

Andrew D. Gitlin
Robert V. Bullough Jr.*
University of Utah

Improving Teaching Through Peer Evaluation

Abstract

Horizontal evaluation, a model for teacher evaluation based upon the structuring of dialogue in order to help teachers better understand the nature and influence of their decisions, is described and the results of a pilot study using the model presented. Through horizontal evaluation pairs of teachers analyzed the assumptions they held about teaching and learning and began to understand, critically, taken-for-granted aspects of schooling including problems of inequality, contextual constraints on teachers' work, and legitimacy of program goals.

Introduction

Teacher evaluation is widely understood as a key element in the improvement of teaching. The dominant model of evaluation arises from process-product research where the concern is to establish statistically significant relationships between teacher behaviour and tested student achievement (Brophy and Evertson, 1984). The effective teacher is the one who demonstrates those behaviours that have been shown mathematically to relate to achievement. The Florida Performance Management System is typical of this approach. A trained evaluator observes a teacher at work and notes the presence or absence of a list of 121 competencies (Florida, 1983). The Kalamazoo evaluation program is yet another example, but with a twist: teachers evaluate other teachers using an administrator developed rating scale of desirable teaching behaviours (Bartz, 1978).

*The authors contributed jointly and equally to this article.

Both approaches are seriously limited as means for improving teaching: The narrow emphasis on behavioural change – behaviours presumably connected with increasing standardized test scores – reduces teaching to a technical rather than moral or ethical activity. The emphasis of process-product research on identifying general laws governing behaviour means that context variables which have a profound impact not only on what teachers do but on what students learn are ignored. And, questions relating to how teachers understand and think about their work, especially what they see as problematic, are relegated to insignificance. By ignoring how teachers think about and understand their work, these approaches to evaluation have the added consequence of curtailing teacher professionalism while encouraging the better and brighter teachers to leave the field in search of occupations that will allow more opportunities to express their initiative, creativity, and intelligence.

Fortunately, not all evaluation schemes are cut out of the same frayed cloth. Clinical supervision, at least in theory, takes a different slant, aiming to "enhance the learner's self-sufficiency and freedom to act" (Goldhammer, 1969, p. 55). To achieve this aim, clinical supervision relies upon a five-stage interaction of teacher and peer observer: Interaction begins with a preobservation conference during which time the observed teacher is expected to share aims. The classroom is observed and then the evaluator analyzes what took place and decides on a strategy for presenting the data that is most likely to affect desired changes. The observer and teacher then meet together in a conference during which perceptions are shared and plans made. The final stage is a post-conference analysis where the two teachers assess the quality of their interaction.

Unfortunately, as practiced within the public schools, clinical supervision frequently fails to achieve its aims. McFaul and Cooper (1983), for example, report that when peer teachers used the clinical supervision model, discussion centered on the tricks of trade while more substantial questions about the aims of their practice or about how they understood their work were ignored. They conclude that this failure was the result of the overwhelming negative influence of school structure. But, this may not be the only cause. The model itself might be partially to blame because it does not contain either a sufficient number or the right kind of safeguards to discourage it from being collapsed into a discussion only of teaching techniques. Specifically, the emphasis in clinical supervision on objective observation, along with the requirement that supervisors leave the teacher with concrete suggestions for improvement, and the failure of the model to provide an explicit means for analyzing the relationship between what a teacher intends to do and what actually transpires, likely elevates the importance of technical questions and issues.

If evaluation is to have a hand in developing teachers who know how and when to apply or change a recommended technique in order to achieve a desired aim; who are enthusiastic about their work, independent but willing, even eager, to talk with and learn from others about it; who analyze school aims and means in relationship to ethical social standards – who are, in short, professional educators – then we need to reconsider other models. Within this paper we describe a model of evaluation, horizontal evaluation¹, that represents a severe departure from the dominant model while building on the strengths of clinical supervision and correcting its weaknesses. This model links the concern for behaviour to what teachers think and feel and to the institutional constraints under which they work. To make our case, we present the results of a recent pilot study conducted using the model.

The Horizontal Evaluation Model

The aim of horizontal evaluation is to help teachers become students of their own practice, individuals capable of recognizing areas in need of improvement and of making appropriate changes. To accomplish this aim requires, first, that teachers together, both evaluated and evaluator, gain understanding of their practice. In order to make ethically responsible and reasonable decisions about practice, it is necessary to throw into relief the values and beliefs and the contextual constraints that influence teacher decision-making. The aim is not to gain compliance with a pre-established list of desirable behaviours, but rather to help teachers see practice as reflecting choices about educational values over which they have a measure of control. In a school this suggests that teachers should consider with others who share their interest in educating the young why they make particular decisions and the desirability of these decisions in relation to an articulated, although inevitably fluid, normative framework that requires constant justification.

Intention and practice

To accomplish this aim, it is necessary to structure teacher dialogue. Much like clinical supervision, horizontal evaluation begins with the teacher sharing intentions. Intentions are shared as a first step toward helping both participants better understand and critically assess the values underlying statements of purpose. Once intentions are shared and clarified, they are analyzed in relation to practice. However, teacher intentions should not be viewed as reified concepts, fixed forever more. Rather, emphasis should be placed on understanding aims as evolving in relationship to practice; the relationship between aims and means in teaching is dynamic and interactive. Accordingly, the discovery of discrepancies between ends and means is not an occasion for apology, nor necessarily for an adjustment in practice, rather, it is an opportunity to reconsider both. Moreover, the

discovery of congruence is not necessarily a moment to be celebrated but rather an opportunity to ask a different set of questions. Congruence is only desirable if the assumptions underlying both practice and intention are exposed, considered, and found to be ethically and educationally justifiable.

To make the consideration of assumptions more likely, horizontal evaluation provides several tools that guide the discussion of the relation between ends and means. These include communication analysis, historical perspective, and alternative methods.

Communication analysis

Communication analysis refers specifically to a process that uncovers the prejudgments embedded in speech. One relevant model for guiding analysis of what teachers say is Habermas' (1976) "Universal Pragmatics." Habermas has identified four validity claims present in every speech act: comprehensibility, truth, sincerity, and appropriateness (p. 1-68). These four claims establish the range of questions that can be asked of any utterance.

The first claim raises the question of whether or not a teacher's statement is understandable. Considering the truth of a statement such as "The students behaved as they did because they are in the low group" is a more complex undertaking. The issue is whether, or in what ways, the sentence can be a true description of those involved and of what happened. To get at this issue, a teacher might raise a series of questions about the accuracy of describing the group as "low." Coming to agreement on the "facts" of a situation is an essential element in establishing understanding.

The sincerity claim leads to questions about whether or not what is stated reflects the teacher's actual perceptions and feelings. What I say sincerely, therefore, indicates both my understanding and that which I accept as natural, common sense. Categorizing groups of students as high or low, for example, may well occur because such labels are taken for granted as having obvious meaning. A way of getting at this is by examining the feelings we hold about the categorizations we use to organize the social world which are revealed in our sincere statements. Potentially, our talk, when scrutinized, may disclose tensions or discrepancies between feelings and understandings and how we communicate them. Through this type of discovery, teachers can lay hold of the systematic distortions built into their utterances.

The final validity claim, is a statement appropriate, helps in separating what *is* from what *ought to be* – Is the labeling justifiable,

right? Inherent in this question is the belief that teaching is more than a matter of applying techniques artfully; it is a moral activity about which moral judgments are made.

When probing an utterance to gain understanding, it is not always necessary to ask questions which reflect each of the four validity claims: comprehensiveness, truth, sincerity, and rightness. It is often sufficient to work with one claim while being aware of the others. In practice, the direction of the discussion will suggest which of the claims need to be raised.

Historical perspective

In addition to raising questions about the validity of a statement, horizontal evaluation encourages teachers to see current practice in relation to past events. For example, if a teacher is working in an open classroom, discussion could include comparison of the past and current forms of the open classroom as well as the evolution of the idea in relation to developments within society.

Since educational ideas are filtered through perceptions, historical perspective includes analysis of the histories of teachers and students. If a teacher, for example, decides to make the students stay in for recess because they haven't finished their work, this decision can be partially understood as a response that has historical roots. In this instance, the focus of dialogue is not to find a better solution to the problem *per se*, but to expose the general themes that help mold classroom life over time.

Alternative methods

The concern of many evaluation schemes is to help teachers identify alternative practices that "work." The suggestion of alternatives is also part of horizontal evaluation; however, the purpose is different – to illuminate the educational implications of taking a different approach. For example, if the teacher is conducting a large group discussion, the observer might suggest that the class be broken into small groups. One reason for making this suggestion would be because the observer believes it is better suited to the achievement of the teacher's aims. But, the more important reason is to help both participants consider the educational possibilities inherent in a given situation and to help them make decisions based upon an understanding of the values reflected in particular choices.

Challenge statements

The final horizontal evaluation technique is the issuance of challenge statements. The purpose of challenge statements is to push teachers to

consider issues related to Habermas' rightness or appropriateness claim. A challenge statement is a counter-claim or contradictory explanation of an event that throws into question the ethical or moral justification of the challenged action. For example, if a teacher wants students to obey her classroom rules in all circumstances, the observer could raise questions about the appropriateness of her demand in the light of student and teacher rights. By so doing, the observed teacher should gain a deeper understanding of the normative framework that underlies classroom decisions.

It is important to keep in mind that challenge statements, like all horizontal evaluation methods, are not only issued by the observer but also by the observed teacher. Both share the task of constructing a dialogue that will enhance their understanding of teaching and learning.

The Study

To begin exploring the model and its potential for helping teachers come to understand teaching differently, we conducted a pilot study involving twelve elementary school teachers in a gifted and talented program, *Horizons*. To be candid, it should be clearly stated that because of the size and type of the teaching population studied, our findings are suggestive but limited in their generalizability. Six dyads were formed of the teachers' own choosing; four completed the study. In October, 1984, the teachers were given four hours of training in the use of horizontal evaluation methods. After the training, the teachers participated in a trial observation and conference in which they used the methods. These conferences were audiotaped and discussed in order to clarify confusion and provide feedback on the model's use. Each team conducted four observations and four follow-up conferences throughout the remainder of the school year which were audiotaped for later transcription and analysis. And, following the conclusion of the study, each participant was interviewed to gain additional understanding of the effect of the model. Transcripts were analyzed independently by the research team and three graduate students to identify patterns and discrepancies.

Results and Discussion

Three of the four dyads that completed the pilot study engaged in critical dialogue about practice. They raised a wide range of issues beyond the technical – how to do something – and they increasingly questioned intentions and the assumptions underlying practice. Moreover, as their understanding of practice changed, they modified what they did in the classroom. One dyad, however, failed, choosing instead to focus almost entirely on technical concerns. We have included data from both the

successful and unsuccessful dyads in order to understand better why horizontal evaluation might be effective with some pairs and not with others.

The successful dyads

The horizontal evaluation method most often used by the teachers was analysis of the relation of intention to practice. This method helped the teachers to assess critically taken-for-granted aspects of teaching practice. In one conference, for example, a question was raised about why teachers so frequently emphasized tests, reports, and writing samples. The observer wondered if centering a unit on products was consistent with her colleague's long-term aim of developing students' learning processes: "This is just something that came to my mind. We keep [teaching] these units. I think, sometimes, we have a disparity between wanting students to get material and our intention to teach [cognitive] processes. I think it's certainly a dilemma . . ." The discussion that followed centred on the conflict that often arises between a process aim and an instruction emphasis on products. Both teachers had been unaware of this tension. Once the problem was identified and discussed, they began to consider other educational possibilities.

In another conference the observer commented that the observed teacher seemed a bit frustrated when "Nobody got the one about what problems might [exist] if you have smaller children in your house with dangers." The observer posed a challenge to the teacher by suggesting that the students didn't respond because the teacher was fishing for a right answer and they didn't want to risk being wrong. Realizing that she had undermined her own stated intentions, the observed teacher commented that the students "must pick up on the fact that even though I'm having [a] very open-ended [discussion], in all the open-mindedness there is a right answer."

Relating intentions to practice revealed other inconsistencies between what teachers say and do. One conference focused on the teacher's intention to have students "be able to find sources for learning." The observer focused the discussion on a troubling habit: "I don't know if you're like me, but a number of times you said, 'We need to go look that up. We need to look into that further.' I say that a lot and then I never go do it." By being openly self-critical, an event common to horizontal evaluation but rare in other schemes, the observer opened the way for her partner to reflect more fully on her stated intention. "We got into the big magnetic field thing because I asked the 5th graders if they knew what a magnetic field was and they didn't. I said, well then, maybe *we* [emphasis added] need to find out . . . but I also need to find out." Through their interaction, both teachers were able to see conflicts previously hidden from view.

Discussion of the relation of intention and practice also led to consideration of problems of inequality. In one conference, for example, the observer noted that, "In terms of your intents, the first one [to make students feel worthwhile], you do that really nicely with the verbal kids." In approaching the issue in this manner, she confirmed that her partner had been successful in accomplishing at least part of her aim. Taking the hint, the observed teacher asked, "But not with the non-verbal kids?" The discussion that followed revealed an additional problem: most of the non-verbal students were girls. Through analyzing the situation together, the observer was then able to see a previously hidden inequality and consider possibilities for improving the situation.

In some dialogues, discussion of the relation of intentions to practice pointed to the influence of contextual constraints. For instance, several discussions centred on the procedures for admission into the gifted and talented program. Specifically, students were selected into the program through classroom teacher recommendations and high achievement test scores. One teacher commented that this procedure made it virtually impossible to achieve her aim "to encourage the abilities of all children in a sort of even-handed way [because] I can't take as many girls for math because they don't seem to have the interest or the motivation in that area." Awareness of inequality of this kind is a necessary first step toward confronting it. Not being able to alter admission procedures, the teacher commented that "When I do have a girl that's good in math, I really try to encourage her. . ."

These dialogue segments suggest that by carefully considering the relation of intentions to practice, teachers not only are able to identify unintended outcomes, but also to see how their aims are shaped by school structure and by their own values and beliefs. Through analysis new educational possibilities open up, institutional constraints are identified, and taken-for-granted assumptions about teaching and learning are thrown into doubt.

Communication analysis and historical perspective were also present in the dialogue of the more successful dyads. These methods encouraged the teachers to clarify program goals, such as developing higher level thinking skills, and lead to consideration of the proper role of a teacher in a gifted and talented program.

Most of the teachers in the study held as the primary aim of instruction that of developing higher level thinking skills. This said, they showed remarkable disagreement about what they meant. In one conference the teachers used communication analysis to clarify this intention. "If we consider knowledge and comprehension lower-level thinking, then application, analysis, evaluation, and synthesis are going to be higher level

thinking. . . It seems so easy to define higher-level thinking by using Bloom's taxonomy. But, for me, it's just getting beyond the memorization and regurgitation of knowledge." The observer responded: "In that sense, just making choices, even though you don't go through a long involved process of using criteria, and why is it that you chose this, would be higher level." Although the meaning of "higher level thinking" was not explored in any depth, the teachers did establish a shared understanding that allowed discussion to continue with greater clarity.

Communication analysis was used to understand better the internalized role of teacher. One observer noted, for example, that the peer teacher referred to herself in the third person, "You refer to yourself as Mrs. T. rather than saying, 'I.' Were you aware of that?" The peer teacher said that she wasn't and wondered if that tendency came from being a mother. The observer agreed, but also noted that "it removes you one step from [the students], it's real subtle." Understanding teaching as mothering legitimates certain kinds of teacher actions while making unlikely other actions. By exploring the use of language in this way, the observed teacher was able to consider how she was communicating to students much more than she consciously intended. Clearly, this was a realization she would not have made without being observed by a trusted peer. "I can't believe that [I refer to myself in the third person]. This is why [the evaluation process is] so beautiful because I would have never picked up on doing that."

Historical perspective also was used to understand better teacher role. In one discussion the focus was on the importance of maintaining classroom control. The observed teacher commented: "I have to say the only thing I felt uncomfortable about is I came from an educational system, the Catholic school, where the teacher was in control. When I have independent study going on like that and I don't know what everybody is doing, I get nervous wondering if people are wasting time or what not." After considering how she had been socialized into a particular view of what was appropriate teacher control, this teacher noted, with some pleasure, that despite her fears, the students did not seem to be wasting their time: "When I left [the room], they were working very well independently." Although she did not change her basic view of the need for teachers always to be in control, her view softened: "Maybe I need to give them credit for [working independently]."

These dialogue segments, and many others, indicated that teachers can critically reflect on their work and that perceptions and practice do change as a result. Moreover, when given the opportunity and the tools necessary for reflection, they are very much interested in issues beyond the technical. The teachers did question their taken-for-granted assumptions about the need to maintain tight control over students and they became

increasingly aware of the importance of attending to the unintended implications of practice. And, they saw issues contextually – as resulting from their own and their students' histories and that of the schools within which they worked. In sum, they understood teaching more broadly than ever before. They saw problems where formerly they had seen none, and they saw possibilities for improvement never before imagined.

The failed dyad

One dyad stands in stark contrast to the openness, trust, and non-defensiveness characterizing the three successful dyads. Ignoring the stated intention of horizontal evaluation and the training during which time the limitations of rating scales were noted, one observer created a scale to determine how effectively her partner achieved her aims. Mechanically, the observer assessed each stated goal and its achievement:

In meeting each one of your goals, I just gave you a rating. . . on your first long-term objective which was to foster joy in learning. I gave you a 'superior.' On your second long-term objective, which was to expose students to ideas they might not meet elsewhere, I gave you 'excellent.' On mere passive acceptance of information, I gave you a 'superior.' Now, on today's objectives, to find a subject that each child will be interested in learning and writing about, I gave you an 'excellent.' To share appreciation for things that are good and beautiful, a 'superior.'

Her partner did a similar kind of thing. Objectives were listed, and quick assessments were made.

In effect, the two teachers transformed the model to make it consistent with process-product forms of evaluation where hierarchical relations are seen as necessary and desirable. The teachers took turns playing the expert which made it virtually impossible to use the methods of horizontal evaluation appropriately. One result was that, unlike the other dyads, these teachers' relationship actually deteriorated as they continued to meet. If they ever respected one another, by study's end all respect was lost: "Now, in my case and Penny's (fictitious name), we're two different kinds of teachers. I'm more of a practitioner and I think Penny's more of a theoretician. And, she's a very, very intelligent person. I found myself thinking Penny might be getting more out of this experience if she were working with someone who was her intellectual equal." Neither teacher was able to break away from the belief that evaluation is fault-finding and expert-driven.

Penny did not view Samantha (fictitious name) as a legitimate source of feedback and therefore discounted discussion as a source of increased understanding. Moreover, she was unwilling to give Samantha useful feedback. Nevertheless, because they had committed themselves to complete the study, they kept meeting and talking. When problems were identified and explanations required, common sense and excuses ruled the day. When they occasionally did discuss a personal concern, they focused on issues which were of a technical nature. These issues were safe in the sense that teachers can talk about them endlessly without having to question any of the educational assumptions they hold. All that matters is that a solution "works." There were numerous examples of this kind of talk in their post-observation transcripts: "Maybe I tried to do too much in one lesson. I've begun to think – maybe I have too many objectives to try to accomplish in one lesson." "One of my greatest weaknesses is that I always bite off more than I can chew. I always do too much at one time." And, "It turned out to be an exciting experience except that it was just too much for one hour." The issue here is how to fit more content into less time, a technique question. Other technical issues were also prominent: When it is best to reinforce a student's response and how to use techniques picked up in district-sponsored workshops. Unfortunately, because these are safe issues, neither habit nor common sense were challenged.

By comparing and contrasting this dyad with the first three, we concluded that if the aim of evaluation is to increase teachers' understanding of their work through peer interaction, then it is necessary to establish non-hierarchical relations. We are not altogether certain why the failed dyad insisted on establishing hierarchy. Perhaps Samantha was correct – a different pairing may have produced better results. Clearly, teachers must have or must develop respect for one another if they are to learn from one another. We believe that had we provided feedback to Penny and Samantha about the quality of their relation they might have altered it. And too, it may have been that the training was inadequate. We will attend to each of these possibilities in our future work with teachers.

Conclusion

For three of four dyads, horizontal evaluation proved to be an effective means for helping participants question their taken-for-granted assumptions about teaching and learning. In the successful dyads, each participant had a stake in helping the other improve practice by coming to understand more fully what they were doing, why they were doing it, and the moral and ethical consequences of their actions.

Approaching the improvement of practice in this way leads to increased professional development. Part of being professional is taking on

the responsibility of monitoring members' work (Ritzer, 1977). After participating in the study, the teachers began to "see the role of the evaluator as part of being a teacher," an expanded role that they viewed positively. As one teacher put it: "I was pleasantly surprised how easy it was. I think before [the study] I had somewhat of a stigma about evaluation, but I found . . . when two people want it, when two people are trying to improve their teaching, it's a fascinating experience." Professionalism also requires that teachers see themselves as connected, less protective of selfish interests, and less competitive (Darling-Hammond, 1985). One teacher noted how horizontal evaluation might provide such a possibility: "If nothing else [it] could provide an atmosphere of awareness of what we're all doing together, and we will hopefully be unified."

Another teacher echoed this idea: "If teachers used horizontal evaluation, there might be less of a 'I have to protect my little group' [mentality]. I mean, like the first, second and third grade teachers do their thing, and the fourth, fifth and sixth will do their thing. I think there might be some efforts in terms of [establishing] unity." In addition, professionalism means having some control over one's work conditions (Darling-Hammond, 1985). After the study one teacher felt she had moved closer to this position: "You know that no one really has the right or the knowledge to tell us what to do. I guess this process would essentially be giving us back the power that we should have . . . and I think that this is the power we deserve to have as professionals."

Through participation in the study, the teachers believed they became more professional. This is an exciting outcome especially since evaluation, as typically conceived, is scorned by most teachers. It is scorned and feared because teachers generally have so little control over the process and receive so few benefits from it. Our findings are promising: When teachers are encouraged to help one another to understand better teaching, they not only improve technically, but also begin to develop and internalize a professional role which is an essential element in improving the quality of education.

NOTES

1. For a more complete description of horizontal evaluation, including its theoretical framework, see Gitlin, A. and Goldstein, S. (1987), A dialogical approach to understanding: Horizontal evaluation. *Educational Theory*, Winter.

REFERENCES

- Bartz, D. (1978). A Multi-component approach to evaluation for teacher improvement. A paper presented at the National Association of Secondary School Principals, February, Anaheim, California.
- Brophy, J.E., & Evertson, C.M. (1974). Process-product correlations in the Texas Teacher Effectiveness Study: Final report (Report No. 74-4). Austin: University of Texas, R&D Center for Teacher Education.
- Darling-Hammond, L. (1986). Valuing teachers: The making of a profession. *Teachers College Record*, 87, 2, 205-219.
- Florida Coalition for the Development of a Performance Management System. (1983). *Domains: Knowledge base of the Florida performance measurement system*. Tallahassee, Florida: Office of Teacher Education, Certification, and Inservice Staff Development.
- Goldhammer, R. (1969). *Clinical supervision: Special methods for the supervision of teachers*. New York: Holt, Rinehart and Winston.
- Habermas, J. (1976). *Communication and the evolution of society*. Boston: Beacon Press.
- McFaul, S., & Cooper, J. (1983). Peer clinical supervision in an urban elementary school. *Journal of Teacher Education*, 34, 5, 34-39.
- Ritzer, G. (1977). *Working: Conflict and change*. New Jersey: Prentice-Hall.