

Rodgers, E. & Hung, C. (In press). Systematic Observation of Teaching. In A. Rodgers & E. Rodgers (Eds). *Coaching Conversations*. Teachers College Press.

Chapter 4

Systematic Observation of Teaching

with Carrie Hung

Even though we know that expertise in teaching is clearly related to student progress in school (Block, Oakar, & Hurt, 2002; Darling-Hammond, 1996), there is little research about what it is that effective literacy coaches do to develop that expertise (Dole, 2004). In fact, even though coaching has become a key component in the professional development repertoire of many school districts across the United States, we know little about how coaches coach. Joyce and Showers' Student Achievement Through Staff Development (1995) is often referenced in this regard, but their work is mostly at the system level; they consider how a district support system for teacher professional development might be designed and make recommendations about the content of those programs. The

best evidenced-based descriptions of how coaches coach comes from Lyons and Pinnell (2001) and Lyons (2002). Our goal in this chapter is to add to their work by describing our findings about how coaches observe and analyze teaching.

The Value of Systematic, Direct Observation of Teaching

We defined learning in Chapter 2 as changing participation in an activity. It makes sense, then, that to look for evidence of learning we should look at the nature of teachers' participation in instructional activities. Participation refers to how the learning activity is structured, what the teacher says and does during the activity, and how all of these things change over time. For example, coaches can look for changes in the way a guided reading lesson is conducted or for shifts in how content area reading is integrated throughout the day. The most effective way to determine changes in instruction is through direct, systematic observation of the teaching in action.

Why not rely on teacher self-reports of changes in teaching to inform the coaching process? After all, it is time consuming to observe teaching. In addition to the travel time involved, the coach has to sit through a lesson

and observe. Wouldn't it be easier and simpler to ask teachers to report what they have been doing differently?

Self-reports are not very reliable. According to Hops, Davis, and Longoria (1995), researchers in the field of clinical psychology began moving away from self-reports in the 1960s in favor of direct observation because it was thought that subjects' reports about their own behavior were likely to be biased and, therefore, unreliable. In some cases, for example, individuals who were asked to provide reports of their behavior focused on negative experiences and ignored positive ones.

Direct observation is also a commonly used research approach in education. It has been used to study teaching since the 1930s when researchers began to explore teacher-student interactions and other classroom behaviors (Evertson & Green, 1986). Like researchers in clinical psychology, education researchers realized that there can be little doubt as to the instructional activities in the classroom when they are directly observed. A teacher's self-report might be influenced by personal bias, perception, and an under- or overestimation of how new instructional practices are taken on (although in our experience we have found that teachers consistently underestimate how well they are doing). In any case, direct

observation by a coach likely yields more reliable information about the instructional practices than a teacher's self-reports about how things are going.

The value placed on direct observation in education is evident in the way teachers are assessed by the National Board for Professional Teaching Standards for national board certification. In addition to submitting a portfolio of artifacts that represents their teaching and their reflections on their teaching, applicants also submit a collection of teaching videotapes. While portfolios are highly regarded as an effective assessment tool, they still require an evaluator to make inferences from the materials about the teaching. Videotapes, on the other hand, provide evaluators with a way to directly observe teaching. As Hops and colleagues (1995) noted, "Unlike traditional assessment in which a person's score on an instrument represents some level of an underlying, covert trait, little inference from direct observation is required because data are collected on the specific behaviors of interest" (p. 194).

The challenge with direct observation, of course, lies in learning how to observe practice in productive ways that will enrich the feedback and joint inquiry that follows the observation. In this chapter we plan to share findings from the small body of research that presently exists about how

effective coaches observe teacher practice and make decisions about coaching. We begin with a description of the coaching context for this study and then discuss what we learned about how effective coaches coach.

Context for the Study: Literacy Collaborative

Lyons and Pinnell (2001) studied coaching practices within the context of Literacy Collaborative, a research-based comprehensive literacy framework to guide reading, writing, phonics, and word study instruction in K-6 classrooms. In addition to its instructional framework, a central feature of Literacy Collaborative is the ongoing work of a school-based literacy coach, called a literacy coordinator, who works closely with individual teachers to help them create positive learning environments, organize powerful teaching opportunities, and make instructional decisions based upon the needs of individual students.

This model of professional development and literacy instruction assumes that teachers grow professionally and learn by doing with the continuing support of these school-based coaches who also teach courses, model lessons, observe classroom teaching, and provide coaching aimed at shifting teaching to a higher level of precision based on students' needs.

Lyons and Pinnell (2001) described literacy coordinators' decision making as analytical and recursive in nature, turning on the coach's systematic observation and analysis of teaching. The coach works from these observations to select powerful coaching points that lead the teacher to deeper conceptual understandings and new learning (Lyons, 2002; Lyons & Pinnell, 2001).

Emily and Carrie, a doctoral student at Ohio State, used this coaching setting to learn more about how effective coaches coach. They used the following two questions to guide their inquiry:

1) How do coaches analyze teaching?

2) What are their understandings of quality teaching?

Exploring these questions will bring us closer to the heart of coaching: the analysis that forms the basis for the coaches' interactions with teachers.

Conducting the Study

Literacy Collaborative coaches were invited to participate in a pilot project to develop observational rubrics. Rubrics are written descriptions of levels of enactment, usually from the least idealized form to the most idealized form. Rubrics for essay writing, for example, might include four descriptions: one of a poorly constructed essay, one of a below-average essay, one of an

average essay, and the fourth of an ideal essay. A teacher evaluates the student's essay writing sample by comparing it against the rubric to see which features are present.

Rubrics are useful not only for evaluation purposes, but also for teaching. They communicate the range of kinds of performance to teachers and students, allowing participants to set concrete goals for themselves for improvement. Teachers often find writing rubrics a valuable activity because it requires them to think about what an ideal example would look like; it is an even more valuable activity when developed with colleagues.

The observational rubrics for this study were initially developed by a group of faculty involved with coaching coaches for the Literacy Collaborative project. We wrote approximately 60 rubrics to describe general instructional practices and specific teaching practices related to each lesson component. Here is an example of one of three rubrics that we wrote for the lesson component called Interactive Read-Aloud. Notice how each description moves closer to an idealized form:

- Teacher begins read-aloud with limited or no interaction about the book or topic to prepare students as listeners.

- Teacher provides some opportunity for students to think about the book to be read which may be unfocused or only marginally related to the book topic.
- Teacher engages in some preliminary interaction about the book which at least partially engages students' attention and prepares them to listen.
- Teacher engages attention of the students prior to reading with brief comments or questions, preparing students for active listening and response.

The long-term goal of our study was to establish the reliability and validity of the observational rubrics. Once developed, these rubrics would eventually be used by the literacy coordinators as a tool to observe teaching and support coaching decisions. We do not share all of the rubrics in this volume because they are specific to the Literacy Collaborative instructional framework; our focus here is to describe what we learned from piloting these observational rubrics about how effective coaches observe and analyze teaching.

The coaches involved with piloting these rubrics received 2 days' training on how to use them. They viewed multiple videotapes of instructional practices from the Literacy Collaborative framework and used the rubrics to observe, analyze, and rate the instruction. (We should

mention here that typically we would not advise coaches to rate teaching. Ratings were necessary in this case in order to establish the reliability and validity of the observational rubrics.) Following their observations of each videotape, the coaches compared their ratings with one another and discussed rationales for why they rated the instruction as they had. Although we continued to collect data for the rest of the year on how the coaches used the observational rubrics, the findings reported in this chapter come from the data that Emily and Carrie collected and analyzed from the initial 2-day training session.

To address the two questions posed earlier about how coaches analyze teaching and what they understand about quality teaching, we transcribed and analyzed their small group discussions about their ratings of each instructional element. We also collected the rubrics and analyzed the coaches' anecdotal comments written directly onto the rubrics as they observed the videotapes.

Analyzing the Data

We used a constant comparative method of data analysis, as informed by Strauss and Corbin (1998), to discern common themes across the coaches' handwritten comments on the rubrics and within the transcriptions of their discussions about their rubric ratings. Each

researcher analyzed the data set independently and then met to discuss and reach agreement on emerging categories and themes. Patton (1990) refers to this kind of data analysis—when two people look at the same data set—as a form of analytical triangulation of data analysis. This independent and shared analysis with multiple passes at the data contributes to the trustworthiness of our analysis.

In our first phase of data analysis we compiled all of the handwritten notes that the coaches made on the rubrics while they observed the videotapes. We regarded these handwritten comments as “notes to self” because the coaches were not asked to write anything down while they viewed the videotapes. We assume that the coaches made notes about the kinds of things that they wanted to recall later.

We organized our observations into three categories of comments. We noted that many of the written notes were descriptive in nature, with the coach actually scripting the teachers’ talk and the students’ responses. We coded these kinds of descriptive comments as What I Observed. We searched through the remaining notes to see if the coaches evaluated what they observed and found that although they did, their notes were only about instructional practices that they thought did not go well. None of the comments described what they thought went well. We called this

second category of notes What Went Wrong. The third category of comment to emerge was Rubric Reflections, in which the coaches noted questions or confusions that they had about the content of the rubrics. This last category did not occur as frequently as the first two categories.

With these categories in mind, we began our second phase of analysis. We transcribed the small-group discussions of rubric ratings and then read through the transcripts for themes related to our research questions. We found that the coaches' talk as they compared and discussed their ratings could be grouped around two related themes: teacher decision making and student response.

We then coded the coaches' talk in terms of what they were saying about teacher decision making and student responses. We found that the coaches' analyses of teacher decision making could be categorized in terms of the responsiveness of the teacher's actions and the relevance of the actions. In other words, it seemed that the coaches' analyses revolved around two questions:

- 1) Do the students need to know this; is it relevant?
- 2) Is the teacher being responsive to the students?

This data analysis process led us to the patterns that we describe in the next section about how coaches observe

and analyze teaching as well as what they consider to be quality teaching.

Observation and Analysis of Teaching

In response to our first research question regarding how coaches analyzed teaching, we found that coaches' observations were almost clinical in nature, often scripting exactly what they were seeing and hearing. Our second research question addressed how coaches analyzed quality teaching. Our analysis of the coaches' discussions about their ratings on the rubrics led us to identify three lenses that influenced their observations: the pedagogy lens, the responsiveness lens, and the relevancy lens. We discuss both of these findings in depth in the sections which follow.

Systematic observation. As they observed the same videotapes of teaching, coaches made handwritten notes on their rubrics to script exactly what the teacher said and how the students responded. Their notes were so systematic that they might even be regarded as transcriptions of the interactions which they observed. It was as though they wanted a written record of everything that was said, but only at specific times in the lesson they observed.

In fact, when we compared their scripted notes afterwards, we were surprised to find that the coaches tended to select the very same teaching interactions to script. Most coaches, for example, transcribed the teacher's introduction to a new book, writing down the vocabulary that the teacher discussed before reading the story. Most coaches also wrote down exactly what the teacher said the story was about. This similarity was interesting to us because even though the coaches observed the lesson at the same time, they recorded their observations independently. There was no way that any of them could know what other coaches were writing.

As we further analyzed the content of the coaches' transcriptions, we noted that they tended to script specific teaching points that they did not agree with; in other words, they scripted interactions that they thought did not go well. We inferred that had these been actual coaching sessions, the coaches would have wanted access to a written record of those particular interactions that they perhaps would return to later.

How do we account for the similarities across the coaches' observations? It would not make sense to write down every word that was uttered, yet when they wrote they tended to record the same things. It was as though the

coaches were making sense of their observations in much the same way, using a similar filter or lens to sift and sort their observations.

The pedagogy lens. Adrian conceptualized a mechanism called a pedagogical lens to help understand how the same teaching might be viewed and interpreted differently by different people (A. Rodgers, 2002). A literacy lesson that included explicit instruction in phonics, for example, might be considered timely and instructive by one person but lacking in quality by someone else who thinks that phonics instruction should also be systematic, preplanned, and by necessity decontextualized.

It may be that because most people have had experience with formal schooling, we all—educators, coaches, parents, and policymakers alike—have our own pedagogical lens through which we understand and evaluate teaching. What is interesting in this present study is not the variability of the coaches' observations and analyses, but its similarity. The coaches seemed to use a very similar lens with which to view instruction, judging by the commonalities across their scripted observations about how the lesson was going. It's as though they had a shared pedagogical lens which

suggested something like, "This is a guided reading lesson, therefore I expect to see..."

No doubt the coaches shared a similar pedagogy lens because they had all taken extensive professional development training to be Literacy Collaborative coaches. We suspect that no matter the training or preparation a coach has received to prepare for the role, there will always be this pedagogy lens that will influence their observations and analysis of teaching. In the example which follows, the pedagogy lens is clearly at work in the coach's observations. In other words, the coach clearly is observing and analyzing the teaching in terms of how the teacher should teach word solving.

I thought she gave a lot of clues. She should have had have them attempt the word. She kept pointing to the picture and one boy looked at her and she said, "Look at the word." She never really did get them to look at any of the sounds; there wasn't any word solving at all. (Sheila, Literacy Coordinator, December 2004)

This coach's observation suggests that she thought there was undue attention given to using the pictures as clues to identify unfamiliar words in the story; there should have been more instruction in word solving, specifically learning more about letter-sound relationships. It seems

that the coach's observation of how the teaching was going did not fit with her pedagogical lens or what she expected to see.

The transcripts which follow provide two more examples of coaches analyzing teaching in terms of how they expect the teaching to proceed. Both coaches are reflecting on the same videotape. The similarity across their observation is evident: both think the teacher should have involved the students more in thinking about the meaning of the story along with the attention to decoding words. One coach said,

I felt that the students needed to be engaged more in the conversation because a lot of it was more like, "What will you see in cry?" or "What does went begin like?" [She should have worked] more with the story rather than just parts of what you hear. (Keisha, Literacy Coordinator, December 2004)

The coach in this next example, commenting on the same excerpt, also addressed the how of teaching in her observation and comes to much the same conclusion:

In that particular book maybe, if she maybe alluded to why the boy become the lamb's mother and not other animals'. It's just like going beyond the literal a little bit—maybe just not really having a full-blown discussion, but still alluding to the fact that there

is a turn of events here, what would be the rationale for it, let's find out about it. She got them engaged in the text, but not to anticipating the deeper thinking. (Jackie, Literacy Coordinator, December 2004)

In our discussion thus far of how literacy coaches in this study observed teaching, we have described what we call the pedagogy lens; that is, the coaches seemed to take into account the how of teaching. In a way, this measure is similar to a test of face validity. It's as though the coaches asked themselves, "Does this look like what I'd expect this particular component (teaching for fluency, word solving, comprehension, etc.) to look like? Another way of summarizing this kind of analysis would be to say, "Yes, that looks like a guided reading lesson" or "That's not the way a book introduction should go."

The responsiveness lens. When coaches discussed rationales for their ratings, they often referred to what we call the responsiveness of teaching as a factor in their analysis. For example, coaches cited a lack of teacher responsiveness to how students were learning as a rationale for giving a low rating to a video segment of instruction. By lack of responsiveness the coaches meant that the

teacher seemed to be following a set agenda without noticing whether the teaching was too easy or too hard for the students. One coach summed up the responsiveness lens nicely when she said, "What you're really looking for is, is she meeting the needs of those kids?"

Another coach had this to say about a lesson, after giving the book introduction a low rating on the observational rubric. Notice her comment that the teacher was not being responsive to the students, but instead was teaching in a routine way:

I was thinking that her introduction was a bit lengthy; she gave the whole story away. I don't think she did it because they needed it. Maybe it was just something she usually did and she just got on a roll.

(Kiearra, Literacy Coordinator, December 2004)

Another coach in the same small discussion group agreed with that observation, saying

I kind of agree. I thought there must be some students there who were learning to speak English, and that's why the teacher was being so supportive with labels, concepts, and vocabulary in the story she was introducing. So I thought maybe it's an appropriate level after all. (Gail, Literacy Coordinator, December 2004)

In fact, the responsiveness lens played such a large role in the coaches' observations and analysis of teaching that they often formed hypotheses like the one in the previous example to explain the mismatch between the teaching they were observing and what they thought the students could do. Another coach, commenting on the same videotape of teaching, said

I felt that her book introduction was too heavy; she gave so, so much support in order for the kids to take on the reading, I thought maybe it could be the first time they read that level so you always would want to give a higher amount of support at the beginning...I just felt like the book introduction was very lengthy so I was concerned about the text selection. It was too hard for them. (Gabriella, Literacy Coordinator, December 2004)

Clearly coaches were taking into account the degree of responsiveness present in the teaching. They expected teaching to be matched to how students were responding and what they were able to do. The coaches noted whether they thought a teacher was doing something that the student already knew how to do, or couldn't do even with help, and this factor played into their analysis of the teaching.

The relevancy lens. A third dimension of coaches' observations had to do with the relevancy of teaching—what we refer to as the relevancy lens. Their reflections on their ratings also dealt with whether or not they thought students could profit from the teaching; this could only happen if what the teacher was teaching matched what students could use to learn. One coach, for example, made the following comment when sharing her rationale for her rubric rating:

The teacher said, "You did what good readers do, you went back and fixed it," but she really didn't go back into the text and say, "This is what you did." The student went all the way back here and reread. I'm not sure he made the connection about what the teacher meant: to go back and reread. (Sonya Gail, Literacy Coordinator, December 2004)

According to the coach's analysis, the teacher's comment was likely not relevant to the student because the student did not know what the teacher was referring to in her comment. The teacher wasn't really teaching exactly what it was the student needed to know. In other words, the coach analyzed a gap between what the teacher was doing and what the student could use. Specifically, the coach thought that the teacher's comment, "You went back and fixed it" was

probably too vague to be useful and relevant for the student at this time. More specific teaching was needed. The teacher might have said something like, "Good job checking on yourself when you fixed up that word [perhaps pointing to the wrongly read word]. Were you right?" If the coach was correct that the student needed more specific teaching at this point in order to know what "went back and fixed it" meant, then this alternative teaching might be more relevant and useful to the student.

Another coach, reflecting on a different videotape, gave the following rationale for her rubric rating. Note how the relevancy lens comes into play as the coach describes how she analyzed the teaching:

She just let things happen which I know that's good to do, but she didn't have a focus. The word random came to mind. By bringing the student up to the chart paper, she needed to have a teaching point. But instead she just got into talking about how the student didn't have any front teeth. So there was no teaching, even though she could have said something that she [the student] was doing well and make some teaching from that. (Marjorie, Literacy Coordinator, December 2004)

In this example, we can see how relevancy plays a role in the coach's observation of the teaching interaction. The coach expects the teacher to really teach, questions teaching moves that are not meaningful or useful for the students, and notes missed opportunities for learning.

Discussion

We began this chapter by pointing out that there is very little research on what it is that effective literacy coaches do (Dole, 2004). We had a unique opportunity in this study to gain insight into this very question by having coaches view and rate teaching videotapes, and then share and discuss their ratings in small groups. Again, we point out that while we believe that coaching teachers is not about rating their teaching performance, coaches rated teaching in this study to help us understand how they interpreted teaching. The coaches' rating of teachers that we report here is not what we recommend coaches should do; rather the rating represents a research tool for us to better understand what coaches do when they observe teaching.

The coaches in this study were not following a manual or checklist to tell them what teachers ought to be saying or doing at particular times throughout a lesson. Instead, the coaches observed teaching in a systematic way, even

going so far as to script specific teaching interactions. We were interested in how they decided what to script and how we could account for the similarities in their teacher ratings.

Our analysis of the coaches' discussion following their analysis led us to conceptualize three lenses through which they viewed the teaching. The pedagogy lens filtered their observations of how the teaching was going. The responsiveness and relevancy lenses led them to judge whether the teaching was relevant to the children's needs and whether it was pitched at an appropriate level.

It was interesting to note that even though the coaches were observing teaching, their focus was not limited to the teacher and what the teacher was saying and doing. Instead, their observations included the students in terms of how they were responding and participating in the lesson. We might say that the coaches' way in to analyzing the teaching was through the students. The coaches noted whether the students were about to profit from the instruction, whether the teacher was keeping it easy enough for the students to learn, and whether the focus of the teaching was what students needed to learn at that particular time. As Lyons (2002) suggests, "coaching should never focus on what the teacher is doing right or wrong,

but how the students are learning as a result of teaching” (p. 96).

The coach’s lenses to observe and analyze teaching remind us of Wood’s (1996) description of the nature of scaffolding. He said that teaching decisions take into account when to teach, what to teach, and how to teach; and that these decisions are often made on a moment-by-moment basis as the teacher decides what to say or what to do next. The what and how to teach are informed by knowledge of the subject matter and knowledge about how to teach. The when decision, according to Wood, is probably the most difficult to make because it requires the teacher to match the content to the student. We are not referring here to the larger curriculum-related decisions about what material to cover when during the year, but the moment-by-moment decisions teachers make about when students are ready for a new concept or a new procedure. This decision is informed by the student’s participation and can be summed up by the question, “Have I prepared my students to learn like this?” and “Will I introduce this concept now?”

Similarly the coach’s lenses filter their observations to consider the how (pedagogy), the what (relevance), and the when (responsiveness) of teaching. It is as though the dimension of scaffolding became the yardstick that guides

coaches' observations and led them to distinguish quality teaching from less-than-quality teaching. This is an extremely significant finding. Districts should avoid hiring coaches based only on years of service or previous administrative backgrounds; instead they must consider many other factors in the hiring of coaches. Our finding suggests that those school districts which insist on hiring coaches who are not highly trained in both content and how to teach that content will only get simplistic approaches to coaching—approaches that may not be effective in changing teaching and affecting student learning.

One thing seems certain. Coaches need to be well prepared in the subject matter so that they can observe teaching in terms of the what, how, and when kinds of discussions that teachers need to make. No doubt a coach with minimal training can observe teaching in terms of what is being taught. This is particularly true, and no doubt particularly easy, if the teachers are following a set curriculum that mandates what should be taught and when it should be taught. In that case, coaching can be reduced to observing with a checklist to see if particular components are being taught as specified by the core curriculum. Does the teacher include a lesson on word solving? Does the teacher provide the specified amount of time for familiar

rereading? We wouldn't call this teaching, though, nor would we call it coaching.

Teachers need a theoretical understanding of what and how they're teaching in order to be able to make teaching decisions in the moment. As Lyons et al. (1993) noted, without theoretical understandings "lists of good teaching behaviors performed in mechanical ways in response to supervisor's observations, are useless" (p. 43). Recently we were reading an account of coaching, and it struck us as hopeful but also naive:

One teacher wrote in a reflection "I was really nervous about demonstrating a guided reading lesson for a colleague and then coaching her through similar lessons. We co-planned the lessons and discussed our experiences after each lesson. We both felt that this process was very helpful. I am convinced that this approach will make a positive impact in our school."

(Shaw, Smith, Chesler, & Romeo, 2005, p. 10)

We admire the positive attitude and the dedication implicit in this observation, but all of our experience tells us that coaching is both delicate and sophisticated work. We need more than good intentions, more than a readiness to work hard, and more than a willingness to take risks. In place of this naive assertion, we conclude that coaches not

only need knowledge of what should be taught, but a theory of literacy and learning and an understanding of effective teaching. We concur with the International Reading Association's (2004) stance that if reading coaches are not already reading specialists when they are hired, they should work toward that goal within 3 years.

Conclusion

We know in general that the work of coaching spans a broad range of activity types, from demonstrating model lessons for teachers to suggesting material for teachers to read. They might work one-to-one with teachers or in groups such as grade-level meetings (Bean, 2004a). In short, coaches are multipurposed, school-based staff developers who tailor their activities around the needs of particular teachers and groups of teachers. Their work is contextualized around the state of practice in a school and around the level of expertise of teachers.

We expect to develop a more detailed evidence base about the actual work of coaches in terms of what they focus on when they observe teaching and how they make decisions about quality teaching. Advancing conceptualizations about this work is important in its own right because insights that grow out of this research will

make more detailed specifications of what works about coaching.