ technical properties were consistent with those studied in the massive Measures of Effective Teaching study sponsored by the Seattle-based Bill & Melinda Gates Foundation. But they also discovered some troubling patterns.

For one, the researchers found a strong statistical link between teachers’ observation scores and the achievement levels of the students they instructed. Just 9 percent of teachers of the lowest-achieving students received a top observation score, for example, while 29 percent of such teachers received a ranking in the bottom 20 percent. By contrast, 37 percent of teachers of the highest performing students got a top observation score, and only 11 percent received the lowest score. News reports indicate that teachers in some districts, including the school system for the District of Columbia, have fretted about similar patterns.

Fix Proffered
The cause of the bias isn’t examined in the report, but the authors surmised that some often-measured teaching skills—such as leading a discussion with lots of questioning—may be more difficult with students who are underprepared or not fluent in English.

And the study suggests that the problem may be fixable. The authors applied a handicap of sorts based on student demographics, giving a boost to teachers with many lower...
performing students and depressing the scores of those with students who tend to score well. That method would more evenly distribute observation scores across teachers of different groups of students, the paper shows.

Such an idea could be controversial for states and districts to implement because of its assumptions about how much various subgroups of students can progress. But without it, Mr. Whitehurst contends, teacher-evaluation systems may have unintended consequences—such as making working with the neediest students less attractive for teachers.

“Either we must have observations designed to be immune from this kind of bias, or we have to adjust for it,” Mr. Whitehurst said. “I don’t see any other way out, if we want teachers to teach where we need them to teach, and to be valued for what they do.”

Other observers said that the new research adds yet another question mark to the contested policy push for revamped teacher evaluations.

“I think this is going to be another bad-news story for the supporters of teacher evaluation, and for the [Obama] administration,” said Michael Petrilli, the executive vice president of the Thomas B. Fordham Institute, a Washington think tank. “Rather than trying to find some kind of technocratic solution, we need to get back to common sense—trusting principals to make judgments. If we don’t do that, none of our school-reform efforts are going to work.”

Teachers’ unions have tended to be far more critical of the value-added approach than classroom observations.

Segun Eubanks, the director of teacher quality for the 3 million-member National Education Association, said that, on the one hand, the new analysis from Brookings confirms a general sense among teachers that they’re put at a disadvantage by choosing to work with the most at-risk students.

On the other hand, the notion of adjusting observation scores seems premature without further research, he said.

“My first instinct would be to help to put observational data into a context-specific realm. You need to train folks to see what teaching performance looks like when you’re teaching students who have low achievement,” Mr. Eubanks said. “The look-fors are different; the way the standards are applied are different. We have to find ways to do that before we start going for handicapping.”
Schoolwide Gauge Questioned
The Brookings report also examines a host of other aspects of the newly designed teacher-evaluation systems.

Consistent with the Gates research findings, it suggests that more observations improve the accuracy of the systems, and that outside observers tend to give ratings that are more predictive of teaching quality than principals.

Also, the report takes aim at evaluation systems that use a “schoolwide” value-added measure, in which all teachers are judged in part on the progress of the school as a whole. Such a policy, the report notes, tends to bring down the scores of even good teachers in schools with lots of low-achieving students—and to inflate the scores of weaker teachers who were in high-performing schools.

“It creates a system that is demonstrably and palpably unfair to teachers, given that they have little control over the performance of the whole school,” the report states.

Vol. 33, Issue 32, Pages 1,10-11

What do you think is a solution to classroom-observation bias?
Pocket Guide to Probing Questions

Developed by Gene Thompson-Grove, Edorah Frazer, Faith Dunne and further revised by Edorah Frazer.

The distinction between clarifying questions and probing questions is very difficult for most people working with protocols. So is the distinction between probing questions and recommendations for action. The basic distinctions are:

**Clarifying Questions** are simple questions of fact. They clarify the dilemma and provide the nuts and bolts so that the participants can ask good probing questions and provide useful feedback later in the protocol. Clarifying questions are for the participants, and should not go beyond the boundaries of the presenter’s dilemma. They have brief, factual answers, and don’t provide any new “food for thought” for the presenter. The litmus test for a clarifying question is: Does the presenter have to think before s/he answers? If so, it’s almost certainly a probing question.

*Some examples of clarifying questions:*
- How much time does the project take?
- How were the students grouped?
- What resources did the students have available for this project?

**Probing Questions** are intended to help the presenter think more deeply about the issue at hand. If a probing question doesn’t have that effect, it is either a clarifying question or a recommendation with an upward inflection at the end. If you find yourself saying “Don’t you think you should ...?” you’ve gone beyond probing questions. The presenter often doesn’t have a ready answer to a genuine probing question. Since probing questions are the hardest to create productively, we offer the following suggestions:

- Check to see if you have a “right” answer in mind. If so, delete the judgment from the question, or don’t ask it.
- Refer to the presenter’s original question/focus point. What did s/he ask for your help with? Check your probing questions for relevance.
- Check to see if you are asserting your own agenda. If so, return to the presenter’s agenda.
- Sometimes a simple “why...?” asked as an advocate for the presenter’s success can be very effective, as can several why questions asked in a row.
- Think about the concentric circles of comfort, risk and danger. Use these as a barometer. Don’t avoid risk, but don’t push the presenter into the “danger zone.”
- Think of probing questions as being on a continuum, from recommendation to most effective probing question. For example [on next page — from an actual Consultancy session in which a teacher was trying to figure out why the strongest math students in the class weren’t buying in and doing their best work on what seemed to be interesting math “problems of the week”]:
  1) Could you have students use the rubric to assess their own papers? (recommendation re-stated as a question)
2) What would happen if students used the rubric to assess their own work? (recommendation re-stated as a probing question)

3) What do the students think is an interesting math problem? (good probing question)

4) What would have to change for students to work more for themselves? (better probing question)

In summary, good probing questions:
- are general and widely useful
- don’t place blame on anyone
- allow for multiple responses
- help create a paradigm shift
- empower the person with the dilemma to solve his or her own problem (rather than deferring to someone with greater or different expertise)
- avoid yes/no responses
- are usually brief
- elicit a slow response
- move thinking from reaction to reflection
- encourage taking another party’s perspective

Some final hints for crafting probing questions. Try the following questions and/or question stems. Some of them come from Charlotte Danielson’s Pathwise work, in which she refers to them as “mediational questions.”

- Why do you think this is the case?
- What would have to change for…?
- What do you feel is right in your heart?
- What do you wish…?
- What’s another way you might…?
- What would it look like if…?
- What do you think would happen if…?
- How was…different from…?
- What sort of an impact do you think…?
- What criteria did you use to…?
- When have you done/experienced something like this before?
- What might you see happening in your classroom if…?
- How did you decide/determine/conclude…?
- What is your hunch about…?
- What was your intention when…?
- What do you assume to be true about…?
- What is the connection between…and…?
- What if the opposite were true? Then what?
- How might your assumptions about…have influenced how you are thinking about…?
- Why is this such a dilemma for you?

Some Examples of Probing Questions:
- Why is a “stand-and-deliver” format the best way to introduce this concept?
- How do you think your own comfort with the material has influenced your choice of instructional strategies?
- What do the students think is quality work?
- You have observed that this student’s work lacks focus – what makes you say that?

Protocols are most powerful and effective when used within an ongoing professional learning community such as a Critical Friends Group® and facilitated by a skilled coach. To learn more about professional learning communities and seminars for new or experienced coaches, please visit the National School Reform Faculty website at www.nsrfharmony.org.
• What would the students involved say about this issue?
• How have your perspectives on current events influenced how you have structured this activity?
• Why aren’t the science teachers involved in planning this unit?
• Why do you think the team hasn’t moved to interdisciplinary curriculum planning?
• What would be understanding of this mathematical concept look like? How would you know students have “gotten it”?
• Why did allowing students to create their own study questions cause a problem for you?
• Why do you think the expected outcomes of this unit weren’t communicated to parents?
• What was your intention when you assigned students to oversee the group activity in this assignment?
• What evidence do you have from this student’s work that her ability to reach substantiated conclusions has improved?
• How might your assumptions about the reasons why parents aren’t involved have influenced what you have tried so far?
• How do you think your expectations for students might have influenced their work on this project?
• What do you think would happen if you restated your professional goals as questions?
• What other approaches have you considered for communicating with parents about their children’s progress?
### Developmental Supervision: Approaches

<table>
<thead>
<tr>
<th>Teacher Input in Outcome</th>
<th>Supervisor Approach</th>
<th>Directive Control</th>
<th>Directive Informational</th>
<th>Collaborative</th>
<th>Nondirective</th>
</tr>
</thead>
<tbody>
<tr>
<td>When to Use (expertise, responsibility, care)</td>
<td>Minimum teacher choice</td>
<td>Selected teacher choice</td>
<td>Mutual choice</td>
<td>Maximum teacher choice</td>
<td></td>
</tr>
<tr>
<td>Teacher functions at very low developmental levels</td>
<td>Teacher functions at fairly low developmental levels</td>
<td>Teacher functions at moderate/mixed developmental levels</td>
<td>Teacher functions at high developmental levels</td>
<td>Teacher or group possesses knowledge and expertise</td>
<td></td>
</tr>
<tr>
<td>Teacher does not have awareness or knowledge</td>
<td>Supervisor possesses knowledge</td>
<td>Teacher and supervisor have same degree of expertise</td>
<td>Teacher or group has full responsibility for carrying out decision</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Supervisor totally accountable</td>
<td>Teacher confused/inexperienced</td>
<td>Both accountable and involved in carrying out decision</td>
<td>Teacher or group cares about the problem</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Supervisor committed to resolving the issue; teacher is not</td>
<td>Supervisor willing to take responsibility</td>
<td>Both committed to solving the problem</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>In an emergency</td>
<td>Teacher believes supervisor is credible</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### Examples of Supervisory Behavior

- **Direction of Teacher Development**
  - **Low**
  - **High**

#### Supervisory Behavior Continuum

- Listening
- Clarifying
- Encouraging
- Reflecting
- Presenting
- Problem Solving
- Negotiating
- Directing
- Standardizing
- Reinforcing

Excerpt from Forward by Arthur L. Costa

As a child, I can remember my mother telling me, “If you can’t say something nice, don’t say anything at all.” When my older brother teased me, when other kids crowded in line in front of me, when my teachers disciplined me without warrant, I learned to “turn the other cheek,” to hold my feelings inside, and to say nothing. I felt I was gaining an “inner strength” to withhold my thoughts, my emotions, and my words.

In my early professional life, these behavioral patterns sustained me. Educators are noted for avoiding conflict, for not rocking the boat or ruffling feathers, for holding personal relationships above organizational goals, and for striving to be perceived as “nice guys.” While I might be bristling inside because of detecting an inequity in practice or finding an inconsistency between what was being advocated and my personal value system, if I couldn’t say something nice, I’d say nothing at all.
The Importance of Confronting Behavior That Violates Purpose and Priorities

“Day after day in schools across America, change initiatives, instructional improvement, and better results for children are blocked, sabotaged, or killed through silence and inaction . . . this lack of follow-through results from the avoidance or inability to face conflict openly.” (Saphier, 2005, p. 37)

The need to confront resistance is "one of the toughest truths of change in school" (Evans, 2001, p. 276) because "confrontation forms a matching bookend with clarity and focus" (p. 288).

"Leaders must care enough to confront. Many people avoid confrontation. Others are afraid confrontation will make things worse by creating anger and resentment in the person they confront. When a person's behavior is inappropriate, avoiding confrontation always worsens the situation." (Maxwell, 1995, p. 125)

"Leaders who don't have the courage to force team members to step up to the requirements of teamwork . . . would be wiser to avoid the concept altogether." (Lencioni, 2003, p. 39)

"If you are unwilling to go to the mat when people violate a core value (such as giving their best effort), that value loses its moral force in the organization. On the other hand, you send a powerful message about your values when you hold people accountable." (Patterson, Grenny, Maxfield, McMillan, & Switzler, 2008, p. 216)

"A principal's seeming willingness to tolerate both incompetence and a lack of commitment within the faculty undermines his relational trust with parents, community leaders, and his own teachers." (Bryk & Schneider, 2002, p. 53)

"At the beginning of the change process, adults at featured schools were obliged to cooperate for change even before they believed that success would be the result. Nonetheless, core groups of leaders used their positional authority and their personal influence to engage colleagues in professional learning activities. There were confrontations, and there were courageous conversations. The biggest breakthroughs in reduced resistance came with 'seeing-is-believing' experiences . . . Ultimately, trust and success win hearts and minds. But it is the sense of responsibility among a small group of leaders, their hard work and their determination not to be put off by resistance that builds the trust and wins the initial successes." (Ferguson, Hackman, Hanna, & Ballantine, 2009, p. 32)

"Persuasion, consensus building, and all the other parts of influence don't always do the job. Sometimes it simply comes down to using the power or one's position to get people to act. A common failing of leaders from supervisors to top executives is the failure to be emphatically assertive when necessary." (Goleman, 1998, p. 190)
**Processing Guide to Prepare for Post-Observation Conversation**

1. Place existing evidence for the two components of most concern in the appropriate rating box.
2. Place desired behaviors/expectations for improvement in the box to the right.
3. Is there any evidence missing that would provide additional information to the evaluator?

<table>
<thead>
<tr>
<th>Component</th>
<th>Level 1 (Unsatisfactory)</th>
<th>Level 2 (Needs Improvement)</th>
<th>Level 3 (Proficient)</th>
<th>Level 4 (Excellent)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

4. After reflection on the information above, what areas of growth does that staff member believe are needed to improve his/her teacher practice?  
(Can the staff member identify his/her own behaviors/expectations for improvement?)

*Modified from* Having Hard Conversations* by Jennifer Abrams  
*Corwin Press, 2009  
Page 1 of 2*
5. What does the staff member need to know and be able to do to implement the desired behaviors/expectations?

(In order to implement these behaviors/expectations, what knowledge, skills, or awareness would this person need?)

6. What are some strategies you could use to help build up his or her resources and implement the desired behaviors/expectations?

(What are some of the specific things you could do to address the needs? Given what you know about the person, what language or actions might help him or her with the desired behaviors/expectations?)

7. What are some of the resources you need in order to execute the strategies above? Do you need to learn or relearn in order to support this person?

(What type of personal support do you need? What is your hunch about what emotion or value into which you need to tap to be most effective in your support?)

Timeline for implementation and follow up

<table>
<thead>
<tr>
<th>Date</th>
<th>Action</th>
<th>Who is responsible?</th>
<th>Outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Modified from Having Hard Conversations by Jennifer Abrams
Corwin Press, 2009
Page 2 of 2