
In Practice

On Evaluating Mediators

Christopher Honeyman

Most people expect a mediator to be able to persuade. But what is meant by persuasion? Cajoling, begging, threatening, browbeating, arguing on the merits, arranging a demonstration, and any number of other specific tactics all fit into the art of persuasion. At any given moment in a dispute one or all of these approaches may be inappropriate. Moreover, a mediator may be personally uncomfortable with certain tactics, and some mediation programs devoutly assert that *their* mediators wouldn't try to persuade anyone to do anything, stating as a matter of policy that it is for the parties to persuade themselves and each other. Thus even when only one aspect of mediation is considered, a wide variety of possible strategies emerges—none clearly superior to any other. The choice of methods in mediation is therefore so personal, and so specific to the goals and attitudes of the individuals involved, that evaluation of a mediator's effectiveness would seem to defy detailed analysis.

But if a mediator's performance cannot be evaluated according to intellectually respectable standards, we have no dependable way to select those who are best suited to do this work. Furthermore, it becomes difficult to explain what a given mediator could do to improve effectiveness, and a mediator's own claims of unfairness become impossible either to advance or to defend on logical grounds. As the practice of mediation becomes more widespread, and as more people come in contact with it, our inability to define standards of quality may well result in increasing numbers of people who are adversely affected by mediation and who consequently become opponents of the practice. Such opposition could ultimately prevent the adoption of mediation as a dispute resolution mechanism on the scale, for instance, of the litigation system.

The following is an attempt to come to grips with the problems involved in evaluating mediators. It is not intended as a definitive statement, but rather

Christopher Honeyman has played a variety of dispute resolution roles including mediator, arbitrator, and administrative law judge in cases involving labor and environmental issues, for the State of Wisconsin Employment Relations Commission (P.O. Box 7870, Madison, Wis. 53707-7870) and for other programs.

as a basis for discussion. Because mediation encompasses such a vast range of activity under such varied circumstances, any attempt to apply the same criteria across the board is immediately suspect (see, for instance, Luban, 1988). Therefore, I have chosen to direct these observations primarily to situations in which an organized program is responsible for providing competent mediators to assist parties on a "case" basis. Only in part will this discussion be relevant to individual mediators who offer their services directly, or to mediators who operate in the context of a continuing presence in the parties' lives. I will, however, try to address aspects of evaluation that in one way or another contribute to a program manager's effort to deliver *quality of service*, as well as aspects that contribute to an individual mediator's attempt to improve his or her *personal skills*.

Considerations governing the evaluation of entire programs constitute a separate subject—one that is currently receiving great attention (e.g., Luban, 1988; Tyler, 1988; Baruch Bush, 1988; Esser, 1988; Silbey and Sarat, 1988). The focus here is on the performance of individual mediators, and the thesis is that they share certain factors. Using these common elements, it should be possible to construct a rational overall approach to evaluating the means by which mediators achieve their goals; specific programs can then adapt it to their particular needs.

An Unpopular Subject

Almost no one greets the prospect of being evaluated with unalloyed joy—and the psychological pressures work both ways, so that the evaluator, too, has sound reasons to avoid the enterprise. The consequence of this general air of skepticism, I believe, is that programs are often unable to develop their mediators' skills to the utmost or to make sufficiently well-informed judgments as to whom to assign to particularly difficult cases. The potential for staff improvement and "program control" is thus rarely realized. In addition, there is a significant public interest in the evaluation of those individuals who are in a position to exercise influence in public disputes.

At present, mediators tend to be poorly evaluated or even, in effect, not evaluated at all. For example, one researcher recently proposed a system of evaluation for a seasoned group of mediators: In his published article he argues that this large-scale program does not have *any* coherent evaluation system now (Hannon, 1988). Another long-established program recently considered a proposal suggesting that the agency implement a system of evaluation *for the first time* (Bass and Mael, 1988). Yet another mediation service all but abandoned evaluation of its mediators several years ago, in the face of the mediators' concerns that the evaluations were biased and unhelpful.

These are not failing programs propping themselves up in the public consciousness by hiding their faults. They are, I submit, atypical only in their willingness at least to attempt to apply some standard of logic to evaluation, whatever the results may be. Other programs are proceeding down conventional paths which, their advocates maintain, represent the best of all possible worlds in determining what their mediators should do and what they are doing. Anecdotal evidence, at least, suggests that their publics, their funding sources and the mediators themselves remain unconvinced.

There are good reasons for such doubts. Conventional practice tends to identify three criteria by which a mediator's performance can be judged: rate of settlement of disputes, opinions of the parties, and general reputation among the mediator's peers. Each of these has obvious flaws.

Rate of Settlement

Reliance on settlement rates (compared with those of mediators in similar situations) immediately raises the objection that since no two cases are the same, the mediators cannot be fairly compared. For example, some parties have hidden agendas while others do not, and by definition the presence or absence of such complicating factors may be obscure to both the mediator and the evaluator. This objection alone undermines the use of this measurement for mediators who do not mediate regularly.

Even when a mediator's caseload is large, and its distribution random, the use of settlement rates to determine competence begs the question of what *kind* of settlement the mediator has helped parties to reach. In at least some cases, settlements can be obtained by arm-twisting. Or a mediator may help settle a dispute by coming up with a facile idea that hasn't been thought through well enough to work for the parties in the long run. It's not that these are always inappropriate solutions; rather, much of the doubt about the utility of evaluation goes precisely to the point that different approaches to a dispute may yield vastly different results. "More" does not always mean "better".

It is fair to note that in some instances, such as certain court-affiliated programs, the settlement rate may be the essential determinant of a program's survival; under these circumstances it would be unrealistic to expect the program to give primacy to assessments of the *quality* of those settlements. But more secure programs might properly find a well-thought-out, practical settlement on most issues—even if a few remaining issues must be litigated or otherwise disputed—preferable to a comprehensive settlement that leaves all parties hungry for another crack at their opponents.

Opinions of the Parties

Reliance on the parties' opinions introduces other problems. Certainly people do develop strong opinions about particular mediators, but they are often unfamiliar with what may properly be expected of a mediator. And their opinions are just as firm after a single, abnormal case as they would be after twenty years of day-in, day-out exposure to the mediator. Another drawback to relying on parties' opinions lies in the fact that parties to disputes are unlikely, on any routine basis, to devote the time and effort required to give careful answers to detailed questionnaires. Also, they are not privy to what may have happened in a mediator's caucus meeting with another party, so that their point of view is necessarily limited. And they are, of course, partisan; a mediator who has effectively dislodged a recalcitrant party from a beloved position, and thereby helped to settle the dispute, may not be thanked for those efforts.

Yet if a program decides to make a survey of parties' opinions, even parties with scant exposure to a given mediator cannot automatically be excluded: What if this is the only mediator an experienced advocate has ever refused to work with a second time? What if the mediator's conduct somehow favors "repeat players" such as professional advocates, over the one-time-only client? The advocate contacted in a survey would surely respond favorably to a mediator who got him or her out of a jam—even if it were at the client's expense.

These are proper matters for a program to concern itself with; but they are difficult to distinguish from the other motives for similar favorable or adverse responses already noted.

Building a Reputation

Meanwhile, a sterling reputation remains a kind of Holy Grail, much sought after by all mediation practitioners and, I'm afraid, equally difficult to attain. Not only does it take a long time and many cases to arrive at a professional reputation of any consequence in this line of work—in which the “product” simply isn't as clear or public as, say, an architect's—but most of us harbor doubts about the actual competence of one or another highly touted “expert” we've seen at close quarters. This particular form of skepticism assumes special significance in a field in which manipulation of people's perceptions is arguably a common tactic.

A Common Set of Factors

I have already offered the proposition that a mediator's basic talents can usefully be distinguished as five different types of skill: investigation, empathy, invention, persuasion and distraction (Honeyman, 1988). I refer to these five as the component skills of mediation, and will argue that they, together with two kinds of experience discussed below, provide a basis for evaluation.

One of the key problems in evaluation is the difficulty of convincing the mediator and others that the opinions rendered represent something more than a raw application of the evaluator's biases—a problem that has its roots in disagreements over the proper role of a mediator. (Individuals even within the same program may differ, for example, on the degree to which a mediator should try to help out the weaker or less skilled party.)

An initial focus on the component skills of mediation, along with certain other relatively ascertainable criteria, will help an observer at a case, or one of the parties, or even the mediator to assess performance adequately in key aspects. Giving first attention to such nuts-and-bolts questions helps the evaluator avoid clouding the issue. By requiring that each of these skills be considered separately, the evaluator will find it easier to integrate his or her judgments, rather than merely imposing a personal, subjective view of what sort of mediation is best.

There are bound to be differences of opinion as to whether particularly empathetic, aggressive, interventionist, restrained, entertaining, or inquisitive mediators best fit a program's general needs—or the specific requirements of a particular dispute. I have therefore used the component skills to develop examples of standardized *evaluation scales*. These can be used to encourage the evaluator to keep an open mind about such value-laden judgments, at least until preliminary assessments have been made as to specific skill and knowledge levels within defined areas. These standardized scales could be a substantial factor in acquiring the “consent of the evaluated” which is necessary if the mediator is to take to heart whatever suggestions are ultimately made.

The following evaluation scales are an attempt to draw distinctions between various skills that are relevant in at least some kinds of mediation. Though each skill is assessed according to numerical rankings that coincide with apparently concrete descriptions of what constitutes good work, these are mere devices and are therefore in some sense misleading. Please note: *The implied values*

must be rewritten for each type of mediation practice. For that reason, I have deliberately avoided placing any relative weight among the scales. That emphasis will be unique to each mediation program, as it seeks to balance the perceived needs of its particular public against the real restrictions of its budget, the experience of its mediators, and other practical factors. Programs may not, in fact, find it useful to add these scales up to a combined total; such a tally might inspire the objection—with some justice—that the evaluation implicitly downgrades some styles of mediation.

It is also worth noting that in some settings the mediator may not need to possess all of these skills personally, because others present have and exercise them in a way complementary to the mediator's own skills. In some extremely large-scale and important disputes, such as international conflicts, the "mediator" may even consist, by design, of a team of specialists. Thus a few mediators may have available to them the services of assistants who gather facts relevant to the dispute, or even of a "social director" skilled in the arts of distraction. The talents actually required of any one member of such a team would vary accordingly.

What is offered here, then, is not a full-blown system of evaluation applicable to all mediators in all situations, but rather a kind of kit of parts, from which a given program might *assemble* a system suited to its particular needs.

The specific characteristics listed under each scale are designed to focus the inquiry and to reduce the role of rank prejudice. They direct attention first to the relatively ascertainable topic of "What happened?" and encourage the evaluator to register a number of factual observations before tackling the philosophically more difficult question of "What does it mean?" This specificity also encourages the mediator to go along with the enterprise. That matters even from a hard-boiled agency perspective, because mediators are pretty good at diverting attention away from subjects they don't want to address. A *procedure* must be followed if the evaluator's biases are to be kept under control; given this approach, the mediator should have more confidence in the evaluator's ultimate opinion. At the least, a method that helps both the evaluator and the mediator to take the mediator's style into account encourages an honest and adequately grounded discussion.

Seven Parameters of Effectiveness

The first five scales are derived from the mediation hiring examination described in Honeyman, 1988; the latter two represent types of experience not relevant in that exercise. Because they are drafted from the point of view of a caseload in labor relations, they are not appropriate, without adaptation, to all settings. Some examples are given of the kinds of changes that may be necessary, and a discussion of some principles governing the drafting of scales suitable to other programs appears below.

Investigation: Effectiveness in identifying and seeking out relevant information pertinent to the case.

- 3 Asked many relevant and insightful questions, especially early in the process. Vigorously sought to understand facts, reasons, and interests behind initial positions and counter-proposals of the parties. Sought clarification through relevant and important follow-up questions. Systematic, thorough approach

to questioning. Kept track of new information and changing positions (e.g., via notetaking or other mechanism). Subtle analysis of facts being presented.

- 2 Asked at least the obvious questions. Case data was used, but did miss some issues or avenues of questioning. Generally appeared to discover and comprehend the case facts, though not with great depth or precision. Missed at least some aspects of the underlying facts, reasons, or interests of one side or the other. Missed some aspects of settlement possibilities for either side.
- 1 Asked few or mostly irrelevant questions. Appeared at a loss as to what to ask in follow-up questions. Was easily overwhelmed with new information or trapped by faster thinkers. Disorganized or haphazard questioning, filled with gaps and untimely changes in direction. Did not explore the settlement possibilities for both sides on most or all issues.

Empathy: Conspicuous awareness and consideration of the needs of others.

- 3 Avoided appearance of bias or favoritism for or against either party. Asked tough questions of parties, but did so in a sympathetic manner. Demonstrated concern for parties' feelings. Effectively fostered working relationship with parties through actions and attitudes. Listened politely to others and responded with understanding. Conspicuously recognized good points, and the importance of problems and issues, raised by others. Encouraged parties in making their own decisions, did not foist mediator's ideas on the parties unnecessarily.
- 2 Listened to others and did not antagonize them. Conveyed, at least, some appreciation of parties' priorities. Avoided asking some tough questions, thus sidestepping putting self and others in difficult situations at the cost of missing possible opportunities for joint gains. Helped when asked, but missed opportunities to volunteer.
- 1 Asked misleading, loaded, or unfair questions exhibiting bias. Engaged in oppressive questioning to the disadvantage of one of the parties. Threatened more than persuaded. Came into the discussion abruptly to challenge others. Disregarded others' warnings. Saw others' problems as of their own making and did not want to be bothered.

Inventiveness and problem-solving: Pursuit of collaborative solutions, and generation of ideas and proposals consistent with case facts and workable for opposing parties. (Some programs and individual mediators believe that substantive ideas and proposals should only emanate from the parties. But creative results may be, if anything, more difficult to achieve under these conditions. Those working with this restriction may therefore wish to consider rewriting this scale to focus on the mediator's skill at *creating an environment* within which the parties can create the substantive proposals needed, rather than rejecting this element entirely.)

- 3 Avoided commitment to solutions early in process. Recognized underlying problems as opposed to symptoms. Invented and recommended unusual but workable solutions consistent with case facts. Vigorously pursued avenues of collaboration between the parties. Encouraged parties themselves to seek and develop new solutions. Thought and acted without being urged.

-
- 2 Interrelated at least some proposals and compromises with ideas of other party. Worked well with solutions parties suggested, but did not pursue inventive or collaborative solutions. Appeared to comprehend case facts/problems as they developed, though not with great depth. Allowed collaborative problem solving, but did not stimulate it.
 - 1 Prematurely tried to come up with solutions, pushing to judgment prior to establishing essential facts. Ideas were ineffective and unworkable. Waited for things to happen. Blocked efforts at seeking collaborative solutions. Did not initiate suggestions; required considerable help from the parties.

Persuasion and presentation skills: Effectiveness of verbal expression, gesture, and "body language" (e.g., eye contact) in communicating with parties. (With persuasion, again, there is a sharp difference of opinion between programs operating in different areas as to what degree or kind of activity is desirable. Some programs may wish to rephrase this scale in terms of the mediator's ability to create an environment conducive to the parties' attempts to alter each other's and their own preconceived opinions.)

- 3 Demonstrated particular skill, confidence and persuasiveness in verbal communications throughout. Data presented and manner of presentation effectively altered positions of parties. Remained unflustered from start to finish; articulate and enthusiastic. Maintained eye contact and positive gesture; competently used all tools of communication. Was easily understood and logically organized.
- 2 Generally clear and concise communications. Choices of what to present and manner of presentation did not compromise goals of resolution. Generally but not always at ease with situations presented. Points and comments were sufficiently well organized and presented; but not particularly forceful. Eye contact and other gestures used adequately.
- 1 Presentations not well related to goals of resolution. Was difficult to understand or unclear in expression. Had little or no impact and did not persuade. Appeared flustered and uncomfortable most of the time. Readily withdrew when challenged or questioned. Little or no confidence expressed. Halting gestures, poor eye contact.

Distraction: Effectiveness at reducing tensions at appropriate times by temporarily diverting parties' attention. (This scale is drafted for situations in which the parties consider themselves professionals or are relatively detached for other reasons. There is evidence that the use of humor is quite dangerous in settings where the parties are more emotionally charged, such as divorce mediation [Orbeton, 1989], and this may also apply where the mediator is not intimately familiar with all of the "micro-cultures" present in the particular dispute. Under these conditions it may be appropriate to omit any reference to humor and/or to combine the remaining aspects of distraction with the scale for "managing the interaction," which follows.)

- 3 Demonstrated acute sense of rising tension; invariably had quip or other tactic ready to disarm situation. When allowed tension to rise, did so to good purpose, such as to provide for venting of emotions or to demonstrate hollowness of a proposal or position. Had wide variety of techniques

for redirecting parties' focus away from sullen or otherwise unproductive colloquies.

- 2 Generally recognized signs that discussion had turned sour, took action to try to redirect it. Appropriate use of humor, anecdotes, breaks, or switching to different subjects of discussion. Not always effective at lightening the atmosphere.
- 1 Made little or no effort to provide perspective on the parties' problems or to engineer lighter moments. Little or no sense of humor apparent. Opportunities for breaks ignored, or constant hammering away at difficult subjects made it hard for parties to see their dispute as a small part of their lives.

Managing the interaction¹: Effectiveness in developing strategy, managing the process, coping with conflicts between clients and professional representatives. (As noted above, this scale could be redrafted to include aspects of distraction in settings where humor is considered inappropriate or too risky.)

- 3 Made all decisions about caucusing, order of presentation, etc., consistent with rationale for progress toward resolution. Managed all client-representative relationships present effectively. Handled emotional tensions and outbursts so as to encourage settlement. Gave appearance of being ready to cope with any exigency.
- 2 Controlled process, but decisions did not reflect a strategy for resolution. Did not dominate, but was not overwhelmed by, factual or legal complexities. Did not allow bullying by clients or representatives.
- 1 Decisions on procedure and presentation were unjustified. Was confused or overwhelmed by factual or legal complexities. Allowed clients or representatives to control process in ways counterproductive to resolution.

Substantive knowledge: Expertise in the issues and type of dispute. (It is not established that substantive knowledge is an essential part of a mediator's background. Like the parties, an experienced mediator could tend to overlook the existence of certain assumptions that are no longer valid; someone not burdened with ingrained ideas may be able to bring a fresh approach. There may well be circumstances in which, for instance, a program would deliberately choose to assign a complex dispute to a mediator lacking substantive knowledge but known to be an effective investigator.)

- 3 An authority in the subject field. Demonstrated knowledge without arrogance. Was able to identify all or most known solutions to common problems and adapt them to fit present circumstances. (Note: This is distinguished from inventiveness.)
- 2 Possessed good working knowledge of field, equivalent to that of an ordinary practitioner in that field. Did not necessarily identify subtleties or limits of field's capacity or development. Demonstrated knowledge through reasonably coherent explanations.
- 1 Demonstrated little knowledge of specific field in which dispute takes place. (Note: This is distinguished from knowledge of unrelated fields).

The seventh scale is, like the sixth, related more to age and experience than to raw talent. These last two scales are therefore more useful for evaluating experienced mediators as opposed to those in training. But some programs may find in them a means for distinguishing a mediator who is familiar with simple problems from one who has high potential unaccompanied by experience, which the other scales by themselves do not necessarily provide. Such assessments could be helpful both in the selection of mediators and in training. For example, one program confronted with two potential mediators might, in the light of an adequate present staff, opt for the "high potential" mediator with an eye to its long-term needs, while another with an immediate case overload might prefer the mediator better prepared to leap into the fray forthwith. Or, differences on these scales between two trainees, both equally highly regarded overall, might well lead a program to vary the emphasis of its training between the two. (In some circumstances, the application given to these scales will seem slightly perverse at first glance; of two mediators who rate equally well on the first five scales, the one to watch over the long haul is probably the one who managed that feat *without* the kind of background that would lead to high ratings on the last two scales.)

The standard in all of the scales is deliberately set high, in order to make them relevant to the most competent and well-rounded mediator. Unfortunately this means that, initially, they will appear rather daunting to anyone else. It is only fair to point out that although I have known a good variety of mediators, I do not know of any to whom all of the highest "behavioral statements" would apply on all these scales. There are inherent conflicts among some of the character traits involved, so that, for instance, the most inventive of mediators is rarely the most empathetic.

The five exemplary mediators described in Honeyman (1988), for example, would each come in high on two or three of these scales (different ones in each case), and at least at "2" on the others; and that is as close to perfection as you are likely to meet. (I regret to admit that I myself wouldn't do as well as they.) Nevertheless, this kind of standard provides a target for any mediator to contemplate; it neither encourages complacency nor, with the caveat noted already, is it utterly unattainable. And there is nothing to prevent any program, in redrafting the scales to address its particular needs, from moderating the expected standard of performance.

Redrafting must also take account of ethical differences between programs, including such questions as whether the mediator "empowered" the weaker party—or should have. Empowering the weaker party is considered by some programs to be a proper responsibility of the mediator in attempting to achieve fair and lasting results; in others, the same action is seen as pernicious meddling in a relationship in which the application of power is an accepted fact (SPIDR 1986.)

One other note concerning the scales seems worth emphasizing. These rankings are general statements, and the numerical scores are intended solely as a means to encourage the evaluator to make the difficult summation of what may be varied opinions of the mediator's actions among the various aspects of each quality noted. Thus under "empathy" a highly empathetic mediator might match all the characteristics of the top group of statements except for the quirk of never showing a speaker that he or she had made a good point.

Adjusting a rating to account for this helps tell the mediator how important the evaluator thought it was. At the same time, the full range of characteristics laid out in each scale helps the evaluator clarify impressions, make notes, and explain to the mediator what the evaluator was pleased or concerned about. Assuming that a forthright discussion ensues, the scales should help the evaluator to be thorough and, at the same time, enhance the mediator's ability to see his or her actions through another's eyes.

A Choice of Evaluators

I have already noted that the mere mention of evaluation can be enough to raise mediators' hackles. Yet it clearly has the potential to be a positive process, one that can help the mediator improve his or her skills. Whether any given attempt lives up to that potential, or instead becomes an exercise in mutual bitterness and mistrust, depends on how and by whom it is done.

As discussed earlier, the procedure used in evaluating a mediator can be made reasonably objective by deferring questions of what type of mediation is best to a discussion separated from judgments of the mediator's capacity in each of the component elements. But we're hardly out of the woods at that point, for the wrong choice of observer/evaluator can inhibit any good that might result. There are several types of evaluator, and each has particular strengths and weaknesses.

Program managers or, in larger organizations, supervisors, are most commonly called upon to evaluate, if only because everyone seems to assume that it can't be done by anyone else. They also offer certain advantages in what is likely to be an intricate and time-consuming process. They are, presumably, readily available, interested in the problem, and already budgeted for. And they are—again, presumably—familiar with the ins and outs of the program involved, and understand the circumstances within which a particular mediator must operate.

But from several angles the use of program managers as evaluators seems less advantageous. First, the system of evaluation proposed here is heavily dependent on observation. Observation is labor-intensive, and unless the program is blessed with managers who have some time on their hands, the prospect for actual implementation of a system requiring such a substantial commitment may be illusory. Consider, for example, that programs in existence at the time did not generally adopt even so relatively effortless a performance-measurement tool as the "group performance test" advocated by Gellhorn and Brody (1948).

Not only is the expected "availability" argument thus open to question in practice, but when managers do show up, their presence may lead to unintended distortions. Any observer, say the scientists, classically alters the thing observed; and mediation is a remarkably mutable process. Parties may "play to the gallery"; the mediator may act out of character; and managers who are themselves mediators have been known to find the observer's role of noncommittal silence unendurable and have ended up interfering. Moreover, the presence of a program manager, unless convincingly explained as part of a routine rotation, could be seen by the parties as a signal that their dispute has special significance, or that the mediator lacks management's confidence.

And these are what could be called benign distortions. More worrisome is the fact that not every program manager sees eye-to-eye with every mediator

on the proper direction of the program's efforts; personal dislike or distrust may also be involved. Direct observation by a manager could be seen by a mediator already disenchanted with that manager as a pernicious attempt at intimidation or an indication of some other trouble in the employer-employee relationship. For all of these reasons, other avenues of evaluation deserve more than passing comment.

The parties constitute the second most common group of potential evaluators. Some of the problems presented by their use were noted earlier, but it must be admitted that to the extent that any program is market-driven it should provide for parties' opinions of its efforts to be given some weight. It is difficult to postulate, however, that parties could themselves use effectively the specific tools advocated here—if for no other reason than because, in general, each party sees only a limited view of what the mediator is doing.

For example, Party A, who is out of the room, can't hear the mediator come up with a clever solution to Issue X and try it on Party B. If Party B rejects it out of hand, the mediator may never mention the idea to Party A in the next caucus. But if Party A had itself thought of the same idea, and was for tactical reasons waiting for it to be offered by the mediator, the mediator's silence would logically be read as lack of inventiveness. This would then show up in a reduced score on the associated scale. Thus doth confidentiality make apparent idiots of us all.

Clearly then, mechanisms should be developed to permit program managers to register parties' views in the context of judgments of the mediator's skills reached independently. Otherwise, with all goodwill, the mediator will rely on his or her own self-image, the program manager will have other views, and the parties will go their own way, without any of the inevitable contradictions being resolved.

The difficulties managers present as evaluators make it necessary at least to consider the pluses and minuses of using some sort of *independent consultant*. If the mediator agrees to an outsider's participation, the perspective such a person can provide becomes valuable. It seems probable, also, that someone not directly in control of the program would be less likely to engender out-of-character behavior by the parties. But while low-budget programs may sometimes be able to call on volunteers to help in this way, a consultant able to inspire confidence is unlikely to come cheap. Thus in practice the limited availability of qualified volunteers and the extraordinary cost of professional consultants restrict the use of such "outside experts." Still, in such an instance as a claim of unfairness in an earlier evaluation, the possibility should not be overlooked.

One other point deserves mention. There is considerable variation between even experienced observers in their "focus" during a mediation session, as a recent experiment by the Law and Society Association established (Honeyman and Nielsen, 1989). Observers at a mediation role play differed substantially in their impressions of the parties' options, the mediator's actions, the background facts remembered, and the quality of the results achieved. There is always some doubt, then, about whether even the most assiduous and competent observer is really using the same facts, or following the same train of thought as the mediator. Common sense suggests that this problem is greater for outsiders, but with any observer it argues for a discussion with the mediator before the evaluator commits him- or herself to any judgment, lest the passage of time

and pride of authorship make correction difficult. An observer probably cannot, therefore, simply go home after a case and write up a finished evaluation if he or she expects the opinions rendered to be either valid or heeded.

Because of the solitary nature of mediation, *self-evaluation* seems to me a particularly appropriate tool, at least for “improvement” rather than “program control” purposes. Any mediator can be given a set of grading scales similar to those discussed above and invited to consider his or her own performance on a particular case (perhaps privately) against each scale. Merely supplying copies of such objective criteria to mediators can give an astute mediator something to think about when trying to come up with an approach to a case problem. And if a peer is present, that “evaluator’s” role then is reinforced as being primarily to help the mediator rather than to help enforce the norms of the program. An observer assisting in this way could, I think, press questions relating to performance without offending the mediator in the way that a flat judgment would be likely to.

Anecdotal evidence from the use of self-evaluation in other fields suggests to me that when it is implemented in the right spirit, people respond accordingly—and that they become far more critical of themselves than they would tolerate from anyone else. Self-evaluation palpably depends on a sense of professionalism, but it also fosters one. It also offers some other benefits. Recrimination is avoided; the cash cost is minimal; and it is probably the only system of evaluation that can be implemented in a “politically” acceptable fashion within a voluntary association, such as an uncompensated group of volunteers or a partnership.

There are, of course, drawbacks too. There seems to be a curious tendency for the most self-critical performers to be some of the best—perhaps because they can afford to admit their failings, or perhaps because they were self-critical to begin with, listened to their inner voices, and then improved until they *became* the best. This raises the problem of the converse—the mediator who becomes defensive and can’t or won’t admit failings. I would argue that such a person is, at any rate, even more likely to avoid criticism from others, who can be passed off as biased or as not knowing the facts. Use of self-evaluation cannot be guaranteed to avoid undue self-praise, but if an observer is used in conjunction, the observer’s presence tends to encourage honesty.

There is, moreover, the problem of emphasis. Mediators tend to rely on their own particular strengths; to do so is rational for the purposes of any given case. But it does mean that, for instance, an empathetic mediator is not likely to treat even a recognized (relative) inability to be hard-nosed as very important, while disputes unfortunately cannot be predicted to call for the one skill to the exclusion of the other.

This suggests that a system of self-evaluation must involve the use of a standardized and comprehensive set of questions. Otherwise, even the most committed mediator will tend to be self-critical in areas where relatively little objection to his or her performance can be made—and will overlook more glaring faults in areas that mediator, by disposition and habit, thinks less significant.

For these reasons, an effective system of self-evaluation would appear to require the presence of an observer. The process of elimination leads to the proposal that this should ordinarily be a *peer*, who would be present specifi-

cally to help the mediator confront weak points. In other fields, peer review is a time-honored practice, offering most of the advantages of independent consultants with little of the expense. Because the adoption of this method as a system promises mutuality in such an effort, I believe it would be acceptable to the mediators involved; but as with any approach to evaluation, it requires tact and consistency in the execution.

Within the confines of a particular program, there is always the possibility of favoritism among the practitioners, of lackadaisical work by the evaluator, or of outright collusion. And some legitimate purposes of evaluation cannot effectively be accomplished by someone who has a peer relationship with the person evaluated: If "hard choices" are in the offing, it would be unrealistic for management to foist that responsibility on a person who may not be able to afford it in terms of continuing personal relationships. Peer evaluation can never entirely supplant program managers' review, but for continuing training purposes, it has much to offer. Its adoption on a regular basis indicates an attitude of professionalism on management's part, which is likely to be reciprocated by the mediators themselves. The costs are minimal. The necessary observation is more likely to be performed on the necessary frequent basis than it would be if managers were to retain that function for themselves. And the results are likely to be heard by the mediators with less resentment, and in turn responded to more constructively.

The fact that peers cannot serve as the sole evaluators in management's stead must be recognized. One way to reinforce this point is to separate the function of program control from that of skill-building, and to identify peer evaluators as colleagues who are working primarily for the *mediator* as opposed to management. Ideally, this would involve specifying their output as private communication, and relying more on other evaluation sources for program control purposes. One program experimenting with the evaluation approach suggested here has established a small, specially trained group of experienced mediators, within the larger group of about fifty, for essentially this purpose (Orbeton, 1989). While the attempt is too new to be appraised yet, the intent is to make communication more forthright, and observation much more frequent and thorough, than would be possible if managers had to act as the sole evaluators.

Conclusion

Evaluating mediators is a complex process, but not an impossible one. While no single solution is likely to be found, a set of options is emerging. Any new refinement, admittedly, brings with it new difficulties, and the options laid out here are themselves complex to administer. An adroit program management may be able to put together relatively quickly a workable, efficient, and fair approach to evaluation that is tailored to its own circumstances. But most likely, the process of developing evaluation tools will require sustained effort, justified partly by recognition that only trial and error will eventually produce a result keyed to the program's, the parties', and the mediators' diverse needs.

Nevertheless, it should be apparent that avoidance of the problems is no longer an acceptable strategy. In an era when rational standards for judging the elements of mediators' effectiveness are becoming more refined, and when mediation itself is becoming an increasingly common option for resolving all

kinds of disputes, retaining public confidence in any program will demand that the program devote time and effort to evaluating and strengthening its most important resources.

NOTES

An earlier version of this article was presented under the title "Problems in Evaluating Mediators," at the North American Conference on Peacemaking and Conflict Resolution, Montreal, Quebec, March, 1989.

Christina Sickles Merchant, Byron Yaffe, Stephen Goldberg, Jeanne Brett and Martha Askins offered detailed and helpful critiques of earlier drafts of this work. I am particularly grateful to the mediators of Maine's Court Mediation Service and to its directors, Jane Orbeton and the late Lincoln Clark, for their willingness to take the risks of applying an untried theory and their help in developing it. And once again, my colleagues at the Wisconsin Employment Relations Commission provided numerous and significant comments and criticisms which have corrected my thinking on a number of points. However, the opinions expressed here are the author's, and do not necessarily reflect the policy of the WERC.

1. The scale for "managing the interaction" was developed by the managers of the Suffolk County (Mass.) Superior Court's program for mediating civil litigation. I am indebted to them for its use here; see Honoroff, Matz and O'Connor (1990).

REFERENCES

- Baruch Bush, R.** (1988). "Defining quality in dispute resolution: Taxonomies and anti-taxonomies of quality arguments." Madison, Wis.: Working Paper Series, Disputes Processing Research Program, Institute for Legal Studies, University of Wisconsin.
- Bass, A. and Mael, E.** (1988). "Report to Michigan Employment Relations Commission." Unpublished paper.
- Esser, J.** (1988). "Evaluations of dispute processing: We don't know what we think and we don't think what we know." Madison, Wis.: Working Paper Series, Disputes Processing Research Program, Institute for Legal Studies, University of Wisconsin.
- Gellhorn, W. and Brody, W.** (1948). "Selecting supervisory mediators through trial by combat." *Public Administration Review*, autumn, 259-266.
- Hannon, J.** (1988). "Performance appraisal for federal mediators: A scientific approach to assessing the art." *Labor Law Journal* 39: 91-100.
- Honeyman, C.** (1988). "Five elements of mediation." *Negotiation Journal* 4: 149-160.
- Honeyman, C. and Nielsen, D.** (1989). "Roadster meets dent: An inquiry into research in mediation," a videotape and teaching notes. Cambridge, Mass.: Clearinghouse, Program on Negotiation at Harvard Law School.
- Honoroff, B., Matz, D. and O'Connor, D.** (1990). "Putting mediation skills to the test." *Negotiation Journal* 6: 37-46.
- Luban, D.** (1988). "The quality of justice." Madison, Wis.: Working Paper Series, Disputes Processing Research Program, Institute for Legal Studies, University of Wisconsin.
- Orbeton, J.** (1989). "Commentary on applying a new theory of evaluation." Presented at the North American Conference on Peacemaking and Conflict Resolution, Montreal, Quebec, March, 1989.
- Silbey, S. and Sarat, A.** (1988). "Dispute processing in law and legal scholarship: From institutional critique to the reconstitution of the judicial subject." Madison, Wis.: Working Paper Series, Disputes Processing Research Program, Institute for Legal Studies, University of Wisconsin.
- SPIDR** (1986). Code of Professional Responsibility. Washington, D.C.: Committee on Ethics, Society of Professionals in Dispute Resolution.
- Tyler, T.** (1988). "The quality of dispute resolution processes and outcomes: Measurement problems and possibilities." Madison, Wis.: Working Paper Series, Disputes Processing Research Program, Institute for Legal Studies, University of Wisconsin.